

Instituto Universitário de Lisboa — ISCTE-IUL
Faculdade de Ciências da Universidade de Lisboa

Departamento de Finanças do ISCTE-IUL
Departamento de Matemática da FCUL



Bonds With Credit Risk and Call Provisions

Pedro Filipe Matos da Cruz

Dissertação
Mestrado em Matemática Financeira

2014

Instituto Universitário de Lisboa — ISCTE-IUL
Faculdade de Ciências da Universidade de Lisboa

Departamento de Finanças do ISCTE-IUL
Departamento de Matemática da FCUL



UNIVERSIDADE
DE LISBOA

ISCTE  Business School
Instituto Universitário de Lisboa

Bonds With Credit Risk and Call Provisions

Pedro Filipe Matos da Cruz

Dissertação
Mestrado em Matemática Financeira

Orientadores:
Professor Doutor João Pedro Nunes
Doutor Pedro Baltazar

2014

Abstract

In this thesis the reduced form models for pricing non-callable and callable corporate bonds are studied. The default and call times are considered as the first jump of a Cox process, as done in Lando [26]. With the recovery of market value assumption of Duffie and Singleton [16] we show that the price of defaultable non-callable and callable bonds can be written as the price of a non-defaultable bond, but considering an adjusted rate. With this we can use the results known from short-rate models, particularly the affine term structural models. We consider the models of Duffie [11], Jarrow et al. [20] and Park and Clark [34] to price non-callable bonds, callable bonds and bonds with make-whole call provisions, respectively. Given the non-linear dependence on the state variables we have to use the results of Kimmel [24] to find a closed-form approximation for the price of bonds with call provisions. We then used extended Kalman filters to estimate the models parameters and see how well they fit to corporate bond data.

Keywords: corporate bonds, callable bonds, Kalman filter, Kimmel series expansion, reduced form models.

Resumo

O preço de um qualquer ativo financeiro depende dos riscos que estão associados aos seus pagamentos prometidos, de modo que uma quantificação de esses riscos é essencial. Um investidor está apenas disposto a incorrer em certos riscos se tiver a perspectiva de ser adequadamente remunerado por eles. Nesta tese tem-se por objetivo a descrição de modelos que quantifiquem o risco de incumprimento e o risco associado às provisões de resgate de obrigações.

Um dos métodos mais utilizados para analisar o risco de incumprimento são os chamados modelos estruturais que ligam a probabilidade de incumprimento ao valor dos ativos da empresa. Estes têm no entanto algumas limitações. Em particular nestes modelos não são possíveis quedas acentuadas no preço de obrigações emitidas por uma qualquer empresa, o que é muitas vezes observado na prática. Isto porque não é possível nos modelos estruturais que a falência do emitente surja como um evento imprevisto.

Outro tipo de modelos utilizados para modelar o risco de incumprimento são os modelos em forma reduzida. Nestes podemos ter de facto a falência de uma empresa como um evento inesperado. O evento de incumprimento é modelado pelo primeiro salto num processo de Cox e podemos assim obter uma redução drástica no processo de uma obrigação. Além disso, também o risco associado à compra de uma obrigação com provisão de resgate pode ser modelado através dos modelos em forma reduzida, uma vez que podemos considerar o evento de resgate como o primeiro salto de um processo de Cox. Os processos de Cox são descritos no Apêndice C enquanto que uma introdução às diferenças entre modelos estruturais e modelos em forma reduzida é feita no Capítulo 1. Os resultados matemáticos mais importantes surgem no Apêndice B, enquanto que definições sobre obrigações se encontram no Apêndice A.

No Capítulo 2 são apresentados os modelos em forma reduzida na perspectiva de Lando [26], ou seja, usando os processos de Cox. Com o pressuposto, devido a Duffie e Singleton [16], de que recuperamos uma fração do valor de mercado da obrigação no momento de incumprimento ou no momento de resgate, mostra-se que o valor da obrigação num dado momento pode ser escrito da mesma forma que o valor de uma obrigação que não está

sujeita a risco de incumprimento, mas agora a taxa de desconto é uma taxa efetiva, no sentido em que pode ser decomposta em várias componentes, sendo uma delas a taxa usada para avaliar obrigações sem risco de incumprimento, mas adicionando agora uma componente devido a esse risco. Isto permite que usemos os métodos já conhecidos para avaliar obrigações do tesouro (sem risco de incumprimento) e aplica-los a emitentes que podem falir. Em particular vamos utilizar os modelos afim na especificação de Dai e Singleton [7], descritos no Apêndice D.

No Capítulo 3 são descritos os modelos que utilizam os modelos afim com o objetivo de descrever as taxas efetivas. Em particular é utilizado o modelo de Dufee [11] para avaliar obrigações com risco de incumprimento mas que não têm provisões de resgate. O autor utiliza uma forma menos geral da que é permitida pelos modelos afim, mas que é muito utilizada na literatura — os modelos CIR. Nestes modelos as variáveis que explicam a taxa efetiva tendem para um valor de equilíbrio, podendo deslocar-se deste dado um forçamento estocástico e que depende do valor da variável. Isto é requerido pelas observações para o caso das taxas do tesouro. Para o caso de obrigações que não só têm risco de incumprimento como também têm provisões de resgate, em que esse resgate é feito a um custo fixo pré-determinado, então utilizamos o modelo de Jarrow et al. [20]. Semelhante ao de Dufee [11], socorre-se dos métodos CIR, mas agora apresentando a taxa efetiva a depender de forma não-linear de uma das variáveis de estado que explicam essa taxa efetiva. Não são conhecidas fórmulas analíticas para a nova forma do fator de desconto dos pagamentos da obrigação. No entanto, os resultados de Kimmel [24] permitem-nos obter uma aproximação em séries de potências que é uniformemente convergente, querendo isto dizer que a convergência da série para a função que descreve o valor esperado do valor descontado dos pagamentos prometidos não depende da maturidade da obrigação que se está a considerar. Estes resultados, devidos a Kimmel [24], são também úteis para o caso do modelo de Park e Clark [34], que tenta modelar a taxa efetiva de uma obrigação com risco de incumprimento e com provisão de resgate, mas agora o preço que o emitente paga pelo resgate não é um valor fixo pré-determinado, mas antes um valor que é o máximo entre o valor facial da obrigação e o valor descontado dos pagamentos restantes prometidos descontados a uma taxa que é a soma da taxa do tesouro que vigora no momento do resgate mais um prémio a que se convencionou chamar de prémio de resgate. Tem-se no entanto a semelhança de uma taxa efetiva não-linear nas variáveis de estado, sendo essa forma de não-linearidade igual ao caso de Jarrow et al. [20] e portanto podem-se aplicar os resultados de Kimmel [24] que funcionam com esta forma de não-linearidade. Com todos estes três modelos para as respetivas taxas efetivas escrevem-se os respetivos valores das obrigações em termos dos parâmetros das dinâmicas das variáveis de estado e dos próprios valores dessas variáveis de estado.

Tendo as fórmulas analíticas para a avaliação das obrigações podemos usar dados observacionais de taxas do tesouro e de preços de obrigações para comparar com os modelos, o que é feito no Capítulo 4. Tem-se que obter o valor dos parâmetros e das variáveis de estado, o que é feito com filtros de Kalman descritos no Apêndice F. Estes consideram que cada observação pode ter um erro observacional associado e, para um determinado vetor de parâmetros, obtêm o valor das variáveis de estado que faz com que o erro de previsão de uma observação dado todos os valores observacionais anteriores seja mínimo. Em conjunto com um método de minimização que funciona em paralelo varia-se o vetor de parâmetros de modo a obter os erros mínimos. Isto é feito em primeiro lugar para taxas de tesouro usando um método CIR com dois fatores. De seguida faz-se o mesmo com preços de obrigações para dois emitentes diferentes, cada um com uma obrigação com risco de incumprimento e sem provisão de resgate, tendo um deles ainda uma obrigação com provisão de resgate a custo fixo e o outro uma obrigação com provisão de resgate a preço estocástico. Em todos os casos podemos, com o valor dos parâmetros e dos valores das variáveis de estado obtidos, criar um vetor de preços através das fórmulas de avaliação dadas no Capítulo 2 e compará-las com os dados observacionais. Além disso, para obrigações com provisão de resgate a custo fixo, considera-se ainda um modelo semelhante ao modelo de Jarrow et al. [20], mas ligeiramente alterado de modo a que seja incluída mais uma dependência não-linear numa variável, o que é fundamentado pelo mesmo argumento que nos leva ao modelo de Park e Clark [34].

No Capítulo 5, e considerando os resultados que se obtiveram no Capítulo 4, conclui-se quanto à capacidade dos modelos usado em explicar as observações.

Palavras-Chave: obrigações, resgate de obrigações, filtros de Kalman, expansão de Kimmel, modelos de forma reduzida.

Acknowledgements

I want to thank my supervisor Pedro Baltazar for all the comments and discussions, and also for all the headaches. I also thank Professor João Pedro Nunes for accepting the co-supervision of this work.

A special thanks to *InvestQuest* for the opportunity and all the time and resources spent in helping me. It certainly has been a great experience.

Mainly, I thank my family and Catarina for all the support.

Contents

Abstract	i
Resumo	iii
Acknowledgements	vii
Contents	ix
1 Introduction	1
2 Defaultable Bonds in the Reduced Form Models	5
2.1 Defaultable Bonds	5
2.2 Call Provision	11
3 Model Specification	15
3.1 Non-defaultable Bonds	15
3.2 Defaultable Non-Callable Bonds	18
3.3 Defaultable Bonds With Call Provisions	20
3.3.1 Defaultable Callable Bonds	22
3.3.2 Defaultable Bonds With Make-Whole Call Provision	23
4 Results	25
4.1 Estimation of the Default Free Process	25
4.2 Estimation of the Default Spread	29
4.3 Make-Whole Callable Bonds	33
4.4 Standard Callable Bonds	36
5 Conclusions	39
A Bond Concepts	41
B Probability Concepts	45
C Cox Process	51
C.1 Default as a Poisson Process	51
C.2 Default as a Cox Process	53

D	Affine Term Structure Models	55
D.1	The Completely Affine Class of ATSM	57
D.1.1	Excess Returns	59
D.2	The Essentially Affine Class of ATSM	61
E	Kimmel Series Expansion	63
E.1	Power Series Solution to the Pricing Problem	65
F	Kalman Filter	71
F.1	The Linear Kalman Filter	71
F.1.1	Kalman Filter With The Multifactor CIR Model	76
F.2	Extended Kalman Filter	78
G	Matlab Codes	79
	Bibliography	81

1

Introduction

price equals expected discounted payoff. The rest is elaboration, special cases, and a closet full of tricks that make the central equation useful for one or another application

– John H. Cochrane, *Asset Price Theory*

Embedded in the price of any financial securities are the risks and uncertainties of the promised pay-off's — The buyer is only willing to bear those risks if he is adequately paid for taking them. One of the most important risk factors is *credit risk* — the possibility of default by the issuer. Together with the time value of money and interest rate risk this explains most of the price of fixed income securities such as corporate bonds. The two most used approaches for modelling credit risk are the structural models and the reduced form models.

In structural models, first developed by Merton [30], we have a fundamental representation of default. This event occurs when the value of the firm assets fall below a deterministic or stochastic boundary, exogenously or endogenously given, linked to the value of the liabilities. This way we have an economical reason for default and the model parameters can be obtained from the firm balance sheet. The probabilities of default and yield spreads of this firm follow as a consequence of this balance sheet data and considering that the asset values follows some stochastic process, usually a geometric Brownian motion.

Giving this fundamental description of default has the inconvenient of having a not so flexible model because default can only occur in very specific occasions. The value of the assets diffuse and hit a barrier, but in practice we sometimes see a drop (or several drops) in market prices of the firm securities that reveal default as a sudden event, as seen in Figure 1 for a subordinated bond of Banco Espírito Santo.



FIGURE 1.1: The market prices of a Banco Espírito Santo bond between February 2014 and August 2014. In early August we see a jump in the price (source: Bloomberg).

Calibration of structural models is also a problem because the assets of the firm are not directly observed, we do not have balance sheet information for every moment in time, and this information may even be unreliable. With this in mind, Duffie and Lando [15] develop a model similar to that of Leland and Toft[28] where the default barrier is given endogenously as a consequence of an optimal decision made by equity holders, but including incomplete information where bond investors do not observe the value of the assets directly. They receive instead periodic and imperfect reports where the investor does not have the correct value of the assets. In this incomplete information setting the existence of a default intensity arises naturally. The existence of a default intensity is a characteristic of reduced form models and because of this the model of Duffie and Lando is considered a hybrid model.

In the *reduced form models* [11, 17, 19, 32] (also called intensity models) default is modelled as the time of the first jump of a Poisson process. The probability of default is modelled not worrying about explaining the fundamentals of default, i.e, not considering the value of the assets directly, and therefore it as been said that these models have little economical content, but the results of Duffie and Lando, to some extent, give an economical substantiation for the reduced form models. These type of models give more flexibility and will also give us mathematical tractability because we will be able to use the machinery used in term structure models to obtain closed form solutions. Calibration is not a problem when we have a liquid market of corporate bonds, but when we do not have it we still have to rely on structural models.

Another risk source appears in contracts that nowadays constitute a high fraction of the US bond market — bonds with call provisions. In these contracts the issuer can redeem a bond at a predetermined cost or through a pre-determined random cost, in the sense that it will pay at the call moment the value of the remaining payments discounted at the risk-free rate plus a fixed premium — the make-whole call premium. In the first case we say we have a standard callable bond and in the second a bond with make-whole call provision.

In this work we model this risk factors using reduced form model. This is done by Jarrow et al. [20] and Park and Clark [34] — the papers in which this thesis is based. These models have the advantage of easily considering market friction and non-optimal call policies, contrary to the approach of Duffie and Singleton [16] where market frictions play no role in the most simple case and the evaluation is done considering optimal decisions. The model of Jarrow et al. is proposed to evaluate callable bonds in general, but Park and Clark notice that the model is only suitable for standard callable bonds and propose an alternative model to evaluate bonds with make-whole call provisions. In both cases the call provision can be of Bermudan or American type, i.e., the bonds can be called at pre-determined times or at any time, but no distinction is made between the two in these models.

In this work the reduced form models are described in the framework developed by Lando [26]. The basic construction of the Cox processes is given in Appendix C and the equations that we will need to solve to price both non-callable and callable bonds are given in Chapter 2. In Chapter 3 the particular models of Jarrow et al. [20] and Park and Clark [34] are used to evaluate these bonds. To find closed form solutions we rely on the affine term structure models of Duffie and Kan [14] in which the short-rate is an affine function of the variables that explain it (see Appendix D). Given that in the models for callable bonds we have the discount rate as an non-linear function of some of the variables, we also need the Kimmel series expansion [24] (explained in Appendix E) to find an approximation for the pricing problem. The numerical implementation is done using extended Kalman filters (Appendix F) and the results of fitting the model to the data are given in Chapter 4. Chapter 5 concludes.

2

Defaultable Bonds in the Reduced Form Models

In the following section the basic formulas to evaluate zero coupon defaultable callable bonds in the reduced form models are given. We consider the case where, at the default time, the holder receives a fraction of the market value just before default. This makes possible to write the value of these bonds as the expected discounted value of the face value of the bond, just as in the case of risk-free bonds, but now with an effective rate that is the sum of the risk-free rate plus a default spread. In the case of callable bonds we will also have to add the call spread.

2.1 Defaultable Bonds

In this section the objective is to evaluate a non-callable defaultable bond. We start by considering a bond that in case of default pays nothing to the bondholder. The payoff at maturity T may be random but it is \mathcal{F}_T -measurable. The main idea is to consider the default time τ_d as the first jump of a Cox process $N_{d,t} \equiv \mathbf{1}_{\{\tau_d > t\}}$. In this case $\lambda_{d,t}^{\mathbb{Q}}$ ¹ represents the default intensity, as it is done in Lando [26].

We use the results of Grasselli and Turd [19], summarized in Appendix C, considering a risk-neutral measure \mathbb{Q} , so that we can apply directly the results of Appendix C to valuation of bonds. Now, as done in Appendix C, we split the full filtration \mathcal{F}_t into two sub-filtrations:

$\mathcal{F}_t = \mathcal{G}_t \vee \mathcal{H}_t$ represents all the available information up to time t ;

¹See appendix C for details

$\mathcal{G}_t = \sigma \{Y_s : 0 \leq s \leq t\}$, the information of the evolution of a state vector Y_t up to time t , i.e, the filtration generated by the state vector that determines the default intensity;

$\mathcal{H}_t = \sigma \{1_{\{\tau_d \leq s\}} : 0 \leq s \leq t\}$, the information of the existence of default up to time t , i.e, the filtration generated by a Cox process $N_{d,t}$.

This partition is of crucial importance in the results of this chapter. Here the state vector determines the default probabilities but it cannot trigger the default event directly, as in structural models, and because the default process does not affect the intensity we cannot have the default of one firm affecting the intensity of another firm. Here we do not treat the case of correlated defaults and the results of this chapter are not straightforward generalized for that case. For an introduction of correlated intensity's you can see the book of David Lando [27].

Based on the theory of Poisson processes we evaluate a defaultable zero coupon bond with zero recovery in the reduced form models.

Proposition 2.1. *The value at time t of a defaultable bond whose only payment is $X1_{\{\tau_d > T\}}$ ($X \in \mathcal{G}_T$) at maturity is given by:*

$$V(t, T) = \mathbf{1}_{\{\tau_d > t\}} E_{\mathbb{Q}} \left[e^{-\int_t^T (r_s + \lambda_{d,s}^{\mathbb{Q}}) ds} X | \mathcal{F}_t \right] \quad (2.1)$$

Proof. With the definition of the \mathbb{Q} -measure, conditioning on $\mathcal{G}_T \vee \mathcal{H}_t$ and using the law of iterated expectations, we have that:

$$V(t, T) = E_{\mathbb{Q}} \left[e^{-\int_t^T r_s ds} X \mathbf{1}_{\{\tau_d > T\}} | \mathcal{F}_t \right] = E_{\mathbb{Q}} \left[E_{\mathbb{Q}}(e^{-\int_t^T r_s ds} X \mathbf{1}_{\{\tau_d > T\}} | \mathcal{G}_T \vee \mathcal{H}_t) | \mathcal{F}_t \right]$$

We have that $e^{-\int_t^T r_s ds}$ and X are \mathcal{G}_T -measurable and therefore, using Proposition B.4 of Appendix B:

$$V(t, T) = E_{\mathbb{Q}} \left[e^{-\int_t^T r_s ds} X E_{\mathbb{Q}}(\mathbf{1}_{\{\tau_d > T\}} | \mathcal{G}_T \vee \mathcal{H}_t) | \mathcal{F}_t \right] \quad (2.2)$$

For the second expectation we just need to use equation (C.3) from Appendix C to write:

$$\begin{aligned} E_{\mathbb{Q}}(\mathbf{1}_{\{\tau_d > T\}} | \mathcal{G}_T \vee \mathcal{H}_t) &= \mathbb{Q}(\tau_d > T | \mathcal{G}_T \vee \mathcal{H}_t) \\ &= e^{-\int_t^T \lambda_{d,u}^{\mathbb{Q}} du} \end{aligned} \quad (2.3)$$

Combining equations (2.2) and (2.3), equation (2.1) follows. \square

This is a really intuitive and practical way of considering the default probability. We now have an effective (adjusted) rate that is the sum of the short rate and the default intensity. The intensity works as a default spread and the expression for the effective rate can be modelled in the same way as the short-rate by using term structure models.

Now we can generalize Proposition 2.1 and consider a bond that can pay some non-zero value at time of default.

Proposition 2.2. *Consider a defaultable zero coupon bond that pays at maturity a value $X \mathbf{1}_{\{\tau_d > T\}}$ ($X \in \mathcal{G}_T$) and a stochastic Z_{τ_d} , \mathcal{G}_t -adapted, pay-off at the time of default ($Z_{\tau_d} = 0, t > T$). The value at time t is given by:*

$$V_t(t, T) = \mathbf{1}_{\{\tau_d > t\}} \left(E_{\mathbb{Q}} \left[e^{-\int_t^T (r_s + \lambda_{d,s}^{\mathbb{Q}}) ds} X | \mathcal{F}_t \right] + E_{\mathbb{Q}} \left[\int_t^T e^{-\int_t^u (r_s + \lambda_{d,s}^{\mathbb{Q}}) ds} Z_u \lambda_{d,u}^{\mathbb{Q}} du | \mathcal{F}_t \right] \right) \quad (2.4)$$

Proof. Again using the \mathbb{Q} -measure and conditional on $\mathcal{G}_T \vee \mathcal{H}_t$:

$$\begin{aligned} V(t, T) &= E_{\mathbb{Q}} \left[e^{-\int_t^T r_s ds} X \mathbf{1}_{\{\tau_d > T\}} + e^{-\int_t^{\tau_d} r_s ds} Z_{\tau_d} | \mathcal{F}_t \right] \\ &= E_{\mathbb{Q}} \left[E_{\mathbb{Q}}(e^{-\int_t^T r_s ds} X \mathbf{1}_{\{\tau_d > T\}} | \mathcal{G}_T \vee \mathcal{H}_t) | \mathcal{F}_t \right] + E_{\mathbb{Q}} \left[E_{\mathbb{Q}}(e^{-\int_t^{\tau_d} r_s ds} Z_{\tau_d} | \mathcal{G}_T \vee \mathcal{H}_t) | \mathcal{F}_t \right] \end{aligned}$$

The first term is proved in Proposition 2.1 and for the second we notice that the random pay-off is zero if default occurs after the maturity of the bond and it is also zero if default occurred before t . Hence,

$$E_{\mathbb{Q}}(e^{-\int_t^{\tau_d} r_s ds} Z_{\tau_d} | \mathcal{G}_T \vee \mathcal{H}_t) = \mathbf{1}_{\{\tau_d > t\}} \int_t^T e^{-\int_t^s r_u du} Z_s f(s) ds$$

where $f(s)$ is the probability density function of Z_{τ_d} .²

For the probability density function we use Bayes' theorem to get:

$$\mathbb{Q}(\tau_d < s | \tau_d > t | \mathcal{G}_T) = \frac{\mathbb{Q}(\tau_d < s \cap \tau_d > t | \mathcal{G}_T)}{\mathbb{Q}(\tau_d > t | \mathcal{G}_T)} = 1 - e^{-\int_t^s \lambda_{d,u}^{\mathbb{Q}} du}$$

and therefore,

²See Definition B.16 of Appendix B.

$$f(s) = \frac{d}{ds} \mathbb{Q}(\tau_d < s | \tau_d > t | \mathcal{G}_T) = \lambda_{d,s}^{\mathbb{Q}} e^{-\int_t^s \lambda_{d,u}^{\mathbb{Q}} du}$$

With this we obtain equation (2.4). □

Equation 2.4 is just the result of Proposition 2.1 (zero recovery) plus a recovery term. Nonetheless this is a more complex expression in which we have to solve an integral. To overcome this difficulty we use a recovery at default assumption that leads to a more simple and tractable expression.

