Dynamic Interest Rate and Credit Risk Models

Adam Saeed Iqbal

Imperial College Business School, Imperial College London

This thesis is submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy of Imperial College London
Declaration of Originality

I declare that the research presented in this thesis is my own and that all else is appropriately referenced.
To

my parents

with

love and unbounded admiration.
Acknowledgements

I thank William Perraudin for his supervision, for his insightful feedback and for allowing me the freedom to pursue my academic interests.

I thank Lara Cathcart for her general comments on my research and, in particular, for her encouragement and guidance towards completing this thesis.

I am grateful to Julie Paranics for her efforts on the administrative side in helping me return to research following an unexpected interruption to studies.

Finally, I thank my PhD colleagues, Leo Evans, Filip Zikes, Ewan Mackie and Robert Metcalfe, for many interesting discussions about research and for their friendship.
This thesis studies the pricing of Treasury bonds, the pricing of corporate bonds and the modelling of portfolios of defaultable debt. By drawing on the related literature, Chapter 1 provides economic background and motivation for the study of each of these topics.

Chapter 2 studies the use of Gaussian affine dynamic term structure models (GDTSMs) for forming forecasts of Treasury yields and conditional decompositions of the yield curve into expectation and risk premium components. Specifically, it proposes market prices of risk that can generate bond price time series that are consistent with the important empirical result of Cochrane and Piazzesi (2005), that a linear combination of forward rates can forecast excess returns to bonds. Since the GDTSM here falls into the essentially affine class (Duffee (2002)), it is analytically tractable.

Chapter 3 studies conditional