The Recovery of Market Value Assumption

We have considered a random payoff at maturity and equation (2.4) is compatible with the recovery of face value (RFV) assumption for bonds with face value $FV = X$. Under the RFV assumption the creditor receives a constant fraction ω of the face value at the time of default, which leads to the last equation with $\omega = Z_{\tau_d}$. For convenience we start to work with bonds with face value $FV = 1$.

The RFV assumption underlying equation (2.4) is difficult to implement in practice and with this in mind Duffie and Singleton [16] consider the recovery of market value (RMV) assumption that results in an easier and more intuitive way of evaluating defaultable bonds. In this case the payoff X_d at default is a fraction δ_t of the market value just before default, where δ_t is a predictable stochastic process:

$$X_d = \delta_{\tau_d} \lim_{s \uparrow \tau_d} V(s, T) = \delta_{\tau_d} V(\tau_d^-, T)$$

Under this assumption we evaluate again the bonds considering an effective rate, just as in equation (2.4), as shown in the following proposition.

Proposition 2.3 (Generalized Lando Formula). *A defaultable zero coupon bond that pays $FV = 1$ at maturity and a stochastic fraction of the market value $\delta_t V_t$ at default, where δ_t is exogenous (does not depend on V_t), has the following value at time t :*

$$V(t, T, 0, \delta_t) = E_{\mathbb{Q}} \left[e^{-\int_t^T (r_s + \lambda_{d,s}^{\mathbb{Q}} (1 - \delta_s)) ds} | \mathcal{F}_t \right] \quad (2.5)$$

Proof. From equation (2.4) the value of this bond is:

$$V(t, T, 0, \delta_t) = \mathbf{1}_{\{\tau_d > t\}} \left(E \left[e^{-\int_t^T (r_s + \lambda_{d,s}^{\mathbb{Q}}) ds} | \mathcal{F}_t \right] + E_{\mathbb{Q}} \left[\int_t^T e^{-\int_t^u (r_s + \lambda_{d,s}^{\mathbb{Q}}) ds} \delta_u V_u \lambda_{d,u}^{\mathbb{Q}} du | \mathcal{F}_t \right] \right)$$

In $(t, T, 0, \delta_t)$, the zero is saying that the bond have coupons of value zero.

Introducing

$$M_t = E_{\mathbb{Q}} \left[e^{-\int_0^T (r_s + \lambda_{d,s}^{\mathbb{Q}}) ds} + \int_0^T e^{-\int_0^u (r_s + \lambda_{d,s}^{\mathbb{Q}}) ds} \delta_u V_u \lambda_{d,u}^{\mathbb{Q}} du \middle| \mathcal{F}_t \right]$$

it is possible to write:

$$V(t, T, 0, \delta_t) = e^{-\int_0^t (r_s + \lambda_{d,s}^{\mathbb{Q}}) ds} \left(M_t - \int_0^t e^{-\int_0^u (r_s + \lambda_{d,s}^{\mathbb{Q}}) ds} \delta_u V_u \lambda_{d,u}^{\mathbb{Q}} du \right)$$

for $0 < s < t$,

$$\begin{aligned} E_{\mathbb{Q}} [M_t | \mathcal{F}_s] &= E_{\mathbb{Q}} \left[E_{\mathbb{Q}} \left[e^{-\int_0^T (r_s + \lambda_{d,s}^{\mathbb{Q}}) ds} + \int_0^T e^{-\int_0^u (r_s + \lambda_{d,s}^{\mathbb{Q}}) ds} \delta_u V_u \lambda_{d,u}^{\mathbb{Q}} du \middle| \mathcal{F}_t \right] \middle| \mathcal{F}_s \right] \\ &= E_{\mathbb{Q}} \left[e^{-\int_0^T (r_s + \lambda_{d,s}^{\mathbb{Q}}) ds} + \int_0^T e^{-\int_0^u (r_s + \lambda_{d,s}^{\mathbb{Q}}) ds} \delta_u V_u \lambda_{d,u}^{\mathbb{Q}} du \middle| \mathcal{F}_s \right] = M_s \end{aligned}$$

therefore, M_t is a martingale.

Defining

$$f(t, M_t) \equiv e^{-\int_0^t (r_s + \lambda_{d,s}^{\mathbb{Q}}(1-\delta_s)) ds} V_t = e^{\int_0^t \delta_s \lambda_{d,s}^{\mathbb{Q}} ds} \left(M_t - \int_0^t e^{-\int_0^u (r_s + \lambda_{d,s}^{\mathbb{Q}}) ds} \delta_u V_u \lambda_{d,u}^{\mathbb{Q}} du \right)$$

we can use the Itô Lemma to obtain:

$$d(e^{-\int_0^t (r_s + \lambda_{d,s}^{\mathbb{Q}}(1-\delta_s)) ds} V_t) = e^{\int_0^t \delta_s \lambda_{d,s}^{\mathbb{Q}} ds} dM_t$$

Integrating in the interval $[t, T]$ and conditioning on \mathcal{F}_t :

$$E_{\mathbb{Q}} [e^{-\int_0^T (r_s + \lambda_{d,s}^{\mathbb{Q}}(1-\delta_s)) ds} V_T - e^{-\int_0^t (r_s + \lambda_{d,s}^{\mathbb{Q}}(1-\delta_s)) ds} V_t | \mathcal{F}_t] = E_{\mathbb{Q}} [e^{\int_0^t \delta_s \lambda_{d,s}^{\mathbb{Q}} ds} (M_T - M_t) | \mathcal{F}_t]$$

Since the second term in the first expectation is \mathcal{F}_t -measurable, $V_T = 1$ and M_t is a martingale:

$$E_{\mathbb{Q}} [e^{-\int_0^T (r_s + \lambda_{d,s}^{\mathbb{Q}}(1-\delta_s)) ds} V_T | \mathcal{F}_t] - e^{-\int_0^t (r_s + \lambda_{d,s}^{\mathbb{Q}}(1-\delta_s)) ds} V_t = 0$$

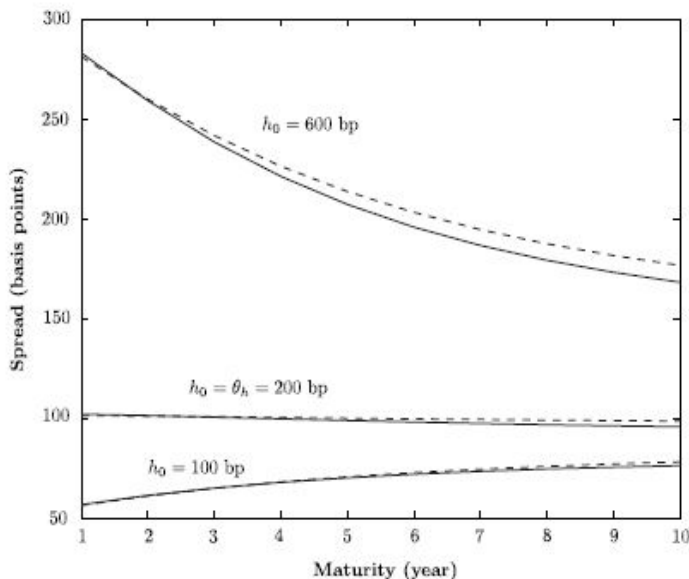


FIGURE 2.1: This figure from Duffie and Singleton shows the difference in the term structure of par-coupon yield spreads for RMV (dashed lines) and RFV (solid)

we get (2.5).

□

According to Duffie and Singleton [16] the recovery of face value assumption has a legal substantiation. For the case of coupon bonds, again with a legal justification, they consider that coupons have zero recovery as well. Despite not having this legal support, and only the advantage of mathematical tractability, the recovery of market value assumption does well comparing to RFV. They find, with a constant δ (recovery of face and market value) and using CIR processes for the state variables that explain the effective rate, small spread differences between the RFV and RMV approaches, at least for coupon bonds trading near par, as shown in Figure 2.2. Based on these results we consider justified the use of RMV in the rest of this work.

With this assumption we can once again model the default spread (now given by $(1 - \delta_s)\lambda_{d,t}^{\text{Q}}$) in the same way as we model the short-rate in term structure models. But in this spread the intensity and the recovery rate appear together and, therefore, we will model them together.

Equation 2.5 is easily generalized for the case of bonds paying coupons of value c at n dates T_i .

$$\begin{aligned}
V(t, T, c, \delta_t) &= E_{\mathbb{Q}} \left[\sum_{t < T_i \leq T} ce^{-\int_t^{T_i} R_{eff,d}^{\mathbb{Q}} du} + e^{-\int_t^T R_{eff,d}^{\mathbb{Q}} du} \right] \\
&= \sum_{t < T_i \leq T} cV(t, T_i, 0, \delta_t) + V(t, T, 0, \delta_t)
\end{aligned} \tag{2.6}$$

Where the effective rate is given by:

$$R_{eff,d}^{\mathbb{Q}} = r_t + \lambda_{d,t}^{\mathbb{Q}}(1 - \delta_t) \tag{2.7}$$

Summarizing, we have now the pricing formulas for both zero-coupon and coupon bonds in the reduced form models. If the bond pays nothing at the time of default we have an effective rate constituted by the short-rate plus a default spread given by the intensity of the Cox process. From interest rate theory we know how to construct stochastic models for this effective rate that gives us a closed form solution for the discount factor. In the case of a recovery term at default, it was found that considering RFV — an assumption supported on legal terms — the solution does not permit the same usage of the models for the short rate. To have again an effective rate that can be modelled with interest rate theory we had to consider RMV, an assumption that was found by Duffie and Singleton [16] to be empirically satisfactory.

2.2 Call Provision

A special case of bonds are those that have call provisions. These type of contracts represent a large share of the US bond market. These provisions allow the issuer to redeem the bond prior to maturity at a fixed cost X_c in the case of callable bonds and at a random price $X_{mw,t}$ in the case of bonds with make-whole call provisions.

With call provisions the issuer has more flexibility managing his debt, in particular he can change the leverage of the firm at any time. Moreover, these bonds also involve less interest rate risk. If interest rates go down the firm can redeem the bond and issue another at lower interest rates. Because of these facts the issuer is willing to pay a higher coupon rate for a callable bond. In the perspective of the buyer, he now does not have an issuer stuck for a deterministic time with a contract. Now is him that have more interest rate risk, because if rates go down the issuer redeems the bond and to apply the correspondent capital for the same return he will have to face a higher risk. Therefore, it is expected a higher return for callable bonds when compared to non-callable ones.

In the spirit of the last section we generalize the valuation of defaultable bonds to callable bonds by considering the call time τ_c as the jump of a Cox process $N_{c,t} = 1_{\{t \geq \tau_c\}}$ with exogenous intensity $\lambda_{c,t}^{\mathbb{Q}}$. This way the value of a defaultable callable bond with random recovery at maturity is obtained in a straightforward way from Proposition 2.2.

Proposition 2.4. *A defaultable zero coupon callable bond that pays $X \mathbf{1}_{\{\tau_d > T\}}$ ($X \in \mathcal{G}_T$) at maturity T , a stochastic payoff $Z_{\tau_d} \mathbf{1}_{\{\tau_c, T > \tau_d\}}$, \mathcal{G}_t -adapted, at the time of default, and a fixed value $X_c \mathbf{1}_{\{T, \tau_d > \tau_c\}}$ at the call time, has the following value at time t :*

$$V(t, T) = E_{\mathbb{Q}} \left[e^{-\int_t^T (r_s + \lambda_{d,s}^{\mathbb{Q}}) ds} X | \mathcal{F}_t \right] + E_{\mathbb{Q}} \left[\int_t^T e^{-\int_t^u (r_s + \lambda_{d,s}^{\mathbb{Q}}) ds} Z_{u_d} \lambda_{d,u}^{\mathbb{Q}} du | \mathcal{F}_t \right] \\ + E_{\mathbb{Q}} \left[\int_t^T e^{-\int_t^u (r_s + \lambda_{c,s}^{\mathbb{Q}}) ds} X_c \lambda_{c,u}^{\mathbb{Q}} du | \mathcal{F}_t \right]$$

Once again, we can not use an effective rate and we have to consider the recovery of market value assumption where the bondholder receives at τ_c a fraction k_t of the market value. This is, of course, a crude simplification. In standard callable bonds the value of recovery at the call time is a known, pre-determined value. We could easily modify Proposition 2.4 in such a way that we would have a recovery of market value at default and maintain the fixed value at the call time, but we would still have a more complex formula than with the RMV assumption. The expression obtained with RMV will make it easier to calibrate the model. In this case we have a generalization of Proposition (2.3), and therefore an effective rate.

Proposition 2.5. *A defaultable callable zero coupon bond that pays $FV = 1$ at maturity, a stochastic fraction of the market value $\delta_t V_t$ at default and a stochastic fraction $k_t V_t$ at call time, where δ_t and k_t are exogenous (not dependent on V_t), has the following value at time t :*

$$V(t, T) = E_{\mathbb{Q}} \left[e^{-\int_t^T (r_s + \lambda_{d,s}^{\mathbb{Q}}(1 - \delta_s) + \lambda_{c,s}^{\mathbb{Q}}(1 - k_s)) ds} | \mathcal{F}_t \right] \quad (2.8)$$

And we write the effective rate as:

$$R_{eff,c}^{\mathbb{Q}} \equiv r_t + \lambda_{d,t}^{\mathbb{Q}}(1 - \delta_t) + \lambda_{c,t}^{\mathbb{Q}}(1 - k_t) \quad (2.9)$$

For the case of bonds with make-whole call provisions the value that we receive at the call time is stochastic and therefore the RMV is less problematic than in the case of standard callable bonds. Now we consider the make-whole call time τ_{cmw} as the first jump of a Cox process $N_{cmw,t} = 1_{\{t \geq \tau_{cmw}\}}$ with exogenous intensity $\lambda_{cmw,t}^{\mathbb{Q}}$. Under the

RMV assumption and with a recovery fraction of \tilde{h}_t , Proposition 2.5 still holds but now the effective rate reads:

$$R_{eff,cmw}^{\mathbb{Q}} = r_t + \lambda_{d,t}^{\mathbb{Q}}(1 - \delta_t) + \lambda_{cmw,t}^{\mathbb{Q}}(1 - \tilde{h}_t) \quad (2.10)$$

Some bonds have both call and make-whole call provisions. Usually the call provisions do not start at the emission of the bond, but just after some time. In practice the bond is MWC up to that point, and then is considered a standard callable, because the issuer can redeem it at the par value that is usually given as the fixed cost of calling the bond. The MWC price would always be equal or greater than the par value.

Having the pricing formulas written in terms of effective rates that resemble what we have for risk-free bonds, it is now time to specify the stochastic processes that these effective rates can follow.

3

Model Specification

The results of Chapter 2, especially those that consider the recovery of market value assumption, make possible a direct comparison between the resulting pricing formulas and the pricing formula for the risk-free bonds. This is so because we write the formulas of defaultable non-callable and callable bonds through an effective rate. This rate contains the stochastic short-rate plus a stochastic spread and we will describe in this chapter the specific models that we use to model these separate processes. The theory of affine term structure models (reviewed in Appendix D) is used because closed-form solutions are known that make easier the estimation process.

3.1 Non-defaultable Bonds

According to Litterman and Scheinkman [28], studying the returns on fixed income securities related to US government bonds, three risk factors explain most of the variation in those returns, and they call these factors level, steepness and curvature, because of the way they affect the term structure, even though, as they say for the steepness: *"it does not correspond exactly to any of the steepness measures commonly used"*. The first factor explains roughly 88% of the variance, while the steepness adds 8% and the curvature 2%. We use a two factor model for the dynamics of the short rate because a two factor model still explains a high percentage of the variance of the returns and will have less parameters to estimate, which will make the estimation procedure faster. In particular we use an affine model with the market price of risk of Dai and Singleton [6] (thus a completely affine model — see Appendix D). These models have known closed-form solutions and exhibit mean reversion phenomena, as required by historical evidence. Empirical data also demands stochastic volatility, and therefore we are restricted to the

$\mathcal{A}_m(N)$ models with $m > 0$ — these consists of an N factor model where m of them contribute to the stochastic volatility.

One of the simplest models that we can consider with these restrictions is the multi-factor CIR model. It is particularly suitable when it is necessary to use the Kimmel[24] series expansion, a method to find approximated closed-form solutions for *particular* non-linear models. This approximation can be used when we have a non-linear model based of a CIR process (see Appendix E). Nonetheless, it is not the most flexible model even when conditional on using the Kimmel series expansion, but is the model used by Jarrow et al.[20] and Park and Clark[34] that are the base for this work, since they contain the model that yields to all types of contracts a closed-form solution that will make the use of the Kalman filter (described in Appendix F) to estimate the model parameters much faster.

With the above considerations we write for the short-rate the following two-factor process:

$$r_t = \alpha_0 + Y_1 + Y_2 \quad (3.1)$$

As said, it is considered a multi-factor CIR model and therefore for each variable we have under the physical measure \mathbb{P} the following dynamics:

$$dY_i = k_{ii}(\theta_i - Y_i)dt + \sigma_i\sqrt{Y_i}dW_i^{\mathbb{P}} \quad (3.2)$$

Where $\theta_i, k_{ii}, \sigma_i \in \mathbb{R}$.

This is consistent with the mean-reverting phenomena of interest rates. This is so because θ_i is a fixed point of the deterministic part of the dynamical system in equation (3.2).¹ Moreover, this fixed point is stable for positive² k_{ii} — a condition imposed by the definition D.3 of Appendix D for this $\mathcal{A}_2(2)$ model. Of course we do not have the system always at the fixed point because we have a random force given by the second term in equation (3.2). This random force depends on the state variable and therefore we have a level dependent volatility, as required from observations [18].

With the market price of risk $\Lambda_t = \sqrt{Y_i\bar{\eta}}$ — corresponding to completely affine models (see Section D.1) — we maintain an affine model and we know a closed-form solution

¹Consider the system $\frac{dx_1}{dt} = f(x_1)$. The point x_1^* is a fixed point if it satisfies $f(x_1^*) = 0$, i.e, it has zero velocity.

²Consider an infinitesimal displacement to the fixed point: $x_1 = x_1^* + \delta x_1$. If $\frac{df}{dx}(\delta x_1) < 0$ the perturbation decays and the fixed point is said to be stable.

for the pricing problems described in Chapter 3. The dynamics of the state variables under the \mathbb{Q} -measure become:

$$dY_i = (k_{ii}\theta_i - (k_{ii} + \eta_i)Y_i)dt + \sigma_i\sqrt{Y_i}dW_i^{\mathbb{Q}} = \dot{k}_{ii}(\dot{\theta}_i - Y_i)dt + \sigma_i\sqrt{Y_i}dW_i^{\mathbb{Q}} \quad (3.3)$$

Where we define $\dot{k}_{ii} = k_{ii} + \eta_i$ and $\dot{\theta}_i = \frac{k_{ii}\theta_i}{k_{ii} + \eta_i}$ so that the dynamics of the state variables are written as square-root processes, making it easier to use known results for these processes.

The constant term α_0 maintains the possibility of non-positive rates even if the latent state variables follow CIR processes with the Feller conditions imposed ($2k_{ii}\theta_i > \sigma_i^2$). These conditions are only needed in the extended class of affine models, and not on the essentially affine ones, but is a usual choice to consider them. In the spirit of Litterman and Scheinkman [29] we can call to Y_1 and Y_2 the level and steepness, respectively. I choose to impose the Feller condition on the first factor because with α_0 we already have the possibility of zero rates. We do not impose the Feller condition on the slope because if this would negate the possibility of a inverted term structure, and we know that this is possible when the market expects the economy to slow down.

With all these considerations we can price a risk-free bond.

Proposition 3.1. *The value at time t of the risk-free zero coupon bond with maturity at T in this model is given by:*

$$V(t, T) = e^{-\alpha_0\tau + \psi_0(\tau) - \psi_1(\tau)Y_{1,t} - \psi_2(\tau)Y_{2,t}} \quad (3.4)$$

where

$$\psi_0(\tau) = \psi_{0,1}(\tau) + \psi_{0,2}(\tau) \quad \psi_{0,i}(\tau) = \frac{2k_{ii}\theta_i}{\sigma_i^2} \ln \left[\frac{2h_i e^{\frac{1}{2}(k_{ii} + \eta_i + h_i)\tau}}{h_i - (k_{ii} + \eta_i) + (k_{ii} + \eta_i + h_i)e^{h_i\tau}} \right]$$

$$\psi_i(\tau) = \frac{2(e^{h_i\tau} - 1)}{h_i - (k_{ii} + \eta_i) + ((k_{ii} + \eta_i + h_i)e^{h_i\tau}} \quad h_i = \sqrt{(k_{ii} + \eta_i)^2 + 2\sigma_i^2}$$

Proof. According to Proposition 6.2.5 of Lamberton and Lapeyre [25] we know that if a state variable Y follows the process (3.3) it satisfies:

$$\mathbb{E}_{\mathbb{Q}} \left[e^{-\int_{t_0}^t r_s ds} \mid \mathcal{F}_t \right] = e^{\phi_{\lambda, \mu}(t-t_0) - r_{t_0} \psi_{\lambda, \mu}(t-t_0)}$$

where:

$$\phi_{\lambda,\mu}(t-t_0) = \frac{2\dot{k}_{ii}\dot{\theta}_i}{\sigma_i^2} \ln \left[\frac{2h_i e^{\frac{1}{2}(\dot{k}_{ii}+h_i)(t-t_0)}}{\sigma_i^2 \lambda (e^{h_i(t-t_0)} - 1) + h_i - \dot{k}_{ii} + (\dot{k}_{ii} + h_i)e^{h_i(t-t_0)}} \right]$$

$$\psi_{\lambda,\mu}(t-t_0) = \frac{\lambda \left[h_i + \dot{k}_{ii} + (h_i - \dot{k}_{ii}) e^{h_i(t-t_0)} \right] + 2\mu(e^{h_i(t-t_0)} - 1)}{\sigma_i^2 \lambda (e^{h_i(t-t_0)} - 1) + h_i - \dot{k}_{ii} + (\dot{k}_{ii} + h_i)e^{h_i(t-t_0)}}$$

Remembering that we have *independent* Brownian motions, then:

$$V(t, T) = e^{-\alpha_0 \tau} \mathbb{E}_{\mathbb{Q}} \left[e^{-\int_t^T Y_{1,s} ds} \mid \mathcal{F}_t \right] \mathbb{E}_{\mathbb{Q}} \left[e^{-\int_t^T Y_{2,s} ds} \mid \mathcal{F}_t \right]$$

The result follows from applying Proposition 6.2.5 of Lamperton and Lapeyre[25].

□

3.2 Defaultable Non-Callable Bonds

It is now time to model the case of non-callable bonds but now with credit risk. In the recovery of market value framework this corresponds to model the effective rate given by equation (2.5). The first term is the short-rate that we modelled in the last section. The second term is called the default spread and is given by a combination of the recovery fraction δ_t and the default intensity $\lambda_{d,t}^{\mathbb{Q}}$. We consider them jointly using a \mathcal{A}_3 (3) model. Equation (2.5) gives the default intensity in the pricing measure, but we start again by considering the intensity in the physical measure and with the same market price of risk from the last section we get the process under the \mathbb{Q} -measure.

We now consider a state variable Y_3 representing the financial health of the corporation. We assume that small values of this variable correspond to a good financial wealth. The variable is random but we still assume mean-reversion. This assumption can be justified as a result of firms trying to maintain constant leverage ratios. The assumption of stochastic volatility is also maintained. We also make the assumption that the risk free interest rate affects the default process, a empirical result due to Dufee [10]. He finds that yield spreads fall when the level of the term structure of treasury rates rises. The relation is bond dependent, in the sense that the correlation depends on the initial rating of the bonds. The correlation declines with increasing credit quality. He finds also a negative correlation between spreads and the steepness of the treasury term structure, but much weaker.

Based on this, we consider for the default spread:

$$(1 - \delta_t)\lambda_{d,t}^{\mathbb{P}} = \gamma_0 + \beta_{d_1}Y_1 + \beta_{d_2}Y_2 + Y_3 \quad (3.5)$$

Where equations (3.2) and (3.3) still apply.

For Y_3 we impose the Feller condition, meaning that the financial health of the firm contributes always something for the default spread. A higher value of Y_3 that implies a larger value of the default spread will also mean higher volatility in the dynamics of the state variable, due to stochastic volatility. We let all the β 's be either positive or negative, and the same happens to γ_0 . The problem is that in this way we make possible the existence of negative default spreads! Even with a positive γ_0 , and with negative β 's, for a high enough short rate we will have negative spreads. This will mean either negative default intensity or fraction of recovery higher than one (possibility that default are happy days for bondholders). The need for this possibility is a problem for reduced form models. One that in the words of Dufee [11]: *(...) is largely ignored if the model accurately prices the relevant instruments.*

Using equations (3.1) and (3.5) for the effective rate in (2.5) we get the result:

Proposition 3.2. *The value at time t of the defaultable zero coupon bond with maturity at T in this model is given by:*

$$V(t, T, 0, \delta_t) = e^{-(\alpha_0 + \gamma_0)\tau + \psi_0(\tau) - \psi_1(\tau)Y_{1,t} - \psi_2(\tau)Y_{2,t} - \psi_3(\tau)Y_{3,t}} \quad (3.6)$$

where

$$\psi_0(\tau) = \psi_{0,1}(\tau) + \psi_{0,2}(\tau) + \psi_{0,3}(\tau) \quad \psi_{0,i}(\tau) = \frac{2k_{ii}\theta_i}{\sigma_i^2} \ln \left[\frac{2h_i e^{\frac{1}{2}(k_{ii} + \eta_i + h_i)\tau}}{h_i - (k_{ii} + \eta_i) + (k_{ii} + \eta_i + h_i)e^{h_i\tau}} \right]$$

$$\psi_i(\tau) = \frac{2\mu_i(e^{h_i\tau} - 1)}{h_i - (k_{ii} + \eta_i) + ((k_{ii} + \eta_i + h_i)e^{h_i\tau}} \quad h_i = \sqrt{(k_{ii} + \eta_i)^2 + 2\mu_i\sigma_i^2}$$

$$\mu_i = \begin{cases} 1 + \beta_{d_i}, & i = 1, 2 \\ 1, & i = 3 \end{cases}$$

Proof. Using equation (2.5) and we have for the effective rate:

$$R_{eff,d}^{\mathbb{Q}} = r_t + \lambda_{d,t}^{\mathbb{Q}}(1 - \delta_t) = (\alpha_0 + \gamma_0) + (1 + \beta_{d_1})Y_1 + (1 + \beta_{d_2})Y_2 + Y_3$$

what leads to:

$$V(t, T) = e^{-(\alpha_0 + \gamma_0)\tau} \mathbb{E}_{\mathbb{Q}} \left[e^{-(1 + \beta_{d_1}) \int_t^T Y_{1,s} ds} \mid \mathcal{F}_t \right] \mathbb{E}_{\mathbb{Q}} \left[e^{-(1 + \beta_{d_2}) \int_t^T Y_{2,s} ds} \mid \mathcal{F}_t \right] \mathbb{E}_{\mathbb{Q}} \left[e^{-\int_t^T Y_{3,s} ds} \mid \mathcal{F}_t \right]$$

Then we use again the Lamperton and Lapeyre [25] result. \square

3.3 Defaultable Bonds With Call Provisions

Now we consider defaultable bonds but with call provisions. With the default spread we need to be more careful. When the issuer decides to call a bond it may not decide to call all of the bonds. So this part of the effective rate also needs to be bond dependent. It is also of crucial importance to understand how the call spread and the make-whole call spread differ, if they differ at all.

To understand the functional form of the dependence we consider a bond that pays a cash-flow CF_i at time T_i and it is emitted at par $FV = 1$. For simplicity we work for now with deterministic interest rates. The cash-flows are discounted at a rate that is the sum of the treasury yield i_t plus a credit spread q_t . Time to maturity is represented by $\tau = T - t$.

The issuer has a reason to redeem the bond if the market price is higher than the buy back price X_c :

$$X_c < V_t$$

$$X_c < \sum_{t < T_i \leq T} CF_i e^{-(i_t, T_i + q_t)(T_i - t)} \sim \sum_{t < T_i \leq T} \frac{CF_i}{1 + (i_t, T_i + q_t)(T_i - t)}$$

With this expression we conclude that calling a bond depends directly on the value of the cash-flows, and therefore on the periodic coupons c , and inversely on the treasury yields and credit spread.

For a bond with make-whole call provisions we first notice that the emission at par corresponds to:

$$1 = \sum_{0 < T_i \leq T} CF_i e^{-(i_0, T_i + q_0)T_i}$$

To call the bond, the following inequality must be verified:

$$X_{cmw,t} < V_t$$

i.e.,

$$\max \left\{ \sum_{t < T_i \leq T} CF_i e^{-(i_0, T_i + q_0) T_i}, \sum_{t < T_i \leq T} CF_i e^{-(i_t, T_i + M_{mw})(T_i - t)} \right\} < \sum_{t < T_i \leq T} CF_i e^{-(i_t, T_i + q_t)(T_i - t)}$$

The make-whole premium M_{mw} needs to satisfy $M_{mw} < q_0$, otherwise the rational to call the bond would be satisfied at emission ($t = 0$).

In the case that the maximum value is the second term:

$$\sum_{n=1}^{\tau} CF_n e^{-(i(t, t+n) + M_{mw})n} < \sum_{n=1}^{\tau} CF_n e^{-(i(t, t+n) + q_t)n} \quad (3.7)$$

For this to happen:

$$q_t < M_{mw} \Rightarrow q_t < q_0$$

Now the coupons are not important. The same happens to the treasury yields. Large values of M_{mw} make it easier to call the bond and the opposite happens with the credit spread q_t .

When the first term is larger we still have equation (3.7) satisfied.

Summarizing, we have for callable bonds and bonds with make-whole call provisions:

- Callable bonds depend directly on the coupon and inversely on the treasury yields and credit spread;
- Bonds with make-whole call provision do not depend on the coupon or treasury yields but have a direct relation with the make-whole premium M_{mw} and inversely with the default spread.

So we must consider callable bonds and make-whole call bonds separately because the respective call processes have different dependences on the several variables involved.

3.3.1 Defaultable Callable Bonds

For the standard callable bonds we need to model the effective rate (2.7). The short rate is given by equation (3.1) and the default spread by (3.5). The call spread remains to be modelled. There is the need to introduce a new variable Y_4 that represents the component of the call spread that includes the non-optimal call policies and market frictions that are observed in practice, that we assume to be stochastic just as the default intensity when considering the financial health of the firm. From the reasoning in the last section we need to add a term that is inversely correlated to the level of the term structure and the default spread, and also directly related to the coupon of the bond. This means that we will have a non-linear expression for the state variables. The results of Appendix D do not apply here. Nonetheless, we know how to find closed form solutions to specific non-linear problems. In particular problems that have an expectation of the form:

$$\mathbb{E} \left[e^{-\int (aY_l + \frac{b}{Y_l}) dt} \right]$$

For a variable Y_l .

The results for this particular problem can be found in example E.1 of Appendix E, as long as we consider a CIR dynamics for the variable Y_l . Moreover, the results of that example only apply if we impose the Feller condition on the dynamics of the variable Y_l , meaning that non-optimal call policies and/or market friction always contribute to the call spread. With these conditions, a closed form approximation in power series of τ exists and is uniformly convergent, i.e, the convergence does not depend of the maturity of the bond. With this in mind we consider the following model for the call spread:

$$(1 - k_t)\lambda_{c,t}^{\mathbb{P}} = \zeta_0 + Y_4 + \zeta_1 \frac{c}{Y_1} + \zeta_3 \frac{c}{Y_3} \quad (3.8)$$

This expression allows the use of the Kimmel series expansion to get a closed-form solutions as we see now.

Proposition 3.3. *The value at time t of the defaultable callable zero coupon bond with maturity at T in this model is given by:*

$$V(t, T, 0, \delta_t, k_t) = e^{-(\alpha_0 + \gamma_0 + \zeta_0)\tau + \psi_0(\tau) - \psi_2(\tau)Y_{2,t} - \psi_4(\tau)Y_{4,t}} \pi(Y_1, t, T) \pi(Y_3, t, T) \quad (3.9)$$

where

$$\psi_0(\tau) = \psi_{0,2}(\tau) + \psi_{0,4}(\tau) \quad \psi_{0,i}(\tau) = \frac{2k_{ii}\theta_i}{\sigma_i^2} \ln \left[\frac{2h_i e^{\frac{1}{2}(k_{ii} + \eta_i + h_i)\tau}}{h_i - (k_{ii} + \eta_i) + (k_{ii} + \eta_i + h_i)e^{h_i\tau}} \right]$$

$$\psi_i(\tau) = \frac{2\mu_i(e^{h_i\tau} - 1)}{h_i - (k_{ii} + \eta_i) + ((k_{ii} + \eta_i + h_i)e^{h_i\tau}} \quad h_i = \sqrt{(k_{ii} + \eta_i)^2 + 2\mu_i\sigma_i^2}$$

$$\mu_i = \begin{cases} 1 + \beta_{d_i}, & i = 2 \\ 1, & i = 4 \end{cases}$$

$$\pi(Y_1, t, T) = \mathbb{E}_{\mathbb{Q}} \left[e^{-\int_t^T (1 + \beta_{d_1}) Y_{1,s} + \zeta_1 \frac{c}{Y_1} ds} \mid \mathcal{F}_t \right] \quad \pi(Y_3, t, T) = \mathbb{E}_{\mathbb{Q}} \left[e^{-\int_t^T Y_{3,s} + \zeta_3 \frac{c}{Y_{3,s}} ds} \mid \mathcal{F}_t \right]$$

Proof. using equation (2.5), we have for the effective rate:

$$\begin{aligned} R_{eff,c}^{\mathbb{Q}} &= r_t + \lambda_{d,t}^{\mathbb{Q}}(1 - \delta_t) + \lambda_{c,t}^{\mathbb{Q}}(1 - k_t) \\ &= (\alpha_0 + \gamma_0 + \zeta_0) + (1 + \beta_{d_2}) Y_2 + Y_4 + \left(Y_3 + \zeta_3 \frac{c}{Y_4} \right) + \left((1 + \beta_{d_1}) Y_1 + \zeta_1 \frac{c}{Y_1} \right) \end{aligned}$$

That leads to:

$$\begin{aligned} V(t, T) &= e^{-(\alpha_0 + \gamma_0 + \zeta_0)\tau} \mathbb{E}_{\mathbb{Q}} \left[e^{-(1 + \beta_{d_2}) \int_t^T Y_{2,s} ds} \mid \mathcal{F}_t \right] \mathbb{E}_{\mathbb{Q}} \left[e^{-\int_t^T Y_{4,s} ds} \mid \mathcal{F}_t \right] \\ &\quad \mathbb{E}_{\mathbb{Q}} \left[e^{-\int_t^T (1 + \beta_{d_1}) Y_{1,s} + \zeta_1 \frac{c}{Y_1} ds} \mid \mathcal{F}_t \right] \mathbb{E}_{\mathbb{Q}} \left[e^{-\int_t^T Y_{3,s} + \zeta_3 \frac{c}{Y_{3,s}} ds} \mid \mathcal{F}_t \right] \end{aligned}$$

Once again we use Lamperton and Lapeyre[25]. For the non-linear part, it is solved in example E.1 of Appendix E with $c_1 = 1 + \beta_{d_1}$ and $c_2 = \zeta_1 c$ for $\pi(Y_1, t, T)$; $c_1 = 1$ and $c_2 = \zeta_3 c$ for $\pi(Y_3, t, T)$. \square

The term that includes the non-linear relation in the credit spread is not used in Jarrow et al.[20], but as we have seen the rationale to include such a term exists. If this more complex model is more accurate than that of Jarrow et al.[20] is a question that we will try to answer later.

3.3.2 Defaultable Bonds With Make-Whole Call Provision

The result for bonds with make-whole call provisions is similar. We just need to model the make-whole spread of equation (2.10). The variable Y_5 is introduced and represents

the possibility of non-optimal make-whole call policies and market frictions. Again a CIR dynamics is chosen and the Feller condition is imposed. There is also the need to add one term that depends directly on the make-whole premium and inversely on the credit spread:

$$(1 - \bar{h}_t)\lambda_{cmw,t}^{\mathbb{P}} = \chi_0 + Y_5 + \chi_3 \frac{M_{mw}}{Y_3} \quad (3.10)$$

From this equation we get the formula to evaluate bonds with credit risk and make-whole call provisions.

Proposition 3.4. *The value at time t of the defaultable make-whole callable zero coupon bond with maturity at T in this model is given by:*

$$V(t, T, 0, \delta_t, \bar{h}_t) = e^{-(\alpha_0 + \gamma_0 + \chi_0)\tau + \psi_0(\tau) - \psi_1(\tau)Y_{1,t} - \psi_2(\tau)Y_{2,t} - \psi_5(\tau)Y_{5,t}} \pi(Y_3, t, T) \quad (3.11)$$

where

$$\psi_0(\tau) = \psi_{0,1}(\tau) + \psi_{0,2}(\tau) + \psi_{0,5}(\tau) \quad \psi_{0,i}(\tau) = \frac{2k_{ii}\theta_i}{\sigma_i^2} \ln \left[\frac{2h_i e^{\frac{1}{2}(k_{ii} + \eta_i + h_i)\tau}}{h_i - (k_{ii} + \eta_i) + (k_{ii} + \eta_i + h_i)e^{h_i\tau}} \right]$$

$$\psi_i(\tau) = \frac{2\mu_i(e^{h_i\tau} - 1)}{h_i - (k_{ii} + \eta_i) + ((k_{ii} + \eta_i + h_i)e^{h_i\tau}} \quad h_i = \sqrt{(k_{ii} + \eta_i)^2 + 2\mu_i\sigma_i^2}$$

$$\mu_i = \begin{cases} 1 + \beta_{d_i}, & i = 1, 2 \\ 1, & i = 5 \end{cases} \quad \pi(Y_3, t, T) = \mathbb{E}_{\mathbb{Q}} \left[e^{-\int_t^T Y_{3,s} + \chi_3 \frac{M_{mw}}{Y_{3,s}} ds} \middle| \mathcal{F}_t \right]$$

Proof. Now we use equation (2.8) and we have for the effective rate:

$$\begin{aligned} R_{eff,cmw}^{\mathbb{Q}} &= r_t + \lambda_{d,t}^{\mathbb{Q}}(1 - \delta_t) + \lambda_{cmw,t}^{\mathbb{Q}}(1 - \bar{h}_t) \\ &= (\alpha_0 + \gamma_0 + \chi_0) + (1 + \beta_{d_1})Y_1 + (1 + \beta_{d_2})Y_2 + Y_5 + \left(Y_3 + \chi_3 \frac{M_{mw}}{Y_3} \right) \end{aligned}$$

We proceed in the same way as in Proposition 2.3, but now $c_1 = 1$ and $c_2 = \chi_3 M_{mw}$ for $\pi(Y_3, t, T)$. \square

Having all the closed form solutions to evaluate non-callable and callable bonds with credit risk, it is time to see how well this models fit the data.

4

Results

In this chapter we evaluate how well the reduced form models developed in Chapters 2 and 3 perform. Given that all the bonds, be it treasury bonds or the different types of corporate bonds, depend of the short-rate process, i.e, on the state variables describe as the level Y_1 and the steepness Y_2 , we first estimate these variables using treasury rates and considering the linear Kalman filter described in Appendix F. The Kalman filter estimate the state vector at each observational time and the time homogeneous parameters. Then, these results are used as inputs for the extended Kalman filter estimation of the processes for non-callable and callable corporate bonds. The estimated state vector and parameters for the corporate bonds result in a price of that bond using the formulas of the previous chapter. We compare this price with the observed price to check if the particular reduced form models are able to price defaultable and callable bond accurately. A comparison between the model of Jarrow et al. [20] for callable-bonds and the model of subsection 3.3.1 is also done.

All the codes used in this chapter were done using MATLAB and are given in Appendix G.

4.1 Estimation of the Default Free Process

We start by estimating the parameters and state vector of equation (3.4) that better fit the observations. As said in Chapter 3, a two factor CIR model is chosen. It still explains a high portion of the term structure and when considering the corporate bonds we will have to know which of the factors is the level and which is the slope. With only two factors this is easier to do, because the two factors will correlate very differently with spreads of yields and the yields themselves, giving a clear picture of who's who. Another

advantage is related to the numerical procedure. With less parameters to estimate it is easier for the maximization of the likelihood (see Appendix F) procedure to find a global maximum.

Data and Estimation Procedure

We choose the linear Kalman filter described in Appendix F to estimate the model, something first proposed by Chen and Scott [3] and that is also used by Duffee [11], Jarrow et al. [20] and Park and Clark [34]. It is particularly useful when we have unobserved/latent variables — the level and the slope. We only need to have a formula that relates these latent factors to the observations. The Kalman filter will estimate an optimal value for the state vector. It also includes the possibility of considering an error in the observations that can result from data entering errors, rounding of values, non-synchronous observations, etc. It is also a fast algorithm that relates our closed form solutions to the observations. Other usual methods are much more time consuming. A Markov Chain Monte Carlo is always possible, but Chatterjee [2] reports the estimation of a similar two-factor model by Lemourx and White using an MCMC algorithm: The estimation procedure took (...) *more than five days on a very sophisticated machine* (...). And even though we are talking about a "very sophisticated machine" by 2002 standards, that is a lot of time when compared with the linear Kalman filter method for a two-factor model that takes only a dozen of minutes to complete or a couple of hours if we consider an extended Kalman filter, on a not so very sophisticated machine.

As described in Appendix F, in the linear Kalman filter we relate the observations z_t with the state variables Y_t and the normal error ν_t by the measurement equation:

$$z_{t_i} = A + HY_{t_i} + \nu_{t_i}$$

For n observations and m state variables, A , z_{t_i} and ν_{t_i} are $n \times 1$ real matrices and H is a $n \times m$ real matrix.

The state variables follow the discretized dynamics:

$$Y_{t_i} = C + FY_{t_{i-1}} + \varepsilon_{t_i} \quad \varepsilon_{t_i} \sim N(0, Q)$$

Where C , Y_{t_i} and ε_{t_i} are $m \times 1$ and F is $m \times m$. All of them are real matrices.

For the case of a two-factor CIR model we have by subsection F.1.1 for a difference of time Δ between the observations:

$$C = \begin{bmatrix} \theta_1 [1 - e^{-k_1 \Delta}] \\ \theta_2 [1 - e^{-k_2 \Delta}] \end{bmatrix} \quad F = \begin{bmatrix} e^{-k_1 \Delta} & 0 \\ 0 & e^{-k_2 \Delta} \end{bmatrix}$$

$$Q = \begin{bmatrix} \frac{\sigma_1^2}{k_1} (1 - e^{-k_1 \Delta}) \left[\frac{\theta_1}{2} (1 - e^{-k_1 \Delta}) + e^{-k_1 \Delta} y_{1t} \right] & 0 \\ 0 & \frac{\sigma_2^2}{k_2} (1 - e^{-k_2 \Delta}) \left[\frac{\theta_2}{2} (1 - e^{-k_2 \Delta}) + e^{-k_2 \Delta} y_{2t} \right] \end{bmatrix}$$

And considering that we have yields on treasury bonds, we have by equation (3.4):

$$A = \alpha_0 - \frac{\psi_0(\tau)}{\tau} \quad H = \frac{1}{\tau} [\psi_1(\tau), \psi_2(\tau)]$$

After the estimation has procedure terminated we use Proposition (3.1) to write the estimated yields and compare them to the observations. This comparison is done using the root mean square error (RMSE), defined as:

$$RMSE_t = \sqrt{\frac{\sum_{j=1}^n (\hat{z}_{t,j} - z_{t,j})^2}{n}}$$

Where $\hat{z}_{t,j}$ represents the yield of maturity t at moment j and $z_{t,j}$ the observed yield.

The data from the Federal Reserve consists of monthly data of Treasury constant maturities for eleven different maturities. Not every maturity has historical data for the same months.

TABLE 4.1: Observation Dates for the Treasury Constant Maturity Rates

1 Month	Jul-01	Jun-14
3 Month	Jan-82	Jun-14
6 Month	Jan-82	Jun-14
1 Year	Apr-53	Jun-14
2 Year	Jun-76	Jun-14
3 Year	Apr-53	Jun-14
5 Year	Apr-53	Jun-14
7 Year	Jul-69	Jun-14
10 Year	Apr-53	Jun-14
20 Year	Apr-53	Jun-14
30 Year	Feb-77	Jun-14

Data available at federalreserve.gov/releases/h15/data.htm

Additionally, in the last ten years we do not have five years of data for the 30 years maturity and seven years for the 20 years. For the 1 month maturity only thirteen years of data are available. With this we choose to drop the maturities of 1 month, 20 and 30 years. For the other maturities, we chose the dates from January 1982 to June 2014, in a total of 390 observations.

The constant maturities form a par curve¹ and to use the linear Kalman filter we needed a zero curve, because only this way the yields are a linear function of the state vector, as seen from equation (3.4). We do this *via* bootstrapping².

Results

After running the linear Kalman filter we get the results for the parameters correspondent to the first two state variables Y_1 and Y_2 . They are given in Table 4.2, together with the value of the likelihood $l(\Theta)$. The Feller conditions hold for both variables but were only imposed for the first factor. The market prices of risk are negative for both factors, what leads, from equation (D.21), to positive excess returns. The value $\alpha_0 = -1$ is something that Dufee [11] also obtained. He says that, in fact, the estimation procedure resulted in a lower value of α_0 , but the fit did not improved when compared with -1 , and therefore he imposed a minimum value of -1 to α_0 . The same thing happened here.

TABLE 4.2: Parameters of the short-rate state variables

θ_1	k_1	σ_1	η_1
0.17008	0.07457	0.04710	-0.00522
θ_2	k_2	σ_2	η_2
0.89815	0.41898	0.01835	-0.00822
α_0	$l(\Theta)$		
-1.00000	15202		

To find which of the factors is related to the level of yields and which is related to the steepness of the term structure we find the correlation between the estimated factor for each month and several yields. The first factor is highly correlated with the yields. In particular it has a correlation of 0.98 with the 10 year yield. Factor 2 is less correlated with the yields, in particular with the yields of short maturity (-0.25 for the 3 month bond). In opposition, it is highly correlated with the spread of the 10 year bond over the 6 month bond. In shows a correlation of -0.77 (0.14 for the first factor). Therefore, we

¹See appendix A

²Using the MATLAB function *pyld2zero*.

identify the first factor with the level, and the second with the steepness. This will be important in subsequent estimations, since that the formula for the call spread depends on the level but not on the steepness.

The comparison between the results of the linear Kalman filter and the observed values, using the root mean square error in basis points, is given in Table 4.3.

TABLE 4.3: Root Mean Square Error (in basis points) for the Constant Maturities Treasury Rates

Bond Maturity	RMSE
3 Month	35.15
6 Month	15.33
1 year	0.04
2 year	13.76
3 year	12.35
5 year	6.22
7 year	5.12
10 year	15.47

The values in Table 4.3 are of the same magnitude than those of Dufee [11] and smaller than those of Park and Clark [34]. An example of the fit is given for the case of the 5 years in Figure 4.1. They are roughly of the order of 10 basis points - 0.1% - and therefore we consider the use of the two-factor model justified. It is a small value and therefore a good starting point for considering more complex contracts.

4.2 Estimation of the Default Spread

The estimation of the default spread is done considering Proposition 3.2. It deals with zero coupon bonds but we can always write a coupon bond as a sum of zero coupon bonds. Equation (3.6) involves the risk-free variables and respective parameters estimated in the last section, and we use the estimated values of the parameters and of the latent variables as inputs for the estimation of the default process. Therefore, we will have one price of a non-callable corporate bond to estimate and a state variable to estimate.

Data and Estimation Procedure

The data corresponds to prices of defaultable corporate bonds. We choose two issuers and one bond from each one. The data must be consistent with the results for the

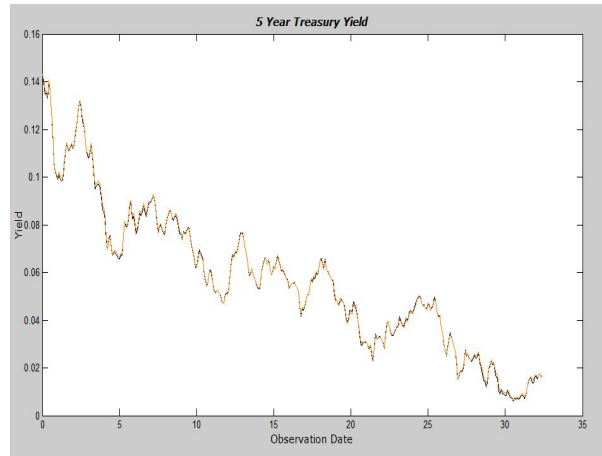


FIGURE 4.1: In black the observations and in orange the yields given by the model for the 5 year treasury yield

risk-free process and, therefore, the dates of the observations must overlap. Since we must have a reasonable number of data points to perform the estimation, it is imposed that the bonds must have at least twenty four months of observations. To control for liquidity³ only bonds emitting more than five million dollars are considered. Finally, they have, of course, to be non-callable. Two bonds from different issuers that satisfy this simple restrictions are described in Table 4.4. They also satisfy other, more restrictive, restrictions that we only explain in following sections.

TABLE 4.4: Corporate Defaultable Non-Callable Bonds

Issuer	Bloomberg ID	Coupon	Maturity	Date Observations
Walt Disney	EC5894893	6.2% S/A ⁴	20/06/2014	Jan2010-Apr2013
Bank of America	ED435463	6.05% S/A	01-06-2034	Mar2011-Jun2014

In fact, the raw data corresponds to weekly observations, but we transform it to monthly data by considering the means of the weeks in a given month.

For the estimation procedure, the equations of the descretization maintain their form with the obvious difference that we only have one unknown variable now, corresponding to the state variable Y_3 . Therefore:

$$C = \theta_3 \left[1 - e^{-k_3 \Delta} \right] \quad F = e^{-k_3 \Delta} \quad Q = \frac{\sigma_3^2}{k_3} \left(1 - e^{-k_3 \Delta} \right) \left[\frac{\theta_3}{2} \left(1 - e^{-k_3 \Delta} \right) + e^{-k_3 \Delta} y_{3t} \right]$$

³Jarrow et al. [20] also consider the risk factor *liquidity* and model it using an intensity.

⁴semi annual

The measurement equation now is not a linear function of the state vector and therefore we need to apply the extended Kalman filter described in Section F.2 of Appendix F. We have:

$$z_{t_i} = \Phi(y_{t_i}) + v_{t_i}$$

with

$$\Phi(y_3(t)) = \sum_{t < T_i \leq T} cV(t, T_i, 0, \delta_t) + V(t, T, 0, \delta_t)$$

Where $V(t, T, 0, \delta_t)$ is given by equation (3.6). The linearisation enables us to use the linear Kalman filter with:

$$\begin{aligned} A &= \sum_{t < T_i \leq T} cf(Y_1, Y_2, t, T) g(\mathbb{E}[y_{3,t_i} | \mathcal{F}_{t_{i-1}}], t, T) \\ &+ \sum_{t < T_i \leq T} c\psi_3 f(Y_1, Y_2, t, T) g(\mathbb{E}[y_{3,t_i} | \mathcal{F}_{t_{i-1}}], t, T) \mathbb{E}[y_{3,t_i} | \mathcal{F}_{t_{i-1}}] \\ &+ (1 + \psi_3 \mathbb{E}[y_{3,t_i} | \mathcal{F}_{t_{i-1}}]) f(Y_1, Y_2, t, T) g(\mathbb{E}[y_{3,t_i} | \mathcal{F}_{t_{i-1}}], t, T) \end{aligned}$$

And :

$$H = \sum_{t < T_i \leq T} c\psi_3 f(Y_1, Y_2, t, T) g(\mathbb{E}[y_{3,t_i} | \mathcal{F}_{t_{i-1}}], t, T) + \psi_3 f(Y_1, Y_2, t, T) g(\mathbb{E}[y_{3,t_i} | \mathcal{F}_{t_{i-1}}], t, T)$$

By Proposition 3.2, we have the following identities:

$$f(Y_1, Y_2, t, T) = e^{-\alpha_0 \tau + \psi_{0,1}(\tau) + \psi_{0,2}(\tau) - \psi_1(\tau)Y_1 - \psi_2(\tau)Y_2} \quad g(Y_3, t, T) = e^{-\gamma_0 \tau + \psi_{0,3}(\tau) - \psi_3(\tau)Y_3}$$

The remaining difference compared to the linear Kalman filter is that the expected value of the observations can now be computed with the non-linear function and not with A and H , i.e, $\mathbb{E}[z_{t_i} | \mathcal{F}_{t_{i-1}}] = \Phi(\mathbb{E}[y_{t_i} | \mathcal{F}_{t_{i-1}}])$ instead of using the linearised version given above.

The RMSE will be calculated for the implicit yield to maturity given by the prices. We could do it using the prices, but this way we can compare with the results of Dufee [11] and Park and Clark [34] since they use the yield-to-maturity (YTM).

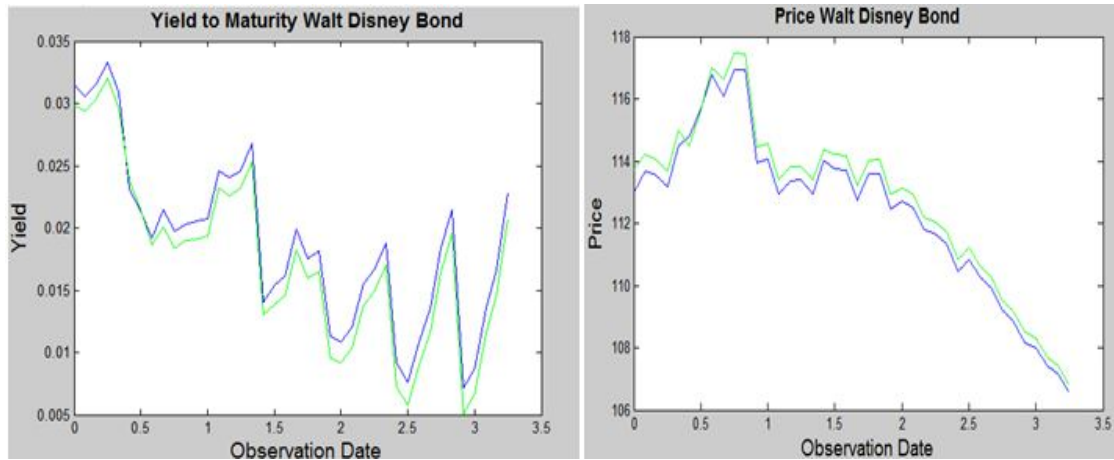


FIGURE 4.2: The results from the estimation process (green) and the real values (blue) for the default process.

Results

In Table 4.5 we have the results of the parameters estimations for the two firms, and not so different values for the dynamics of the state variable related to the financial health of the firm are found, but we have very different dependence on the short-rate factors exists. The Bank of America bond has a much higher dependence on this factors. Maybe this is not surprising for a bank. For γ_0 the estimation preferred positive values, but this was not imposed. We have different time periods for the estimation, but the means of the variable Y_3 for each bond and the respective time period means for Y_1 and Y_2 give us a mean spread, by equation (3.5), of 0.0428 for the Bank of America bond and 0.0121 for the Walt Disney bond.

TABLE 4.5: Default Spread Parameters

TABLE 4.6: Bank of America

θ_3	k_3	σ_3	η_3
0.0215	0.3200	0.1174	-0.2946
γ_0	β_{d_1}	β_{d_2}	$l(\Theta)$
0.6263	-0.6700	-0.6100	92.9991

TABLE 4.7: Walt Disney

θ_3	k_3	σ_3	η_3
0.0500	0.7599	0.2700	-0.1200
γ_0	β_{d_1}	β_{d_2}	$l(\Theta)$
0.0039	-0.0309	-0.0328	127.0989

In Figure 4.2 we have the results of the fit to the price and yield-to-maturity of the Disney bond. The RMSE values in basis points for the yields to maturity are given in Table 4.8:

TABLE 4.8: RMSE for the non-callable bonds

Bank of America	97.62
Walt Disney	15.89

The value of RMSE for the Bank of America bond is quite poor. Almost an order of magnitude higher than the errors found for the treasury rates. Nonetheless, they are not that different from the results of Park and Clark [34]. Their best values are only half the result obtained for the Bank of America bond. In the case of Dufee [11] his results are of the same order of magnitude than that of the Disney bond obtained here, than this order of magnitude coincide with that found for treasury rates and, therefore, we consider this a good result.

The reduced form model considered for defaultable non-callable works well at least for some bonds, but it faces a more difficult test with callable bonds, mainly because we have to fit a non-linear expression.

4.3 Make-Whole Callable Bonds

We estimate the Park and Clark [34] model for bonds with make-whole call provisions using Proposition 3.4. These bonds are increasingly popular and so finding a model that prices them correctly is of particular importance.

Data and Estimation Procedure

There are two ways in which we can estimate the make-whole call spread.

1. Estimate the make-whole call spread and default spread jointly

Here we consider the short-rate estimation of Section 4.1 but not the default spread independently estimated from Section 4.2. It means that we have more parameters to estimate, and that we have at each moment, for a given vector of parameters, one price (of the make-whole callable bond) and two state variables to be estimated in a function of the form:

$$z_{t_i} = \Phi(Y_{3,t_i}, Y_{5,t_i}) + \nu_{t_i}$$

This means that the system is under-identified. Considering Proposition 3.4, the use of the Kimmel series expansion means that the expressions for A and H are more complex than in the previous section. Moreover, H is in this case a vector since we have two state variables to estimate. We also have one of the variables to estimate presented in the series expansion, what implies one more source of complexity. This do not happens in approach 2. We apply the extended Kalman filter with:

$$\begin{aligned}
A &= \sum_{t < T_i \leq T} c (1 + \psi_5 \mathbb{E} [y_{5,t_i} | \mathcal{F}_{t_{i-1}}]) f(Y_1, Y_2, t, T) g(\mathbb{E} [y_{5,t_i} | \mathcal{F}_{t_{i-1}}], t, T) \pi(\mathbb{E} [y_{3,t_i} | \mathcal{F}_{t_{i-1}}], t, T) \\
&\quad - \sum_{t < T_i \leq T} c f(Y_1, Y_2, t, T) g(\mathbb{E} [y_{5,t_i} | \mathcal{F}_{t_{i-1}}], t, T) \frac{\partial \pi}{\partial Y_3} \Big|_{\mathbb{E}[y_{3,t_i} | \mathcal{F}_{t_{i-1}}]} \mathbb{E} [y_{3,t_i} | \mathcal{F}_{t_{i-1}}] \\
&\quad + (1 + \psi_5 \mathbb{E} [y_{5,t_i} | \mathcal{F}_{t_{i-1}}]) f(Y_1, Y_2, t, T) g(\mathbb{E} [y_{5,t_i} | \mathcal{F}_{t_{i-1}}], t, T) \pi(\mathbb{E} [y_{3,t_i} | \mathcal{F}_{t_{i-1}}], t, T) \\
&\quad - f(Y_1, Y_2, t, T) g(\mathbb{E} [y_{5,t_i} | \mathcal{F}_{t_{i-1}}], t, T) \frac{\partial \pi}{\partial Y_3} \Big|_{\mathbb{E}[y_{3,t_i} | \mathcal{F}_{t_{i-1}}]} \mathbb{E} [y_{3,t_i} | \mathcal{F}_{t_{i-1}}]
\end{aligned}$$

And $H = [H_3, H_5]$ where:

$$\begin{aligned}
H_3 &= f(Y_1, Y_2, t, T) g(\mathbb{E} [y_{5,t_i} | \mathcal{F}_{t_{i-1}}], t, T) \frac{\partial \pi}{\partial Y_3} \Big|_{\mathbb{E}[y_{3,t_i} | \mathcal{F}_{t_{i-1}}]} \\
&\quad + \sum_{t < T_i \leq T} f(Y_1, Y_2, t, T) g(\mathbb{E} [y_{5,t_i} | \mathcal{F}_{t_{i-1}}], t, T) \frac{\partial \pi}{\partial Y_3} \Big|_{\mathbb{E}[y_{3,t_i} | \mathcal{F}_{t_{i-1}}]}
\end{aligned}$$

$$\begin{aligned}
H_5 &= -\psi_5 f(Y_1, Y_2, t, T) g(\mathbb{E} [y_{5,t_i} | \mathcal{F}_{t_{i-1}}], t, T) \pi(\mathbb{E} [y_{3,t_i} | \mathcal{F}_{t_{i-1}}], t, T) \\
&\quad - \psi_5 f(Y_1, Y_2, t, T) g(\mathbb{E} [y_{5,t_i} | \mathcal{F}_{t_{i-1}}], t, T) \pi(\mathbb{E} [y_{3,t_i} | \mathcal{F}_{t_{i-1}}], t, T)
\end{aligned}$$

This approach is necessary when we have an issuer that does not have a non-callable bond to be used to estimate the default process. In the perfect scenario, we do not only need a callable and a non-callable bond but also need bonds of the same priority in the sense that if the callable bond is junior the non-callable bond also needs to be junior.

2. Estimation of the make-whole call spread given that the default spread is already estimated

This have been already done when considering non-callable bonds after having estimated the short-rate process. We consider the results given for Y_1 , Y_2 and Y_3 and estimate only the parameters associated with Y_5 . In this case, the Kimmel series expansion in the pricing formula is known given these state variables and respective parameters. The expressions for A and H are simpler than in the previous case:

$$\begin{aligned}
A &= \sum_{t < T_i \leq T} c (1 + \psi_5 \mathbb{E} [y_{5,t_i} | \mathcal{F}_{t_{i-1}}]) f(Y_1, Y_2, t, T) g(\mathbb{E} [y_{5,t_i} | \mathcal{F}_{t_{i-1}}], t, T) \pi(Y_3, t, T) \\
&\quad + (1 + \psi_5 \mathbb{E} [y_{5,t_i} | \mathcal{F}_{t_{i-1}}]) f(Y_1, Y_2, t, T) g(\mathbb{E} [y_{5,t_i} | \mathcal{F}_{t_{i-1}}], t, T) \pi(Y_3, t, T)
\end{aligned}$$

$$H = -\psi_5 f(Y_1, Y_2, t, T) g\left(\mathbb{E}[y_{5,t_i} | \mathcal{F}_{t_{i-1}}], t, T\right) \pi(Y_3, t, T) \\ - \psi_5 f(Y_1, Y_2, t, T) g\left(\mathbb{E}[y_{5,t_i} | \mathcal{F}_{t_{i-1}}], t, T\right) \pi\left(\mathbb{E}[y_{3,t_i} | \mathcal{F}_{t_{i-1}}], t, T\right)$$

In both approaches we face a problem when the value of the state variable Y_3 is zero, since we have in the Kimmel series expansion a term of the form $\frac{1}{Y_3}$. To overcome this difficulty we choose to fix that value at time t_i to the mean of the previous and following values.

To estimate these models, we need to have a make-whole callable bond. We imposed the already considered restriction of emission value in order to be able to use approach 2 we still need a make-whole callable bond that has prices in the same dates of an already considered bond in Section 4.2. To control for maturity premium it is also relevant to consider a callable bond with a similar maturity to the non-callable one.

A callable bond from The Walt Disney Company that meets this criteria is presented in Table 4.9.

TABLE 4.9: Walt Disney Make-Whole Callable Bonds

Bloomberg ID	Coupon	M_{mw}	Maturity	Observation Dates
EG6440481	6.0% S/A	0.15%	7/2017	Jan2010-Apr2013

Results

For the two different approaches is estimated the make-whole call process and only in approach 1 the default process as well.

TABLE 4.10: Disney Make-Whole Callable Bond Parameters

TABLE 4.11: Approach 1

θ_3	k_3	σ_3	η_3
0.0836	0.4340	0.0916	-0.1974
θ_5	k_5	σ_5	η_5
0.0611	0.7631	0.1254	-0.4833
γ_0	χ_0	β_{d_1}	β_{d_2}
-0.9241	0.6819	0.0903	0.0556
χ_3	$l(\Theta)$	RMSE	
0.1751	109.3092	50.33	

TABLE 4.12: Approach 2

θ_5	k_5	σ_5	η_5
0.3598	0.5133	0.0156	-0.3200
χ_0	χ_3	$l(\Theta)$	RMSE
-0.6950	10^{-9}	109.4114	14.03

In approach 1 we obtain values for the default parameters that are different from those obtained by approach 2. Because of the negative value of γ_0 and the overall lower β' s, we

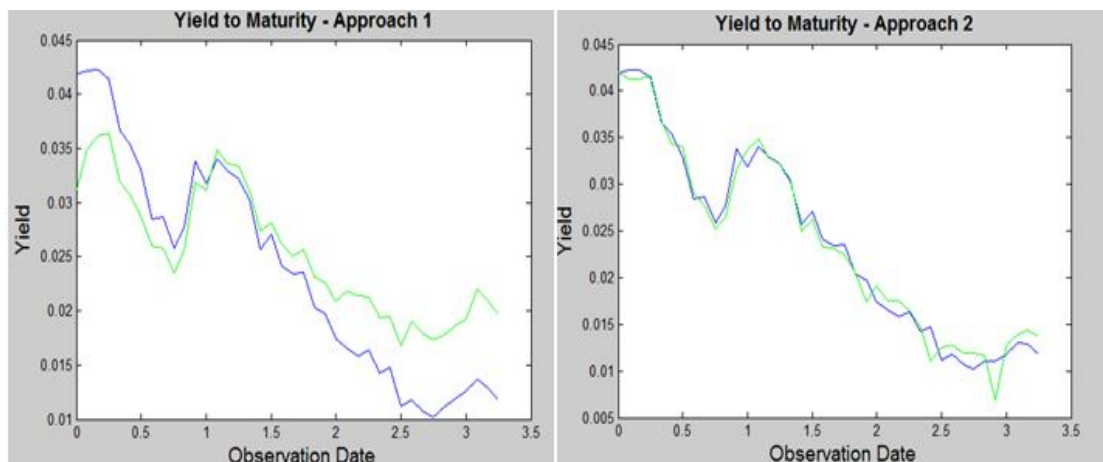


FIGURE 4.3: Yields to maturity given by the model (green) and the 'real' ones (blue) in both approaches.

have a lower default spread in approach 1 than in approach 2. To somewhat compensate, approach 2 has a lower call spread since it has a negative χ_0 in opposition to approach 1. In approach 2 we have a lower error, something I find surprising given the higher constraints of this method. It is nonetheless less surprising when we run the method and see the consequences of having considerably more parameters to estimate, and a state variable to also estimate at each observational time that is found in a non-linear function. In ideal conditions we should use approach 2, but nonetheless the error in approach 1 is not much different from those of Park and Clark [34]. Using an identical model and approach 2 with several issuers, they find errors with the order of magnitude of those found in our approach 1.

When compared with previous results we can say that the results from approach 2 are good, since they have similar errors to those found when estimating the treasury rates and the default parameters of the Disney bonds. In Figure 4.3 we have the fit to the yields-to-maturity in both approaches.

4.4 Standard Callable Bonds

In the case of callable bonds we presented a model that is slightly different from that of Jarrow et al. [20]. In this section we will compare them checking which of them fits better the observations, considering that the model presented in Chapter 3 has one more parameter to be estimated. We compare the two models considering only the three step estimation procedure, i.e, the equivalent of approach 2 of section 4.3 given that approach 1 faces some numerical difficulties, especially in this case where we have two Kimmel series expansions.

Data and Estimation Procedure

Since we just consider approach 2, the results given for this estimation method in section 4.3 still hold, only with the obvious change $Y_5 \rightarrow Y_4$, since Y_4 is the variable associated with the call process and Y_5 is associated with the make-whole call process.

The data restrictions are also the same considered in the last section. Now we have a Bank of America bond described in Table 4.13.

TABLE 4.13: Bank of America Callable Bonds

Bloomberg ID	Coupon	Maturity	Observation Dates
EI406362	5.0% S/A	09/2035	Mar2011-Jun2014

Results

The model of Section 3.3.1 and the model of Jarrow et al. [20] are both considered. The model of Jarrow et al. [20] has $\xi_1 = 0$, and therefore it does not contains in the call spread a non-linear function of the level of treasury yields.

The results for both models are given in Table 4.14. The subscript J means that the parameter considered is that of Jarrow et al. [20].

TABLE 4.14: Parameters of the short-rate state variables

θ_4	k_4	σ_4	η_4
0.0200	0.0100	0.0016	-0.9999
θ_{4J}	k_{4J}	σ_{4J}	η_{4J}
0.0200	0.0100	0.0016	-0.9998
ξ_0	ξ_{0J}	ξ_3	ξ_{3J}
-0.8998	-0.8999	8.00×10^{-12}	8.02×10^{-12}
ξ_1	<i>RMSE</i>	<i>RMSE_J</i>	
0.0998	16.57	30.57	

The common parameters are basically all the same in both models. Surprisingly the parameter ξ_3 , related with the default variable, is almost zero in both cases, even though it appears in the spread with the inverse of a state variable, increasing, therefore, this term of the call spread. This term is completely negligible when compared with the parameter ξ_1 that does not appear in the Jarrow et al. [20] model. Moreover, the value of the error is lower in the modified Jarrow model, and all of it can be explained by the additional term, given that all the others parameters are equal.

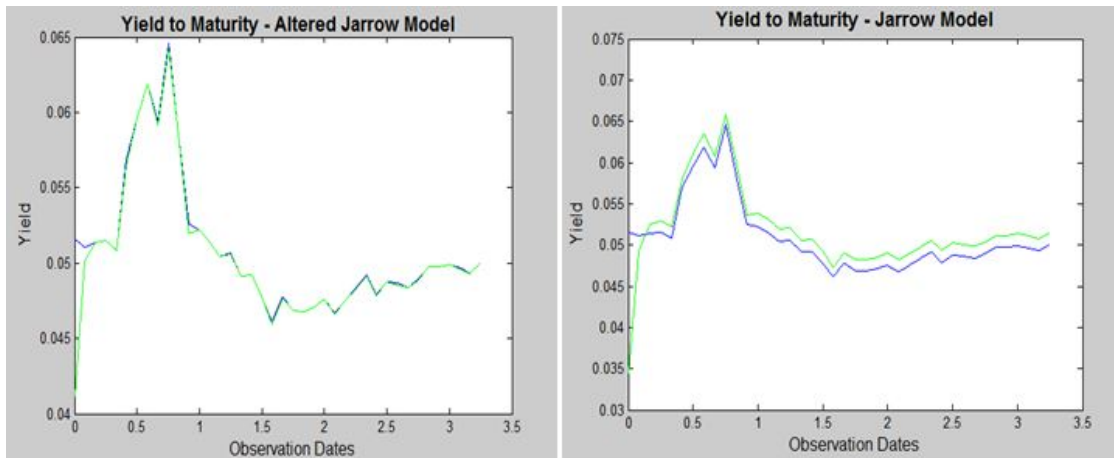


FIGURE 4.4: Yields to maturity given by the model (green) and the 'real' ones (blue) in both approaches.

In Figure 4.4 we have the yield-to-maturity fit of each model. We see by this figure and Table 4.14 that the modified Jarrow model is preferred to the original one given the much lower value of the error. This error is similar to those already found in others estimations, and, therefore, we consider it a good fit, contrary to the fit of the original Jarrow model. This only came at the cost of considering one more parameter.

5

Conclusions

The ability of reduced form models to fit the observed prices of corporate non-callable and callable bonds was tested. To do this, we considered a similar model to that of Duffee [11] to price defaultable non-callable bonds, the model of Jarrow et al. [20] for standard callable bonds, and the model of Park and Clark [34] for bonds with make-whole call provisions. Additionally, for the case of standard callable bonds, a slightly altered Jarrow model, that includes a non-linear dependence on the level of treasury rates, was also considered. The models that were used to price bonds with call provisions contain a non-linear dependence on the state variables and, therefore, closed-form solutions are not known. Fortunately, the results of Kimmel [24] enable us to write a series expansion approximation for the expectation of the discounted values of possible future pay-offs for the particular non-linear dependence considered.

The estimation process consisted of a linear Kalman filter for the case of treasury constant maturities yields and extended Kalman filters when dealing with corporate bond prices. With a two-factor CIR model we found errors roughly in the range [5, 15] basis points for the treasury yields. This represents a small value — around 1% of the average yield — and is used as the benchmark when considering the yields-to-maturity of corporate bonds. For the two non-callable corporate bonds, we found, using the extended Kalman filter, a error of the same order of magnitude for one of them, but for the other we found an error of an order of magnitude higher — a very poor fit.

When dealing with bonds with make-whole call provisions we tested two estimation methods. A two-stage and a three-stage estimation processes. In the first approach we estimated the default and the call spreads together, and in the second we used the values of the already estimated default process as input, an estimated only the make-whole call process. The first approach is computational much more complex and we found smaller errors using approach 2. The error in approach 2 is, in basis points, of the same order

of magnitude of that found in treasury yields and for yield-to-maturities in the default process of non-callable bonds, but, just as for the case of non-callable bonds, this error represents a higher percentage of the yield when compared with the model for treasury yield.

For the case of standard callable bonds only approach 1 was used, and the model of Jarrow et al. [20] performed worse than the one that adds a non-linear dependence on the level of treasury yields. This altered model has the same error than the ones of non-callable and make-whole call bonds.

Overall the reduced form models with the recovery of market value assumption fitted the data worse than a two-factor CIR model fits treasury rates, but the RMSE are smaller than 10% of the average yield-to-maturity.

Appendix A

Bond Concepts

The main objective of this thesis is to obtain prices of defaultable callable and non-callable bonds, and here we give some definitions and concepts involving them.

The most basic is the concept of face value:

Definition A.1 (Face Value). The nominal value of the bond, which is the payment at maturity.

The most important bond when considering pricing problems, but that represents a small percentage of traded bonds, is the zero coupon bond:

Definition A.2 (Zero coupon bond). A bond that has only one payment: the face value at maturity.

Most of the bonds have more than one payment, but ultimately all of them can be represented as a sum of zero coupon bonds.

Definition A.3 (Coupon Bond). A bond that pays a periodic coupon c to the bondholder and the face value at maturity.

We have the possibility that the issuer can default and the bondholder does not receive the promised payments after default. It is now natural to introduce the concept of defaultable bond.

Definition A.4 (Defaultable Bond). Any bond that has its payments uncertain due to the possibility of default.

More complex bonds that represent a large share of the bond market have more complex characteristics, including call options.

Definition A.5 (Callable Bond). These contracts allow the issuer to redeem the bond at pre-determined dates before maturity at a fixed cost.

Definition A.6 (Bond With Make-Whole Call Provision). With these bonds the issuer can redeem the bond at some dates before maturity but, contrary to callable bonds, the cost X_{mw} is not fixed. It is the maximum value between the face value FV and the remaining cash-flows discounted at some stochastic rate r plus a fixed make-whole premium M .

$$X_{mw} = \max \left\{ FV, \sum_{t < T_i \leq T} CF_i e^{-(r_{t,T_i} + M_{mw})(T_i - t)} \right\}$$

When we are discounting cash-flows it is important to have in mind the different types of interest rates.

Definition A.7 (Nominal Rate). An annual interest rate i_n with n compound periods in one year.

Definition A.8 (Effective Interest Rate). A nominal rate r_1 with one compound period in one year.

$$1 + r_1 = \left(1 + \frac{i_n}{n} \right)^n \quad (\text{A.1})$$

The models we consider in this thesis give us the dynamics of the short-rate.

Definition A.9 (Short Rate). The short rate r_t is the risk-free rate for an infinitesimal period.

Associated with the short-rate is the money market account.

Definition A.10 (Money Market Account). A risk-free asset compounded continuously at the short rate r_t . Therefore, the price of this asset is given by:

$$dB = r_t B_t dt$$

Another type of interest rate is the continuously compounded rate that pays an instantaneously rate.

Definition A.11 (Continuously Compounded Rate). For any asset, risky or not, we can define the continuously compounded rate $r_c(0, n)$ for a maturity n considering the same dynamics as for the money market account. It is related with the effective rate for the same period by:

$$e^{r_c(0, n)} = 1 + r_1 \quad (\text{A.2})$$

In short rate models the dynamics of the short rate will ultimately give us the zero curve:

Definition A.12 (Zero Curve). It is a curve where we have the today's value of one monetary unit for a continuous of maturities. This is the discount factor that for n years we denote

$$P(0, n) = \frac{1}{\left[1 + \frac{i_m}{m}\right]^{mn}} = e^{-r_c(0, n)n}$$

Definition A.13 (Par Curve). The curve for a continuous of maturities where we have the value of the coupons that make the bond being traded at par. This way considering m periods per year and a maturity of n years the par yield $y_m(n)$ associated with this maturity is such that:

$$1 = \sum_{j=1}^{mn} \frac{y_m(n)}{m} P\left(0, \frac{j}{m}\right) + P(0, n) \quad (\text{A.3})$$

If $n < 1$ the par curve and the zero curve coincide.

Example A.1 (Bootstrapping). It is a method of extracting the zero curve from the par curve. Consider the following par yields where the compounding is semi-annual:

1Month	3Months	6Months	1year	2years
0.91%	0.96%	1.11%	1.43%	2.07%

The three yields for the three first maturities are already zero yields. We can use (A.1) and (A.2) to have them in other format, but nothing more. For the one year we use the formula (A.3)

$$1 = \frac{y_2(1)}{2} P\left(0, \frac{1}{2}\right) + \frac{y_2(1)}{2} P(0, 1) + P(0, 1)$$

Because we already know $P\left(0, \frac{1}{2}\right)$ it is easy to obtain: $r_c(0, 1) = 1.4260\%$

In a similar way we have:

$$1 = \frac{y_2(1)}{2} P\left(0, \frac{1}{2}\right) + \frac{y_2(1)}{2} P(0, 1) + \frac{y_2(1)}{2} P(0, 1.5) + \frac{y_2(1)}{2} P(0, 2) + P(0, 2)$$

Now we have two unknowns: $P(0, 1.5)$ and $P(0, 2)$, but we can write the rate $r(0, 1.5)$ through the linear interpolation:

$$r(0, 1.5) = r(0, 1) + [r(0, 2) - r(0, 1)] \frac{1.5 - 1}{2 - 1}$$

And now we have only $r(0, 2)$ to obtain, which can be done by some numerical method.

Appendix B

Probability Concepts

We give some basic mathematical definitions and propositions used in the rest of this document. Most of them can be found in [21] and [22].

Definition B.1. A σ -field on a set Ω is a collection \mathcal{F} of subsets of Ω satisfying:

- $\emptyset, \Omega \in \mathcal{F}$
- $X \in \mathcal{F} \Rightarrow \Omega \setminus X \in \mathcal{F}$
- $X_n \in \mathcal{F}, \forall n \in \mathbb{N} \Rightarrow \bigcup_{n=1}^{\infty} X_n \in \mathcal{F}$

Most of the time we are interested on information of one stochastic process available up to some intermediate point.

Definition B.2. A stochastic process is a family of random variables $(Y_t)_{t \in I}$ defined in the same (Ω, \mathcal{F}) .

Definition B.3. The function $(Y_t)_{t \in I} : (\Omega, \mathcal{F}) \rightarrow (\Omega', \mathcal{F}')$ is a random variable with values on Ω' if it is measurable, i.e: $\forall X' \in \mathcal{F}' : Y_t^{-1}(X') \in \mathcal{F}$.

Proposition B.4. With the probability space (Ω, \mathcal{F}, P) and the sub- σ -field \mathcal{G} of \mathcal{F} , if η is a random variable and ξ a \mathcal{G} -measurable random variable:

$$\mathbb{E}[\xi\eta | \mathcal{G}] = \xi \mathbb{E}[\eta | \mathcal{G}]$$

Definition B.5. Let \mathcal{G} be a sub- σ -field of \mathcal{F} ($\mathcal{G} \subseteq \mathcal{F}$) and ξ a random variable. The conditional expectation of ξ given \mathcal{G} is a random variable $\mathbb{E}[\xi | \mathcal{G}]$ such that:

1. $\mathbb{E}[\xi | \mathcal{G}]$ is \mathcal{G} -measurable

2. If $A \in \mathcal{G}$ then $\mathbb{E}[\mathbb{I}_A \mathbb{E}[\xi | \mathcal{G}] | \mathcal{F}] = \mathbb{E}[\mathbb{I}_A \xi | \mathcal{F}]$

A generalization of measurable random variables is the concept of adapted stochastic process.

Definition B.6. A filtration is a family of *sub- σ -fields* of \mathcal{F} , $(\mathcal{F}_t)_{t \in I}$, satisfying:
 $\forall t \in I, \text{ if } 0 \leq s \leq t \Rightarrow \mathcal{F}_s \subset \mathcal{F}_t$

We say that the filtration $(\mathcal{F}_t)_{t \in I}$ satisfies the usual conditions if:

- $X \in \mathcal{F}, \mathbb{P}(X) = 0 \Rightarrow X \in \mathcal{F}_t, \forall t$
- $\mathcal{F}_t = \mathcal{F}_{t^+} := \bigcap_{u > t} \mathcal{F}_u, \forall t$

Definition B.7. A stochastic process $(Y_t)_{t \in I}$ is adapted to $(\mathcal{F}_t)_{t \in I}$ if all the random variables Y_t are \mathcal{F}_t -measurable.

The pricing problem is solved using expected values, and therefore we must define a probability space.

Definition B.8. A probability measure \mathbb{P} on (Ω, \mathcal{F}) is a real-valued function $\mathbb{P} : F \rightarrow [0, 1]$ that satisfy:

- $\mathbb{P}(\Omega) = 1$
- If $X_n \in \mathcal{F}$ satisfy $X_i \cap X_j = \emptyset$ then: $\mathbb{P}(\bigcup X_{n \in \mathbb{N}}) = \sum_{n \in \mathbb{N}} \mathbb{P}(X_n)$

We say that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space.

In the pricing problem we usually change between a physical measure \mathbb{P} and an equivalent pricing measure \mathbb{Q} .

Definition B.9. The measure \mathbb{Q} is equivalent to the probability measure $\mathbb{P}, \mathbb{P} \sim \mathbb{Q}$, if
 $\forall X \in F, \mathbb{Q}(X) > 0 \Leftrightarrow \mathbb{P}(X) > 0$.

Definition B.10 (Pricing Measure). The pricing measure \mathbb{Q} is the measure where a risky asset has the same expected value as the risk-free asset.

The existence of this measure is linked to arbitrage opportunities.

Theorem B.11 (First Fundamental Theorem of Arbitrage Pricing). *The market is free from arbitrage if and only if there is a pricing measure \mathbb{Q} .*

Also useful is the concept of Martingales because they represent the notion of fair-games.

Definition B.12. The stochastic process $(M_t)_{t \in I}$ defined in $(\Omega, \mathcal{F}, \mathbb{P})$ is a martingale if it satisfies:

- $\forall t \in I, \mathbb{E}(|M_t|) < \infty$
- $(M_t)_{t \in I}$ is adapted to $(\mathcal{F}_t)_{t \in I}$
- $\forall s, t \in I$ with $0 \leq s \leq t, \mathbb{E}(M_t | \mathcal{F}_s) = M_s$

The Girsanov theorem says how we will change the dynamics of a stochastic process when we change the measure:

Theorem B.13 (Girsanov). *Let $\{\phi_t\}_{t \in I}$ be a process adapted to the natural filtration of a Wiener process $\{W_t\}_{t \in I}$. Let's assume that z_t is a martingale, where z_t is given by:*

$$z_t = e^{\int_0^t \phi_s dW_s - \frac{1}{2} \int_0^t \phi_s^2 ds} \quad (\text{B.1})$$

A probability measure \mathbb{Q} equivalent to \mathbb{P} can be defined by:

$$\frac{d\mathbb{Q}}{d\mathbb{P}} \Big|_{\mathcal{F}_t} = z_t$$

And the process

$$W_t^{\mathbb{Q}} = W_t - \int_0^t \phi_s ds$$

is a Brownian motion under \mathbb{Q} .

A sufficient requirement to have (B.1) is the Novikov condition:

$$\mathbb{E}_{\mathbb{P}} \left[e^{\frac{1}{2} \int_0^t \phi_s^2 ds} \right] < \infty$$

A useful lemma when dealing with stochastic processes using Brownian motions $(W_t)_{t \geq 0}$ is the Itô lemma that gives us the differential of a function of a stochastic process.

Lemma B.14 (Itô Lemma). *For an n -dimensional Itô drift-diffusion process:*

$$d\bar{Y}_t = \mu(t)dt + \sigma(t)d\bar{W}_t$$

And a differentiable scalar function $f(t, \bar{Y}_t)$, we have:

$$df(t, \bar{Y}_t) = \frac{\partial f}{\partial t} + (\nabla_y^T f) d\bar{Y}_t + \frac{1}{2} (d\bar{Y}_t^T) (\nabla_y^2 f) d\bar{Y}_t$$

The Feynman-Kac formula links stochastic processes with differential equations:

Proposition B.15 (Feynman-Kac Formula). *We suppose that x_t follows the stochastic process:*

$$dx_t = \mu(x_t, t) dt + \sigma(x_t, t) dW_t^{\mathbb{Q}}$$

And let $V(x_t, t)$ be continuous satisfying the differential equation:

$$\frac{\partial V}{\partial t} + \mu(x_t, t) \frac{\partial V}{\partial x} + \frac{1}{2} \sigma(x_t, t)^2 \frac{\partial^2 V}{\partial x^2} - r(x_t, t) V(x_t, t) = 0$$

With terminal condition $V(x_T, T)$ and assuming also that

$$\int_t^T \mathbb{E}_{\mathbb{Q}} \left(\left| \sigma(x_t, t) \frac{\partial V}{\partial x} \right|^2 \right) du < \infty$$

$V(x_t, t)$ has the solution:

$$V(x_t, t) = \mathbb{E}_{\mathbb{Q}} \left[e^{-\int_t^T r(x_u, u) du} V(x_T, T) \mid \mathcal{F}_t \right]$$

Of importance is also the cumulative distribution function that is defined in the following manner

Definition B.16 (Cumulative Distribution Function). Given a real-valued random variable X , the cumulative distribution function is defined as:

$$F_X(h) = \mathbb{P}(X \leq h)$$

Associated with the cumulative distribution function is the probability density function that, instead of giving the probability of the random variable attaining a value below some real number, is the probability of having a value in an infinitesimal interval.

Definition B.17 (Probability Density Function). Given a real-valued random variable X , the probability density function is defined as:

$$f_X(h) = \frac{d}{dh} F_X(h)$$

Example B.1. *A \mathbb{R} -valued random variable x has a Gaussian distribution if the probability density function have the form*

$$f(x) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{\frac{1}{2}}} \exp \left[-\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right] \quad (\text{B.2})$$

Where $|\Sigma|$ is the determinant of the covariance matrix and μ is the expected value of the random variable.

If we sum two Gaussian random variables we still have a Gaussian random variable.

Proposition B.18. *If x and y are independent Gaussian random variables, $x \sim (\mu_x, \Sigma_x)$ and $y \sim N(\mu_y, \Sigma_y)$, then the sum $z = x + y$ is also normally distributed:*

$$z \sim N(\mu_x + \mu_y, \Sigma_x + \Sigma_y)$$

Of special importance when deriving likelihoods for dynamical times series models is the joint probability density function. For two random variables X and Y we have:

$$f_{X,Y}(x, y) = f_{Y|X}(y|x)f_x(x) \tag{B.3}$$

Appendix C

Cox Process

Poisson processes can model systems where the occurrence of some event is unpredictable but has some statistical regularity. If this regularity is time independent, i.e, the probability of an occurrence in a time interval is independent of our place in time, the process is time stationary and the particular Poisson process is an homogeneous Poisson process. If it is not time independent but the dependence is deterministic we will have an inhomogeneous Poisson process. For both cases we will find a mathematical way of describing the probability of occurrence of such event. From there we will generalize the concept to a stochastic probability of occurrence that is used when modelling default in Chapter 2. The mathematical description of Credit Risk in terms of Poisson and Cox processes can be found in [21] and [27].

C.1 Default as a Poisson Process

We model the default time as an unpredictable event that occurs at a time τ_d . This time is the first jump of a counting process $N_{d,t}$ defined as: $N_{d,t} = \mathbf{1}_{\{\tau_d > t\}}$ with $N_{d,0} = 0$. The objective is to obtain the probability that the default will just occur after a time t . For this we define the intensity $\lambda_{d,t}^{\mathbb{P}}$ under the physical measure.

Definition C.1. The default intensity $\lambda_{d,t}^{\mathbb{P}}$ is the instantaneous \mathbb{P} -probability of default in the next instant of time considering that default has not occurred until t .

$$\lambda_{d,t}^{\mathbb{P}} = \lim_{h \rightarrow 0} \frac{\mathbb{P}(\tau_d < t + h | \tau_d > t)}{h}$$

With this definition we can find the probability that default occurs after time t .

Proposition C.2. *The \mathbb{P} -probability that defaults occur after t is given by:*

$$\mathbb{P}(\tau_d > t) = e^{-\int_0^t \lambda_{d,u}^{\mathbb{P}} du} \quad (\text{C.1})$$

Proof. Using Bayes' theorem:

$$\begin{aligned} \lambda_{d,t}^{\mathbb{P}} &= \lim_{h \rightarrow 0} \frac{\mathbb{P}(\tau_d < t+h | \tau_d > t)}{h} = \lim_{h \rightarrow 0} \frac{\mathbb{P}(t < \tau_d < t+h)}{\mathbb{P}(\tau_d > t)h} \\ &= \lim_{h \rightarrow 0} \frac{\mathbb{P}(\tau_d > t) - \mathbb{P}(\tau_d > t+h)}{\mathbb{P}(\tau_d > t)h} = \lim_{h \rightarrow 0} \frac{\mathbb{P}(\tau_d > t) - \mathbb{P}(\tau_d > t+h)}{\mathbb{P}(\tau_d > t)h} \\ &= -\frac{1}{\mathbb{P}(\tau_d > t)} \frac{d}{dt} \mathbb{P}(\tau_d > t) \end{aligned}$$

Therefore we have the first order differential equation:

$$\frac{d}{dt} \mathbb{P}(\tau_d > t) = -\lambda_{d,t}^{\mathbb{P}} \mathbb{P}(\tau_d > t)$$

The condition that default as not occurred at time 0 is imposed and is the same condition used when considering a bond emission at $t=0$:

$$\mathbb{P}(\tau_d > 0) = 1$$

From these expressions we get equation (C.1) □

As a consequence of the definition of the intensity, and using a Taylor expansion in equation (C.1), for infinitesimal time intervals Δt the probability of default in this interval is $\lambda_{d,t}^{\mathbb{P}} \Delta t$, making the probability of default directly related with the size of the time interval and with the constant of proportionality given by the default intensity.

We call $N_{d,t}$ a Poisson process because (C.1) could be obtained by constructing a model via number of occurrences where the increments of the process are Poisson distributed.

$$\mathbb{P}(N_t - N_s = k) = \frac{1}{k!} \left(\int_s^t \lambda(u) du \right)^k e^{-\int_s^t \lambda(u) du} \implies \mathbb{P}(N_t = 0) = \mathbb{P}(\tau_d > t) = e^{-\int_0^t \lambda(u) du}$$

When $\lambda_{d,t}^{\mathbb{P}} \equiv \lambda_d^{\mathbb{P}} > 0$ the counting process $N_{d,t}$ is said to be a *Standard Poisson Process* and when we have $\lambda_{d,t} \equiv \lambda_d^{\mathbb{P}}(t) > 0, \forall t$, $N_{d,t}$ is an *Inhomogeneous Poisson Process*.

C.2 Default as a Cox Process

We can now generalize the results of the previous section to the case of a random intensity $\lambda_{d,t}^{\mathbb{P}} \equiv \lambda_d^{\mathbb{P}}(Y_t)$, where Y_t is a d -dimensional stochastic process. We fix a probability space $(\Omega, \mathcal{F}_t, \mathbb{P})$, where $\mathcal{F}_t = \mathcal{G}_t \vee \mathcal{H}_t$ satisfies the usual conditions and the filtrations are defined as:

$\mathcal{G}_t = \sigma\{Y_s : 0 \leq s \leq t\}$, the information of the evolution of the state vector Y_t up to time t , i.e, the filtration generated by the state vector.

$\mathcal{H}_t = \sigma\{\mathbf{1}_{\{\tau_d \leq s\}} : 0 \leq s \leq t\}$, the information of the existence of default up to time t , i.e, the filtration generated by the counting process $N_{d,t}$.

We are now able to define a Cox process as a generalization of the Poisson processes:

Definition C.3. The counting process $N_{d,t}$ with $N_{d,0} = 0$ is a Cox process with a random, \mathcal{G}_t -adapted, intensity $\lambda_d^{\mathbb{P}}(Y_t)$ (that needs to satisfy $\int_0^t \lambda_{d,u}^{\mathbb{P}} du < \infty$), if conditioned on a particular realization of $\lambda_{d,t}^{\mathbb{P}}$, i.e, knowing \mathcal{G}_t , $N_{d,t}$ is an inhomogeneous Poisson process given that default has not occurred yet.

$$\mathbb{P}(\tau_d > t | \mathcal{G}_t \vee \mathcal{H}_0) = e^{-\int_0^t \lambda_{d,u}^{\mathbb{P}} du}$$

This definition enables us to use the result of Proposition C.2 and get the survival probability in the Cox process.

Proposition C.4. *The \mathbb{P} -probability at time t that defaults occurs after time T in the Cox process is given by:*

$$\mathbb{P}(\tau_d > T | \mathcal{F}_t) = \mathbf{1}_{\{\tau_d > t\}} E_{\mathbb{P}}[e^{-\int_t^T \lambda_{d,u}^{\mathbb{P}} du} | \mathcal{F}_t] \quad (\text{C.2})$$

Proof. We now have two cases because if default had occurred we would not have the process defined through an intensity, and the probability would be simply zero.

Using the law of iterated expectation:

$$\mathbb{P}(\tau_d > T) = \mathbf{1}_{\{\tau_d \leq t\}} * 0 + \mathbf{1}_{\{\tau_d > t\}} E_{\mathbb{P}}[\mathbf{1}_{\{\tau_d > T\}} | \mathcal{F}_t] = \mathbf{1}_{\{\tau_d > t\}} E_{\mathbb{P}}[E_{\mathbb{P}}[\mathbf{1}_{\{\tau_d > T\}} | \mathcal{G}_T \vee \mathcal{H}_t] | \mathcal{F}_t]$$

With definition C.3 and proposition C.2 we get equation (C.2) □

Equation (C.2) is the same we get when dealing with term structure models, and it is known how to calculate the expected value for several dynamics of the default intensity, as shown in Appendix D.

Appendix D

Affine Term Structure Models

The models used to evaluate defaultable bonds and defaultable bonds with call provisions are implemented in this thesis considering the Affine Term Structure Models (ATSM's). These models have implied no-arbitrage restrictions common to every term structure model, and have mathematical and computational tractability. They are called affine because the yields of zero bonds for any maturity are affine functions of the state vector. For some of them there are closed-form solutions for the discount factor.

As usual, we consider the vector of independent standard Brownian motion W , the σ -field \mathcal{F}_t , satisfying the usual conditions, generated by W in the interval $[0, T]$, and the probability space $(\Omega, \mathcal{F}_t, \mathbb{P})$ that characterizes the economical uncertainty.

The more general construction of ATSM's is based on some simple assumptions:

Assumption 1. The short rate r_t is an affine function of the state vector $Y(t)$:

$$r_t = \delta_0 + \sum_{i=1}^N \delta_i Y_i(t) \quad (\text{D.1})$$

where $\delta_i \in \mathbb{R}, \forall i$.

Now we need to specify the dynamics of the state vector. For now we are just concerned with the dynamics in the pricing measure \mathbb{Q} .

Assumption 2. The dynamics of each $Y(t)$ is an Itô diffusion under the pricing measure \mathbb{Q} :

$$dY(t) = \mu^{\mathbb{Q}}(Y_t)dt + \sigma(Y_t)dW_t^{\mathbb{Q}} \quad (\text{D.2})$$

For now we do not specify how to change measure but according to the Girsanov theorem changing measure will only change the drift of the last equation. Hence, under the physical measure \mathbb{P} .

$$dY(t) = \mu^{\mathbb{P}}(Y_t)dt + \sigma(Y_t)dW_t^{\mathbb{P}} \quad (\text{D.3})$$

The last assumption we make is about the form that the drift and diffusion functions can take.

Assumption 3. Under the pricing measure the drift and diffusion functions are affine in the N risk factors

$$\mu^{\mathbb{Q}}(Y_t) = \tilde{K}(\tilde{\Theta} - Y(t))dt \quad (\text{D.4})$$

$$\sigma(Y_t)\sigma^T(Y_t) = g_0 + \sum_{i=1}^N g_i Y_i(t) \quad (\text{D.5})$$

In this way we can write the Itô diffusion in the pricing measure as:

$$dY(t) = \tilde{K}(\tilde{\Theta} - Y(t))dt + \Sigma\sqrt{S(t)}dW^{\mathbb{Q}}(t) \quad (\text{D.6})$$

Where $Y(t)$ is the N -vector of the state variables, \tilde{K} and Σ are $N \times N$ matrices and $S_{ii}(t) = \alpha_i + \beta'_i Y(t)$, being α a N -vector and β_i the line i of an $N \times N$ matrix \mathcal{B} .

Duffie and Kan [14] found a solution for the discount factor in this general ATSM framework.

Proposition D.1. *Considering assumptions 1,2 and 3, the discount factor between t and T , $P(t, T) = E_{\mathbb{Q}}[e^{-\int_t^T r_u du} | \mathcal{F}_t]$, has the solution:*

$$P(t, T) = e^{A(\tau) - B'(\tau)Y(t)} \quad (\text{D.7})$$

Where $A(\tau)$ and $B(\tau)$ satisfy the differential equations:

$$\frac{dA(\tau)}{d\tau} = -\tilde{\theta}\tilde{K}B(\tau) + \frac{1}{2} \sum_{i=1}^N [\Sigma' B(\tau)]_i^2 \alpha_i - \delta_0 \quad (\text{D.8})$$

$$\frac{dB(\tau)}{d\tau} = -\tilde{K}B(\tau) - \frac{1}{2} \sum_{i=1}^N [\Sigma' B(\tau)]_i^2 \beta'_i - \delta_y \quad (\text{D.9})$$

Therefore, the discount factors can be obtained by solving the Riccati system of equations for $B(\tau)$ and afterwards for $A(\tau)$.

It remains to solve the problems of changing between measures. The pricing measure \mathbb{Q} can be defined from the Radon-Nikodym derivative:

$$\frac{d\mathbb{Q}}{d\mathbb{P}} \Big|_{\mathcal{F}_t} = e^{-\int_0^t \Lambda(Y_u) dW_u - \frac{1}{2} \int_0^t \Lambda^T(Y_u) \Lambda(Y_u) du} \quad (\text{D.10})$$

The market price of risk $\Lambda(Y_u)$, from the Girsanov theorem, must have a functional form that makes the Radon-Nykodim derivative a martingale, i.e:¹

$$\mathbb{E}_{\mathbb{P}} \left[e^{-\int_0^t \Lambda(Y_u) dW_u - \frac{1}{2} \int_0^t \Lambda^T(Y_u) \Lambda(Y_u) du} \mid \mathcal{F}_t \right] = 1 \quad (\text{D.11})$$

This way, from the Girsanov theorem, we get the Brownian motion:

$$W_t^{\mathbb{Q}} = W_t^{\mathbb{P}} + \int_0^t \Lambda(Y_u) du \quad (\text{D.12})$$

To complete the construction of a model we need to specify the market price of risk Λ .

D.1 The Completely Affine Class of ATSM

First we use the functional form proposed by Dai and Singleton [7] where this market price of risk is proportional to the volatility level. It is the most simple form of the market price of risk that maintain an affine form for the diffusion in the physical measure. This gives us the first type of models, known as completely affine.

Definition D.2. The completely affine models are affine models with a market price of risk given by $\Lambda_t = \sqrt{S_t} \bar{\eta}$, where $\bar{\eta}$ is an N-vector of constants.

With this we have for equation (D.5):

$\tilde{\mathcal{K}} = \mathcal{K} + \Sigma \phi$, where the i^{th} line of ϕ is given by $\eta_i \beta'_i$ and $\tilde{\Theta} = K^{-1}(K\Theta - \Sigma\psi)$, where the i^{th} element of ψ is $\eta_i \alpha_i$.

Assumption 3 is therefore satisfied. Equation (D.11) will clearly also be satisfied if we have Gaussian models (what we will call the $\mathcal{A}_0(N)$ models) because the Novikov condition holds in this case. In the case of stochastic volatility models is not straightforward to check the validity of equation (D.11) because the Novikov condition does not hold, but results from Karatzas and Shreve [23] in fact guarantee satisfaction of (D.11) for all cases.

The study of Dai and Singleton [7] has as a starting point the verification of the conditions over the parameters that satisfy the admissibility criterion: $\alpha_i + \beta'_i Y(t) \geq 0$. That is, the volatility of each factor Y_i must be well defined. They group the models with N variables as $\mathcal{A}_m(N)$ where m represents the number of variables driving each S_{ij} . This way admissibility is a consequence of having all of the m variables being

¹Because $\frac{d\mathbb{Q}}{d\mathbb{P}} \mid \mathcal{F}_0 = 1$

non-negative. This implies restrictions over the parameters and a trade-off between flexibility on the drift and diffusion parameters. They also consider the econometric under-identification of the model and introduce other constraints on the model parameters imposing a parametrization that guarantees that the model is just-identified (Quoting Dufresne, Goldstein and Jones [6]: *A model is said to be identified if the state vector and parameter vector can be inferred from a particular data set*). They obtain a canonical form that is consequence of admissibility and just-identification of the model.

Definition D.3. The Dai and Singleton [7] canonical form of a completely affine $\mathcal{A}_m(N)$ model with $Y_t' = (Y^{B'}, Y^{D'})$, where Y^B is $m \times 1$ and Y^D is $(N - m) \times 1$ has the elements of (D.3) given by:

$$\mathcal{K} = \begin{bmatrix} \mathcal{K}_{m \times m}^{BB} & 0_{(N-m) \times m} \\ \mathcal{K}_{(N-m) \times m}^{DB} & \mathcal{K}_{(N-m) \times (N-m)}^{DD} \end{bmatrix} \quad \Theta = \begin{bmatrix} \Theta_{m \times 1}^B \\ 0_{(N-m) \times 1} \end{bmatrix} \quad \Sigma = I_{N \times N} \quad (\text{D.13})$$

$$\mathcal{B} = \begin{bmatrix} I_{m \times m} & B_{(N-m) \times m}^{DB} \\ 0_{(N-m) \times m} & 0_{(N-m) \times (N-m)} \end{bmatrix} \quad \alpha = \begin{bmatrix} 0_{m \times 1} \\ 1_{(N-m) \times 1} \end{bmatrix} \quad (\text{D.14})$$

With the restrictions:

$$\delta_i \geq 0, i > m \quad \mathcal{K}_i \Theta > 0, i \leq m \quad K_{ij} \leq 0, j \leq m, j \neq i \quad (\text{D.15})$$

$$\Theta_i \geq 0 \text{ for } i \leq m \quad B_{ij} \geq 0, i \leq m, m + 1 \leq j \leq N \quad (\text{D.16})$$

Example D.1. *The most general $\mathcal{A}_4(4)$ model satisfying admissibility and in the parametrization of Dai and Singleton [7] has the parameters of (D.3) given by:*

$$\mathcal{K} = \begin{bmatrix} k_{11} & k_{12} & k_{13} & k_{14} \\ k_{21} & k_{22} & k_{23} & k_{24} \\ k_{31} & k_{32} & k_{33} & k_{34} \\ k_{41} & k_{42} & k_{43} & k_{44} \end{bmatrix} \quad \Sigma = \mathcal{B} = \mathbb{I}_{4 \times 4} \quad \alpha = 0_{4 \times 4} \quad \Theta' = (\theta_1, \theta_2, \theta_3, \theta_4)$$

Or in a more explicit way:

$$\begin{cases} dY_1 &= [k_{11}(\theta_1 - Y_1) + k_{12}(\theta_2 - Y_2) + k_{13}(\theta_3 - Y_3) + k_{14}(\theta_4 - Y_4)] dt + \sqrt{Y_1} dW_1^{\mathbb{P}} \\ dY_2 &= [k_{21}(\theta_1 - Y_1) + k_{22}(\theta_2 - Y_2) + k_{23}(\theta_3 - Y_3) + k_{24}(\theta_4 - Y_4)] dt + \sqrt{Y_2} dW_2^{\mathbb{P}} \\ dY_3 &= [k_{31}(\theta_1 - Y_1) + k_{32}(\theta_2 - Y_2) + k_{33}(\theta_3 - Y_3) + k_{34}(\theta_4 - Y_4)] dt + \sqrt{Y_3} dW_3^{\mathbb{P}} \\ dY_4 &= [k_{41}(\theta_1 - Y_1) + k_{42}(\theta_2 - Y_2) + k_{43}(\theta_3 - Y_3) + k_{44}(\theta_4 - Y_4)] dt + \sqrt{Y_4} dW_4^{\mathbb{P}} \end{cases}$$

It is nonetheless possible to have another representation for these models and not only the canonical representation. That is possible by means of an invariant transformations, i.e, a transformation that leaves r_t unchanged. An example of an alternative representation used to model the short-rate is given in Mosburger and Schneider [33] where they diagonalize \mathcal{K} , and Σ has now a more complex structure.

Example D.2. In Mosburger and Schneider [33] instead of (D.10) they use:

$$\mathcal{K} = \begin{bmatrix} \mathcal{K}_{m \times m}^{BB} & 0_{m \times (N-M)} \\ 0_{(N-m) \times m} & \mathcal{K}_{(N-m) \times (N-m)}^{DD} \end{bmatrix} \quad \Sigma = \begin{bmatrix} I_{m \times m} & 0_{m \times (N-M)} \\ \Sigma_{(N-m) \times m}^{DB} & \Sigma_{(N-m) \times (N-m)}^{DD} \end{bmatrix}$$

Moving flexibility from \mathcal{K} to Σ , changing the dynamics of the state variables if $N \neq m$.

It is also important to notice that the well known N-factor CIR model is a restricted example of the completely affine class of ATSM's. It's an $\mathcal{A}_N(N)$ model where not only the matrix \mathcal{K} is diagonal, but where we also have usually the Feller conditions imposed. This does not happen in the canonical form of Dai and Singleton where the zero boundary can be reached.

There are some problems with this model, because it cannot explain some empirical facts about the term structure of interest rates. In particular, it does not match the observed excess returns on bonds.

D.1.1 Excess Returns

Excess returns on bonds are return above the short-rate. To have their explicit form we write the dynamics of the discount factor as:

$$\frac{dP(t, T)}{P(t, T)} = \alpha(t, T) dt + \vartheta(t, T) dW_t^{\mathbb{P}} \quad (\text{D.17})$$

To obtain the explicit form of $\alpha(t, T)$ and $\vartheta(t, T)$ we first consider the general case where the discount factor is a function of time and the state vector:

$$P(t, T) = f(t, T, Y_t) \quad (\text{D.18})$$

Applying Itô lemma to this equation and considering (D.3) we get:

$$\alpha(t, T) = \frac{1}{P(t, T)} \left[\frac{\partial P(t, T)}{\partial t} + \mu^{\mathbb{P}}(Y_t) \frac{\partial P(t, T)}{\partial Y_t} + \frac{1}{2} \sigma(Y_t) \frac{\partial^2 P(t, T)}{\partial Y_t^2} \right]$$

and

$$\vartheta(t, T) = \frac{\sigma(Y_t)}{P(t, T)} \frac{\partial P(t, T)}{\partial Y_t}$$

The functional form of the excess return will depend on the choice of the market price of risk, as a result of the following proposition:

Proposition D.4. *Supposing that there is non arbitrage opportunities in the market, the market price of risk must satisfy the following equation:*

$$\alpha(t, T) = \Lambda_t \vartheta(t, T) + r_t \quad (\text{D.19})$$

Because the first term is a return above the short rate, we call it the excess return $ER(t, T)$. With proposition (D.1) this excess return can be written as:

$$ER(t, T) = -B(\tau) \Sigma \sqrt{S_t} \Lambda_t \quad (\text{D.20})$$

In particular, with the model of Dai and Singleton [7] :

$$ER(t, T) = -B(\tau) \Sigma S_t \bar{\eta} \quad (\text{D.21})$$

This means that excess returns, in the completely affine models, depend on the maturity of the bonds and vary in time because the volatility term is not constant² and it does not depend directly on the state vector. This contradicts the observations, what leads to the essential affine models.

²It is in the Gaussian models, but it is another observational fact that the volatility of yields is time varying.

D.2 The Essentially Affine Class of ATSM

The essentially affine models have been proposed by Dufee [12] and try to overcome the problems of excess returns observed in the completely affine models giving a more general formula for the market price of risk.

Definition D.5. The essentially affine models are affine models with a market price of risk given by $\Lambda_t = \sqrt{S_t^-} \bar{\eta} + \sqrt{S_t^-} \eta_2 Y_t$, where $\bar{\eta}$ is an N -vector of constants, S_t is the same matrix of (D.2), η_2 is an $N \times N$ matrix and:

$$S_{ii,t}^- = \begin{cases} (\alpha_i + \beta_i' Y_t)^{-1}, & \text{inf}(\alpha_i + \beta_i' Y_t) > 0 \\ 0 & \text{else} \end{cases}$$

With this we are able to write:

$$dY(t) = \tilde{\mathcal{K}}_+(\tilde{\Theta}_+ - Y(t))dt + \Sigma \sqrt{S(t)} dW^{\mathbb{Q}}(t) \quad (\text{D.22})$$

$$\tilde{\mathcal{K}}_+ = \tilde{\mathcal{K}}(\mathbb{I} + \Sigma I^- \eta_2) \quad \tilde{\Theta}_+ = (\mathbb{I} + \Sigma I^- \eta_2)^{-1} \tilde{\Theta} \quad \text{where } I^- = \begin{cases} \mathbb{I}, & S_t^- \neq 0 \\ 0, & S_t^- = 0 \end{cases} \quad (\text{D.23})$$

The condition in $S_{ii,t}^-$ avoids an infinite market price of risk for zero risk (arbitrage). Furthermore, because the canonical form of Dai and Singleton has $\alpha_i = 0$, $i \leq m$ and the non-Gaussian variables (those that do not enter in each S_{ii}) are always non-negative, in the case $m = N$ every element S_t^- is zero and therefore the market price of risk is the same of the completely affine models. Only for $m \neq N$ the essentially affine models are more general than the completely affine. Proposition (D.1) still holds but with the obvious changes: $\tilde{\mathcal{K}} \rightarrow \tilde{\mathcal{K}}_+$ and $\tilde{\Theta} \rightarrow \tilde{\Theta}_+$. This way we still have assumption 3 and excess returns depend directly on the state vector.

Other more general specifications of the market price of risk are possible, and we still maintain the tractability of the affine models. Feldhütter [18] found, using US Treasury data, that the market price of risk that better fit the observations is the one due to Duarte, with a market price of risk similar to Dufee's [12], where a constant is introduced and makes possible a change in the sign of the excess returns:

$$\Lambda_t = \Sigma^{-1} \eta_0 + \sqrt{S_t^-} \bar{\eta} + \sqrt{S_t^-} \eta_2 Y_t$$

Appendix E

Kimmel Series Expansion

In Appendix D we had the assumption that the short-rate is an affine function of the state variables. Together with other assumptions this leads to a closed form solution for the discount factor. In general we do not know closed form solutions when the short-rate is a non-linear function of the state vector. The objective here is to find an approximation at least for some cases where non-linearity exists.

We are interested in studying N-dimensional diffusions:

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

This usually leads, in pricing problems,¹ to an expected value that depends on this diffusion:

$$f(\tau, x) = E[e^{-\int_t^{t+\tau} r(X_u)du} \psi(X_{t+\tau}) | X_t = x] \quad (\text{E.1})$$

This expected value has a known result for several models, most of them being affine.

By the Feynman-Kac formula we know that this function is the solution to the differential equation:

$$\frac{\partial f}{\partial \tau}(\tau, x) = \sum_{i=1}^N \mu_i(x) \frac{\partial f}{\partial x_i}(\tau, x) + \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \sigma_{ij}^2(x) \frac{\partial^2 f}{\partial x_i \partial x_j}(\tau, x) - r(x)f(\tau, x)$$

Kimmel [24] writes the solution for this differential equation as a power series in τ centred at zero.

¹The discounted value of future expected pay-off's.

$$f(\tau, x) = \sum_{n=0}^{\infty} a_n(x) \frac{\tau^n}{n!}$$

The coefficients are obtained after substituting this expression in the differential equation above:

$$a_0(x) = \psi(x)$$

$$a_n(x) = \sum_{i=1}^N \mu_i(x) \frac{\partial a_{n-1}}{\partial x_i}(x) + \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \sigma_{ij}^2(x) \frac{\partial^2 a_{n-1}}{\partial x_i \partial x_j}(x) - r(x) a_{n-1}(x)$$

But can the function $f(\tau, x)$ be represented by a power series? The problem of finding the expected value in the case where τ represents the maturity of the bond only makes sense for positive values of τ , but we must consider negative and complex values because these complex values determine the region of convergence of the power series.

Definition E.1. A function $f : \Omega \rightarrow \mathbb{C}$ is said to be analytic around a point a if $f(z)$ can be expanded in a power series around a .

$$f(z) = \sum_{n=0}^{\infty} c_n (z - a)^n$$

The necessary and sufficient conditions for a function to be analytic are given by the Cauchy-Riemann equations, but for that we need to know the function $f(z)$ and that is the objective.

And even if the function is analytic and, therefore, has a convergent power series representation, this convergence can be slow for large real values of τ . We are also interested in a convergence independent of the value of τ .

Definition E.2. We say that a series of functions $f_n : \Omega \rightarrow \mathbb{C}$ converges uniformly to a function $f : \Omega \rightarrow \mathbb{C}$ if for every $\epsilon > 0$ there is an $N = N(\epsilon)$ such that for all $n > N$, $|f_n(z) - f(z)| < \epsilon$, for all $z \in \Omega$

This way N depends on ϵ but not on z . Kimmel argues that particular types of pricing problems guarantee an analytic (in τ) function represented by a uniformly convergent power series.

E.1 Power Series Solution to the Pricing Problem

The objective in this section is to find a power series approximation for the expectation (E.1) when $r(X_u)$ is a non-linear function of X_u . To do that we start by transforming the differential equation solved by the function $f(\tau, x)$ to the 'canonical form' — a simpler differential equation. Writing:

$$f(\tau, x) = \eta(x)h(\tau, y(x))$$

By correctly choosing $\eta(x)$ and $y(x)$ we have the transformed problem:

$$\frac{\partial h(\tau, y)}{\partial \tau} = \frac{1}{2} \frac{\partial^2 h(\tau, y)}{\partial y^2} - r_h(y)h(\tau, y) \quad (\text{E.2})$$

$$h(0, y) = g(y) \quad (\text{E.3})$$

Considering the differential equation:

$$\frac{\partial f}{\partial \tau}(\tau, x) = \mu(x) \frac{\partial f}{\partial x_i}(\tau, x) + \frac{\sigma(x)^2}{2} \frac{\partial^2 f}{\partial x_i^2}(\tau, x) - r(x)f(\tau, x)$$

The transformation needed to have the canonical problem is:

$$y = \int^x \frac{du}{\sigma(u)} \quad f(\tau, x) = e^{-\int^x \left[\frac{\mu(u)}{\sigma^2(u)} - \frac{\sigma'(u)}{2\sigma(u)} \right] du} h(\tau, y)$$

We analyse one particular case of the two considered in Kimmel [24]. This is the one that we need to obtain the results of Chapter 3.

$$r_h(y) = \frac{a}{y^2} + \frac{b^2}{2}y^2 + d \quad (\text{E.4})$$

We also use the time transformation:

$$\Delta = \Delta_k(\tau) \equiv 1 - e^{-k\tau}$$

for some $k \neq 0$. Expressing $h(\tau, y)$ as:

$$h(\tau, y) = \nu(\tau, y)w(\Delta(\tau), z(\tau, y))$$

with:

$$\nu(\tau, y) = e^{\lambda\tau}\xi(y) \quad z(\tau, y) = \sqrt{k} \left[\theta + e^{-\frac{k\tau}{2}}(y - \theta) \right]$$

for arbitrary λ , θ and $\xi(y)$.

Kimmel finds a result that guarantees uniform convergence of a power series approximation of $h(\tau, y)$ with just some restrictions over λ and k .

Theorem E.3. *Suppose the solution to $h(\tau, y)$ can be written as:*

$$h(\tau, y) = e^{\lambda\tau}\xi(y)w(\Delta(\tau), z(\tau, y))$$

for some $k \neq 0$, arbitrary λ , θ and $\xi(y)$, where $w(\Delta, z)$ is an analytic function of both variables for all z and for all $|\Delta| < r$ for some $r > 1$. Denote by $w_n(\Delta, z)$ the power series approximation in Δ to $w(\Delta, z)$ including terms up to order n in Δ .

$$w_n(\Delta, z) \equiv \sum_{i=0}^n b_i(z)\Delta^i$$

Define:

$$h_n(\tau, y) = e^{\lambda\tau}\xi(y)w_n(\Delta(\tau), z(\tau, y))$$

Then for a fixed value of y , $h_n(\tau, y)$ converges to $h(\tau, y)$ for all complex τ such that:

$$\operatorname{Re}(k\tau) > -\ln(\cos[\operatorname{Im}(k\tau)] + \sqrt{\cos^2[\operatorname{Im}(k\tau)] + r^2 - 1})$$

Furthermore, for any $1 \leq s < r$ and for any real c , $h_n(\tau, y)$ converges uniformly to $h(\tau, y)$ for all complex τ such that:

$$\operatorname{Re}(k\tau) \geq -\ln(\cos[\operatorname{Im}(k\tau)] + \sqrt{\cos^2[\operatorname{Im}(k\tau)] + s^2 - 1}) \quad \text{and} \quad \operatorname{Re}(\lambda\tau) \leq c$$

If λ is negative and k is positive and $w(\tau, z)$ is analytic in $|\Delta| < r$ for some $r > 1$ and considering real values of τ , then we have uniform convergence in $\tau \in [0, +\infty]$, i.e, for all relevant (positive) maturities. But if k and λ are positive we can construct the interval of convergence $[0, T]$ of τ for any $T > 0$ by just choosing $c = \lambda T$.

We now need to guarantee analyticity for $w(\tau, z)$.

Theorem E.4. *Let $b \neq 0$, a and d be arbitrary numbers and for $\sqrt{2b}$ choose either square root. Let $g_1(y)$ and $g_2(y)$ be even functions that are analytic for all complex y , and let there exist some $c > 0$ and some norm (over the reals) $\|y\|$ such that $g_1(y)$ and $g_2(y)$ satisfy the bounds*

$$\left| e^{\frac{y^2}{4}} g_1\left(\frac{y}{\sqrt{2b}}\right) \right| \leq ce^{\frac{\|y\|^2}{2}} \quad \left| e^{\frac{y^2}{4}} g_2\left(\frac{y}{\sqrt{2b}}\right) \right| \leq ce^{\frac{\|y\|^2}{2}}$$

Then there exist $w_1(\Delta, z)$ and $w_2(\Delta, z)$ analytic for all complex z and $\|\sqrt{\Delta}\|$, that satisfy the partial differential equations with final conditions:

$$\frac{\partial w_1(\Delta, z)}{\partial \Delta} = \frac{1 - \sqrt{1+8a}}{2z} \frac{\partial w_1(\Delta, z)}{\partial z} + \frac{1}{2} \frac{\partial^2 w_1(\Delta, z)}{\partial z^2}$$

$$\frac{\partial w_2(\Delta, z)}{\partial \Delta} = \frac{1 + \sqrt{1+8a}}{2z} \frac{\partial w_2(\Delta, z)}{\partial z} + \frac{1}{2} \frac{\partial^2 w_2(\Delta, z)}{\partial z^2}$$

$$w_1(0, z) = e^{\frac{z^2}{4}} g_1\left(\frac{y}{\sqrt{2b}}\right) \quad w_2(0, z) = e^{\frac{z^2}{4}} g_2\left(\frac{y}{\sqrt{2b}}\right)$$

Furthermore, $h(\tau, y)$, defined by:

$$h(\tau, y) \equiv e^{-\frac{b}{2}y^2 - (\frac{b}{2}+d)\tau} \left[\left(\frac{z}{\sqrt{2b}}\right)^{\frac{1-\sqrt{1+8a}}{2}} w_1(\Delta, z) + \left(\frac{z}{\sqrt{2b}}\right)^{\frac{1+\sqrt{1+8a}}{2}} w_2(\Delta, z) \right]$$

where $\Delta = \Delta_{2b}(\tau)$ and $z = \sqrt{2b}e^{-b\tau}y$, satisfy (E.1), (E.2) and (E.3) with $g(y) = g_1(y)y^{\frac{1-\sqrt{1+8a}}{2}} + g_2(y)y^{\frac{1+\sqrt{1+8a}}{2}}$ for all complex y and τ such that $y \neq 0$ and $\|\sqrt{\Delta_{2b}(\tau)}\| < 1$.

It is now possible to apply theorem E.3 with

$$k = 2b \quad \lambda = -b \left(1 \pm \frac{\sqrt{1+8a}}{2}\right) - d \quad \theta = 0 \quad \xi(y) = e^{-\frac{b}{2}y^2} y^{1 \pm \frac{\sqrt{1+8a}}{2}}$$

The minus sign is used in the $w_1(\Delta, z)$ part and the plus sign in the $w_2(\Delta, z)$ part. This way we have $h(\tau, y)$ of theorem E.4 in the form of $h(\tau, y)$ in theorem E.3.

Example E.1. We consider a CIR dynamics for the state variable and an expectation that is a non linear in this state variable.

$$dX_t = k(\theta - x)dt + \sigma\sqrt{x}dW_t \quad \text{and} \quad f(\tau, x) = E[e^{-\int_t^{t+\tau}(c_1u + \frac{c_2}{u})du}\psi(X_{t+\tau})|X_t = x]$$

with $c_1, c_2 > 0$ and $\frac{2\theta k}{\sigma^2} \geq 1$ (the Feller condition).

That by Feynman-Kac leads to:

$$\frac{\partial f}{\partial \tau}(\tau, x) = k(\theta - x)\frac{\partial f}{\partial x_i}(\tau, x) + \frac{\sigma^2 x}{2}\frac{\partial^2 f}{\partial x_i^2}(\tau, x) - (c_1x + \frac{c_2}{x})f(\tau, x)$$

The transformations described above are:

$$y = \frac{2\sqrt{x}}{\sigma} \quad f(\tau, x) = \left(\frac{4x}{\sigma^2}\right)^{\frac{1}{4} - \frac{\theta k}{\sigma^2}} e^{\frac{xk}{2}} h(\tau, y)$$

This leads to the canonical form with the same specification of r_h as in theorem E.4.

$$\frac{\partial h(\tau, y)}{\partial \tau} = \frac{1}{2}\frac{\partial^2 h(\tau, y)}{\partial y^2} - \left[\frac{a}{y^2} + \frac{b^2}{2}y^2 + d\right] h(\tau, y)$$

$$h(0, y) = y^\alpha e^{-\frac{k}{4}y^2} = y^{\alpha-\gamma} e^{-\frac{k}{4}y^2} y^\gamma$$

The last step in the final condition is needed to get the final condition in the necessary form to apply Theorem E.4.

The constants in the last equations are:

$$a \equiv \frac{1}{2}\frac{4\theta^2 k^2}{\sigma^4} - \frac{2\theta k}{\sigma^2} + \frac{4c_2}{\sigma^2} + \frac{3}{8} \quad b \equiv \frac{\sqrt{k^2 + 2c_1\sigma^2}}{2} \quad d \equiv -\frac{\theta k}{\sigma^2} \quad \alpha \equiv \frac{2\theta k}{\sigma^2} - \frac{1}{2} \quad \gamma \equiv \frac{1 - \sqrt{1 + 8a}}{2}$$

We apply Theorem E.4 with:

$$g_1(y) \equiv e^{-\frac{k}{4}y^2} y^{\alpha-\gamma} \quad g_2(y) \equiv 0 \quad \|y\| \equiv |y| \left[\sqrt{\frac{1}{2} - \frac{k}{4b}} + \epsilon \right] + \left| \frac{y}{\sqrt{2b}} \right| \sqrt{\alpha - \gamma} \quad , \text{ for any } \epsilon > 0$$

Note that $\frac{k}{4b} \in]0, \frac{1}{2}[$.

These are not the conditions of Kimmel [24]. Somehow he states this wrongly but nonetheless reaches the right results. With his expressions for the norm, g_1 and g_2 we can not satisfy the conditions of Theorem E.4.

With these definitions both g_1 and g_2 are even and holomorphic. The norm chosen satisfies the norm axioms² because we are using the modulus, and the growth conditions are satisfied because we have $e^{\delta x^2} > x^\delta$ for both positive x and δ . In our case $\delta = \alpha - \gamma$ and this is positive due to having the Feller conditions satisfied and the minus sign in γ makes us choose $g_2 = 0$ instead of $g_1 = 0$ so that we can apply theorem E.4 directly.

Therefore there are w_1 and w_2 analytic satisfying Theorem E.4. We can now use the series expansion in the differential equations of Theorem E.4 but first we notice that w_2 is everywhere zero. This is true because $g_2 = 0$ and therefore the final condition for the differential equation w_2 is zero. Solving the differential equation by separation of variables and having to satisfy the initial condition we must have the trivial solution.

Applying $w_1(\Delta, z) \equiv \sum_{i=0}^{\infty} b_i(z) \Delta^i$ in the differential equation we find a recursive relation for the coefficients:

$$b_0(z) = \left(\frac{z}{\sqrt{2b}} \right)^{\alpha-\gamma} e^{\frac{z^2}{4} \left(1 - \frac{k}{4b}\right)}$$

$$b_{n+1}(z) = \frac{\gamma}{(n+1)z} \frac{\partial b_n}{\partial z} + \frac{1}{2(n+1)} \frac{\partial^2 b_n}{\partial z^2}$$

Up to second order:

$$w_1(\Delta, z) = \left(\frac{z}{\sqrt{2b}} \right)^{\alpha-\gamma} e^{\frac{z^2}{4} \left(1 - \frac{k}{4b}\right)} \Pi(\Delta) \quad (\text{E.5})$$

²The norm on the space vector Ξ is a function $\|\cdot\|$ that satisfies for all $a \in \mathbb{C}$ and $u, v \in \Xi$:

- $\|av\| = |a| \|v\|$
- $\|u + v\| \leq \|u\| + \|v\|$
- $\|v\| = 0 \Rightarrow v = 0$

Where:

$$\begin{aligned} \Pi(\Delta) = & 1 + \Delta \left[\frac{(\alpha - \gamma)(\alpha + \gamma - 1)}{2z^2} + \left(\frac{2\alpha + 1}{4} \right) \left(1 - \frac{k}{2b} \right) + \frac{z^2}{8} \left(1 - \frac{k}{2b} \right)^2 \right] \\ & + \frac{\Delta^2}{2} \left[\frac{(\alpha - \gamma)(\alpha - \gamma - 2)(\alpha + \gamma - 1)(\alpha + \gamma - 3)}{4z^4} + \frac{(2\alpha - 1)(\alpha - \gamma)(\alpha + \gamma - 1)}{4z^2} \left(1 - \frac{k}{2b} \right) \right] \\ & + \frac{\Delta^2}{2} \left[\left(\frac{(2\alpha + 3)(2\alpha + 1)}{16} + \frac{(\alpha - \gamma)(\alpha + \gamma - 1)}{8} \right) \left(1 - \frac{k}{2b} \right)^2 \right] \\ & + \frac{\Delta^2}{2} \left[\frac{(2\alpha + 3)z^2}{16} \left(1 - \frac{k}{2b} \right)^3 + \frac{z^4}{64} \left(1 - \frac{k}{2b} \right)^4 \right] \end{aligned}$$

As said before we can now apply theorem E.3 with:

$$k = 2b \quad \lambda = -b \left(1 - \frac{\sqrt{1+8a}}{2} \right) - d \quad \theta = 0 \quad \xi(y) = e^{-\frac{b}{2}y^2} y^{\frac{1-\sqrt{1+8a}}{2}}$$

Remembering that we have:

$$\Delta = 1 - e^{-k\Delta\tau} \quad k_{\Delta} = 2b \quad z = \sqrt{2b}e^{-b\tau}y$$

We start to work backwards in order to get the final result:

$$f(\tau, x) = \left(\frac{4x}{\sigma^2} \right)^{\frac{1}{4} - \frac{\theta k}{\sigma^2}} e^{\frac{xk}{2}} h(\tau, y)$$

Where:

$$h(\tau, y) = e^{-\frac{b}{2}y^2 - (\frac{b}{2} + d)\tau} \left(\frac{z}{\sqrt{2b}} \right)^{\alpha} e^{\frac{z^2}{4}(1 - \frac{k}{2b})} \Pi(\Delta)$$

So we have find a power series approximation for a non-linear model. In this particular case the results of Kimmel could be applied because the dynamics was a CIR model and the expectation was of the form:

$$\mathbb{E} \left[e^{-\int (ax + \frac{b}{x}) dt} \right]$$

With the Feller condition imposed to the state variable x .

Appendix F

Kalman Filter

We want to estimate the parameters of our model, but now the state variables are not directly observed. What is observed are the yields associated with bonds of different maturities or the value of coupons bonds. These observables are functions of our unobserved state variables and we assume that the data are observed with errors. The Kalman filter [29] is an algorithm that uses these data to estimate the values of the unobserved variables and estimate the parameters of the model. If the relation of the observables with the state variables is linear we have the linear Kalman filter that is a minimum variance estimator.

F.1 The Linear Kalman Filter

Following Bolder [1] we start with n observables z in N moments t_i . These observables are linked to the state variables y through the linear measurement equation with an additive noise v :

$$z_{t_i} = A + Hy_{t_i} + v_{t_i}$$

Where v_{t_i} is a white noise:

$$v_{t_i} \sim \mathcal{N}(0, R)$$

$$R = \begin{bmatrix} r_1^2 & 0 & \dots & 0 \\ 0 & r_2^2 & \dots & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & 0 & \dots & r_n^2 \end{bmatrix} = \mathbb{E}[v_{t_i}v_{t_i}^T]$$

So, we are assuming measurements with independent normal errors. For the case we are interested, these can be a result of yields being given by a mean of bid-ask values, rounding of values or due to non-synchronous observations.

The dynamics of the state variables follows the transition equation:

$$y_{t_i} = C + Fy_{t_{i-1}} + \epsilon_{t_i}$$

Where ϵ_{t_i} is also a white-noise.

$$\epsilon_{t_i} \sim \mathcal{N}(0, Q)$$

$$Q = \mathbb{E}[\epsilon_{t_i}\epsilon_{t_i}^T]$$

The errors of the two equations are assumed to be independent. We say that a particular model has associated the vector of parameters Θ . The objective is to discover the vector of parameters that better fit the data. At every moment we need the probability of the parameters given the observations.

$$f(z_1, z_2, \dots, z_n; \Theta)$$

From equation (B.3) of Appendix B we can use the prediction error decomposition:

$$\begin{aligned} f(z_1, z_2, \dots, z_n; \Theta) &= f(z_1, |z, \dots, z_n; \Theta) f(z_2, \dots, z_n; \Theta) \\ &= \prod_{t=1}^n f(z_t | z^{t-1}; \Theta) \end{aligned}$$

Where $z^0 = \emptyset$ and $z^{t-1} = (z_1, z_2, \dots, z_{t-1})$.

Using logarithms our objective is to maximize the function:

$$l(\Theta) = \sum_{t=1}^n \ln(f(z_t | z^{t-1}; \Theta))$$

To do that we just need to compute the likelihoods $f(z_t | z^{t-1}; \Theta)$.

First we define the error

$$\zeta_{t_i} = z_{t_i} - \mathbb{E} [z_{t_i} | \mathcal{F}_{t_{i-1}}]$$

So that we have:

$$f(\zeta_t; \Theta) = f(z_t | z^{t-1}; \Theta)$$

This ζ_{t_i} is the error made by making a prediction of z_{t_i} at time t_{i-1} .

The first moment of this error is given by:

$$\mathbb{E} [\zeta_{t_i} | \mathcal{F}_{t_{i-1}}] = 0$$

To write the second moment we first note that:

$$\zeta_{t_i} = H (y_{t_i} - \mathbb{E} [y_{t_i} | \mathcal{F}_{t_{i-1}}]) + \nu_{t_i}$$

And because of the Gaussian assumptions that we made:

$$\zeta_{t_i} \sim \mathcal{N} (0, VAR [\zeta_{t_i} | \mathcal{F}_{t_{i-1}}])$$

Where:

$$VAR [\zeta_{t_i} | \mathcal{F}_{t_{i-1}}] = HVAR [y_{t_i} | \mathcal{F}_{t_{i-1}}] H^T + R$$

With example (B.1) of appendix B and the last equations we know that:

$$\begin{aligned} f(z_{t_i} | z^{t_{i-1}}; \Theta) &= f(\zeta_{t_i}; \Theta) \\ &= \frac{1}{(2\pi)^{\frac{n}{2}} |HVAR [y_{t_i} | \mathcal{F}_{t_{i-1}}] H^T + R|^{\frac{1}{2}}} \exp \left[-\frac{1}{2} \zeta_{t_i}^T (HVAR [y_{t_i} | \mathcal{F}_{t_{i-1}}] H^T + R)^{-1} \zeta_{t_i} \right] \end{aligned}$$

And therefore the function to maximize is

$$l(\Theta) = -\frac{nN \ln(2\pi)}{2} - \frac{1}{2} \sum_{i=1}^N \left[\ln (|VAR [z_{t_i} | \mathcal{F}_{t_{i-1}}]|) + \zeta_{t_i}^T VAR [z_{t_i} | \mathcal{F}_{t_{i-1}}]^{-1} \zeta_{t_i} \right]$$

The Kalman algorithm gives us the necessary values to compute this function.

We start by initializing the state vector, i.e, we give the starting values $\mathbb{E}[y_1|\mathcal{F}_0]$ and $VAR[y_1|\mathcal{F}_0]$.

From the measurement equation with Gaussian errors we have immediately:

$$\mathbb{E}[z_{t_i}|\mathcal{F}_{t_{i-1}}] = A + H\mathbb{E}[y_{t_i}|\mathcal{F}_{t_{i-1}}]$$

For the first variance:

$$\begin{aligned} VAR[z_{t_i}|\mathcal{F}_{t_{i-1}}] &= \mathbb{E}[z_{t_i}z_{t_i}^T|\mathcal{F}_{t_{i-1}}] - \mathbb{E}[z_{t_i}|\mathcal{F}_{t_{i-1}}]\mathbb{E}[z_{t_i}^T|\mathcal{F}_{t_{i-1}}] \\ &= H\mathbb{E}[y_{t_i}y_{t_i}^T|\mathcal{F}_{t_{i-1}}]H^T - \mathbb{E}[y_{t_i}|\mathcal{F}_{t_{i-1}}]\mathbb{E}[y_{t_i}^T|\mathcal{F}_{t_{i-1}}] + \mathbb{E}[v_{t_i}v_{t_i}^T|\mathcal{F}_{t_{i-1}}] \\ &= HVAR[y_{t_i}|\mathcal{F}_{t_{i-1}}]H^T + R \end{aligned}$$

With this we already have the first error and variance, and we notice that this expression equals the variance of ζ_{t_i} . To have the others we now need to update our expectations about the state vector. At this moment we write the expected value of the unobservable state variable as a linear combination of the prior estimate plus an weighted error:

$$\mathbb{E}[y_{t_i}|\mathcal{F}_{t_i}] = \mathbb{E}[y_{t_i}|\mathcal{F}_{t_{i-1}}] + K_{t_i}\zeta_{t_i}$$

Where K_{t_i} is the Kalman gain matrix.

With some algebra we can update the variance of the state variables that are needed for the likelihoods. Using the equation with the Kalman gain and writing the error in terms of the state variables:

$$\begin{aligned} VAR[y_{t_i}|\mathcal{F}_{t_i}] &= \mathbb{E}\left[\left(y_{t_i} - \mathbb{E}[y_{t_i}|\mathcal{F}_{t_{i-1}}]\right)\left(y_{t_i} - \mathbb{E}[y_{t_i}|\mathcal{F}_{t_{i-1}}]\right)^T|\mathcal{F}_{t_{i-1}}\right] \\ &= VAR[y_{t_i}|\mathcal{F}_{t_{i-1}}] - K_{t_i}HVAR[y_{t_i}|\mathcal{F}_{t_{i-1}}] \\ &\quad - VAR[y_{t_i}|\mathcal{F}_{t_{i-1}}]H^TK_{t_i}^T + K_{t_i}VAR[z_{t_i}|\mathcal{F}_{t_{i-1}}]K_{t_i}^T \end{aligned}$$

This is called the Joseph form. Now we have to construct a particular Kalman gain matrix. We build this matrix considering a minimum variance estimator, i.e, we want to minimize the variance of each variable — the trace of the matrix $VAR[y_{t_i}|\mathcal{F}_{t_i}]$.

To do this we notice that the covariance matrices are symmetrical and we have the results from matrix calculus:

$$\frac{\partial \text{tr}(XAB)}{\partial X} = B^T A^T \quad \frac{\partial \text{tr}(AB^T X^T)}{\partial X} = AB^T \quad \frac{\partial \text{tr}(XAX^T)}{\partial X} = XA^T + XA$$

The first order condition for a minimum value:

$$\frac{\partial \text{tr}(\text{VAR}[y_{t_i} | \mathcal{F}_{t_i}])}{\partial K_{t_i}} = -2\text{VAR}[y_{t_i} | \mathcal{F}_{t_{i-1}}] H^T + 2K_{t_i} \text{VAR}[z_{t_i} | \mathcal{F}_{t_{i-1}}] = 0$$

Implies:

$$K_{t_i} = \text{VAR}[y_{t_i} | \mathcal{F}_{t_{i-1}}] H^T \text{VAR}[z_{t_i} | \mathcal{F}_{t_{i-1}}]^{-1}$$

In the Joseph form we can use this last result to rewrite the last term and be able to write:

$$\text{VAR}[y_{t_i} | \mathcal{F}_{t_i}] = (\mathbb{I} - K_{t_i} H) \text{VAR}[y_{t_i} | \mathcal{F}_{t_{i-1}}]$$

With these we are in conditions to forecast the values of the state variables for the next period and use the result to forecast the value of the observable, construct the error and the next likelihood. The following two equations are similar to those of the observables:

$$\mathbb{E}[y_{t_{i+1}} | \mathcal{F}_{t_i}] = C + F \mathbb{E}[y_{t_i} | \mathcal{F}_{t_i}]$$

$$\text{VAR}[y_{t_{i+1}} | \mathcal{F}_{t_i}] = F \text{VAR}[y_{t_i} | \mathcal{F}_{t_i}] F^T + Q$$

Summarizing we have the algorithm:

1. Initialize the state vector:

$$\mathbb{E} [y_1 | \mathcal{F}_0] \quad VAR [y_1 | \mathcal{F}_0]$$

2. Use the observables to construct at each time step the correspondent likelihood:

$$\mathbb{E} [z_{t_i} | \mathcal{F}_{t_{i-1}}] = A + H \mathbb{E} [y_{t_i} | \mathcal{F}_{t_{i-1}}] \quad VAR [z_{t_i} | \mathcal{F}_{t_{i-1}}] = H VAR [y_{t_i} | \mathcal{F}_{t_{i-1}}] H^T + R$$

$$\zeta_{t_i} = z_{t_i} - \mathbb{E} [z_{t_i} | \mathcal{F}_{t_{i-1}}]$$

3. Update the expected values of the states vector (unobservable) according to:

$$\mathbb{E} [y_{t_i} | \mathcal{F}_{t_i}] = \mathbb{E} [y_{t_i} | \mathcal{F}_{t_{i-1}}] + K_{t_i} \zeta_{t_i} \quad K_{t_i} = VAR [y_{t_i} | \mathcal{F}_{t_{i-1}}] H^T VAR [z_{t_i} | \mathcal{F}_{t_{i-1}}]^{-1}$$

$$VAR [y_{t_i} | \mathcal{F}_{t_i}] = (\mathbb{I} - K_{t_i} H) VAR [y_{t_i} | \mathcal{F}_{t_{i-1}}]$$

4. Forecast the two moments of the state vector so that you can forecast the two moments of the observable:

$$\mathbb{E} [y_{t_{i+1}} | \mathcal{F}_{t_i}] = C + F \mathbb{E} [y_{t_i} | \mathcal{F}_{t_i}] \quad VAR [y_{t_{i+1}} | \mathcal{F}_{t_i}] = F VAR [y_{t_i} | \mathcal{F}_{t_i}] F^T + Q$$

5. The function of the parameters to maximize is:

$$l(\Theta) = -\frac{nN \ln(2\pi)}{2} - \frac{1}{2} \sum_{i=1}^N \left[\ln(|VAR [z_{t_i} | \mathcal{F}_{t_{i-1}}]|) + \zeta_{t_i}^T VAR [z_{t_i} | \mathcal{F}_{t_{i-1}}]^{-1} \zeta_{t_i} \right]$$

F.1.1 Kalman Filter With The Multifactor CIR Model

As we know the CIR model is not Gaussian. The transition density is a non central chi-squared. To apply the former results we substitute this transition density by a Gaussian density. This is the basis for the quasi-maximum likelihood estimation (QML). This gives biased estimations for the parameters, but Simonato, using a Monte Carlo method, found that these biases are small for the multi-factor CIR model when he compared the differences between actual values and the estimated ones, and done the same thing for the Vasicek model, which has a Gaussian transition density and, therefore, is a maximum likelihood estimator.

To use the Kalman algorithm with the CIR model we first consider the state variables dynamics:

$$dy_t = k(\theta - y_t) + \sigma\sqrt{y_t}dW_t$$

Applying Itô's lemma to the function $f(t, y) = e^{kt}y$, we get:

$$d(e^{kt}y_t) = e^{kt}k\theta dt + \sigma e^{kt}\sqrt{y_t}dW_t$$

Integrating both sides between t and $t + \Delta$:

$$\begin{aligned} y_{t+\Delta} &= e^{-k\Delta}y_t + \theta [1 - e^{-k\Delta}] + e^{-k(t+\Delta)}\sigma \int_t^{t+\Delta} e^{ku}\sqrt{y_u}dW_u \\ &= a + \Phi(y_t) + v_t \end{aligned}$$

Where:

$$a = \theta [1 - e^{-k\Delta}] \quad \Phi(y_t) = e^{-k\Delta}y_t \quad v_t = e^{-k(t+\Delta)}\sigma \int_t^{t+\Delta} e^{ku}\sqrt{y_u}dW_u$$

From Chapter 3 we know the moment generating function $M_{y_t}(w)$ for the CIR model and then we have:

$$\mathbb{E}[y_{t+\Delta}|\mathcal{F}_t] = \left. \frac{\partial M_{y_t}}{\partial w} \right|_{w=0} = e^{-k\Delta}y_t + \theta [1 - e^{-k\Delta}]$$

$$VAR[y_{t+\Delta}|F_t] = \left. \frac{\partial^2 M_{y_t}}{\partial^2 w} \right|_{w=0} - \left(\left. \frac{\partial M_{y_t}}{\partial w} \right|_{w=0} \right)^2 = \frac{\sigma^2}{k} (1 - e^{-k\Delta}) \left[\frac{\theta}{2} (1 - e^{-k\Delta}) + e^{-k\Delta}y_t \right]$$

In this way we have for the two-factor CIR

$$C = \begin{bmatrix} \theta_1 [1 - e^{-k_1\Delta}] \\ \theta_2 [1 - e^{-k_2\Delta}] \end{bmatrix} \quad F = \begin{bmatrix} e^{-k_1\Delta} & 0 \\ 0 & e^{-k_2\Delta} \end{bmatrix}$$

$$Q = \begin{bmatrix} \frac{\sigma_1^2}{k_1} (1 - e^{-k_1\Delta}) \left[\frac{\theta_1}{2} (1 - e^{-k_1\Delta}) + e^{-k_1\Delta}y_{1t} \right] & 0 \\ 0 & \frac{\sigma_2^2}{k_2} (1 - e^{-k_2\Delta}) \left[\frac{\theta_2}{2} (1 - e^{-k_2\Delta}) + e^{-k_2\Delta}y_{2t} \right] \end{bmatrix}$$

For the initialization of the state vector ($\Delta = 0$) and because the parameters are assumed to be constant over time we use:

$$\mathbb{E}[y_1 | \mathcal{F}_0] = [\theta_1, \theta_2] \quad VAR[y_1 | \mathcal{F}_0] = \begin{bmatrix} \frac{\sigma_1^2 \theta_1}{2k_1} & 0 \\ 0 & \frac{\sigma_1^2 \theta_1}{2k_1} \end{bmatrix}$$

F.2 Extended Kalman Filter

In the case of non-linear systems we can't use the above results. A natural choice to deal with the problem is to use markov chain monte carlo (MCMC) methods. These methods have the inconvenient of being very time-consuming compared to the Kalman filter. One solution is, whenever possible, to linearise the equations through a Taylor expansion around a prediction point. Considering that just the measurement equation is non-linear in the state variables, and considering again additive noise, we have the new measurement equation:

$$z_{t_i} = \Phi(y_{t_i}) + v_{t_i}$$

The algorithm works for the state variables related equations just as in the case of the linear Kalman filter. For the equations related with the measurements we make a Taylor expansion around the predicted value of the state variable:

$$\Phi(y_{t_i}) = \Phi(\mathbb{E}[y_{t_i} | \mathcal{F}_{t_{i-1}}]) + \frac{\partial \Phi(y_{t_i})}{\partial y_{t_i}} \Big|_{\mathbb{E}[y_{t_i} | \mathcal{F}_{t_{i-1}}]} (y_{t_i} - \mathbb{E}[y_{t_i} | \mathcal{F}_{t_{i-1}}])$$

By just rearranging terms we can use the algorithm described in the previous section by substituting A and H by:

$$A = \Phi(\mathbb{E}[y_{t_i} | \mathcal{F}_{t_{i-1}}]) - \frac{\partial \Phi(y_{t_i})}{\partial y_{t_i}} \Big|_{\mathbb{E}[y_{t_i} | \mathcal{F}_{t_{i-1}}]} \mathbb{E}[y_{t_i} | \mathcal{F}_{t_{i-1}}] \quad (\text{F.1})$$

$$H = \frac{\partial \Phi(y_{t_i})}{\partial y_{t_i}} \Big|_{\mathbb{E}[y_{t_i} | \mathcal{F}_{t_{i-1}}]} \quad (\text{F.2})$$

Appendix G

Matlab Codes

The codes used in Chapter 4 are written in MATLAB. These codes use the Kalman filters of Appendix F to estimate state variables values and also the values of the state variable time-homogeneous parameters. For a specific model the codes are divided in two parts. Take as an example the estimation of the default process done in Section 4.2. The code work as follows:

1. The first script, *CIRDefaultumtent.m*, receives three excel files. In them we have the dates of the observations, the values of those observations, the values of the state variables Y_1 and Y_2 for each observational date, the payments dates of the bond and the parameters of the variables Y_1 and Y_2 , plus α_0 of equation (3.1);
2. With this data this code runs a minimization problem for the symmetric of the likelihood. We fix a parameter vector and call the second file — *CIRDefaultent.m* — that uses the Kalman filter algorithm to calculate the state variables values and the likelihood. The minimization routine change the parameter vectors and find the minimum likelihood. For that vector we save the state variables values;
3. We go back to *CIRDefaultumtent.m* and use the results of the minimization algorithm plus the Kalman filter to price the defaultable bond considered in the observation dates. The code then calculates the implicit yield-to-maturity and, comparing with the observations, the associates root mean square error.

The rest of the codes work in a similar fashion, and all of them are available at: <https://github.com/PedroFilipeCruz/AppendixG>

Bibliography

- [1] Bolder, D. J. 2001, “*Affine Term-Structure Models: Theory and Implementation*”, Bank of Canada Working Paper No. 2001-15.
- [2] Chatterjee, S. 2005, ”*Application of the Kalman Filter for Estimating Continuous Time Term Structure Models: The Case of UK and Germany*”, Working Paper.
- [3] Chen, R. and Scott,L. 1995, “*Multi-Factor Cox-Ingersoll-Ross Models of the Term Structure: Estimates and Tests From a Kalman Filter Model*”, Working Paper, University of Georgia.
- [4] Cheridito, R., D. Filipovic, and Kimmel, R. 2007, “*Market Price of Risk Specifications for Affine Models: Theory and Evidence*”, Journal of Financial Economics 83,123–70.
- [5] Cochrane, J.H., “*Asset Pricing*”, New Jersey: Princeton University Press, 2005.
- [6] Collin-Dufresne, P., R. Goldstein, Jones, C. 2008, “*Identification of Maximal Affine Term Structure Models*”, Journal of Finance 63,743–95.
- [7] Dai, Q., and Singleton, K. 2000, “*Specification Analysis of Affine Term Structure Models*, Journal of Finance 55, 1943–1978.
- [8] Dai, Q., and Singleton, K., 2003, “*Term Structure Dynamics in Theory and Reality*”, Review of Financial Studies 16, 631–678.
- [9] Duan, J.C., Simonato, J.G. 1998, ”*Estimating And Testing Exponential Affine Term Structure Models By Kalman Filters*”,Review of Quantitative Finance and Accounting, 13, 111-135
- [10] Duffee, G. 1996, ”*Treasury Yields and Corporate Bond Yields Spreads: An Empirical Analysis*”, Working Paper, Federal Reserve Board.
- [11] Duffee, G. 1999, ”*Estimating The Price Of Default Risk*”, Review of Financial Studies, 12, 197-226.

-
- [12] Duffee, G. 2002, “*Term Premia and Interest Rate Forecasts in Affine Models*”, *Journal of Finance* 57, 405–443.
- [13] Duffie, D., “*Dynamic Asset Pricing Theory*”, New Jersey: Princeton University Press, 2001.
- [14] Duffie, D., and Kan, R. 1996, “*A Yield-Factor Model of Interest Rates*”, *Mathematical Finance* 6,379–406.
- [15] Duffie, D., and Lando, D. 1997, “*The Term Structure of credit Spreads with Incomplete Accounting Information*”, Working Paper, Stanford University
- [16] Duffie, D., and Singleton, K.J. 1999, “*Modeling The Term Structure Of Defaultable Bonds*. *Review of Financial Studies*, 12, 687-720
- [17] Duffie, D. and Singleton, K.J., “*Credit Risk*”, New Jersey: Princeton University Press, 2003.
- [18] Feldhütter, P. 2006, “*Can Affine Models Match the Moments in Bond Yields?*”, Working Paper, Copenhagen Business School.
- [19] Grasselli, M. R., and Hurd, T. R. 2011, “*Math 774 - Credit Risk Modeling*”, McMaster University.
- [20] Jarrow, R., Li, H., Liu, S., Wu, C. 2010, “*Reduced-Form Valuation of Callable Corporate Bonds: Theory and Evidence*”, *Journal of Financial Economics*, 95, 227-248.
- [21] Jeanblanc, M. and Rutkowski, M. 2000, “*Modelling of Default Risk: Mathematical Tools*. Working paper, Warsaw University of Technology.
- [22] Kallenberg, O., “*Foundations of Modern Probability*”, New York: Springer-Verlag, 1997
- [23] Karatzas, I. and Shreve, S.E., “*Brownian Motion and Stochastic Calculus*”, New York: Springer-Verlag, 1991.
- [24] Kimmel, R. 2011, “*Changing Times: The Pricing Problem in Non-Linear Models*”, Working Paper, Ohio State University.
- [25] Lamberton, D. and Lapeyre, B. “*Introduction to Stochastic Calculus Applied To Finance*”, Boca Raton: Chapman & Hall, 1996.
- [26] Lando, D. 1998, “*On Cox Processes and Credit Risk Securities*”, *Review of Derivatives Research*, 2, 99-120.

-
- [27] Lando, D., "*Credit Risk Modelling*", New Jersey: Princeton University Press, 2004.
- [28] Leland, H., and Toft, K. 1996, "*Optimal Capital Structure, Endogenous Bankruptcy, and the Term Structure of Credit Spreads*", *Journal of Finance*, 51, 987-1019.
- [29] Litterman, R., and Scheinkman, J. 1988, "*Common Factors Affecting Bond Returns*". Working Paper, Goldman Sachs
- [30] Lund, J. 1997, "*Non-Linear Kalman Filtering Techniques for Term-Structure Models*", Working Paper, Aarhus School of Business.
- [31] Merton, R. C. 1974, "*On The Pricing Of Corporate Debt: The Risk Structure Of Interest Rates*", *Journal of Finance*, 29, 449-470.
- [32] Morini, M., "*Understanding and Managing Model Risk - A Practical Guide for Quants, Traders and Validators*", Chichester: John Wiley & Sons, 2011.
- [33] Mosburger, G. and Schneider, P. 2005, "*Modelling International Bond Markets with Affine Term Structure Models*". Working paper, University of Vienna, Vienna University of Economics and Business.
- [34] Park, M., Clark, S.P. 2013, "*A Reduced-Form Model for Valuing Bonds with Make-Whole Call Provisions*", Working Paper.
- [35] Powers, E. A., and Tsyplakov, S. 2009, "*What is the Cost of Financial Flexibility? Theory and Evidence for Make-Whole Call Provisions*", *Financial Management*, 37, 485-512.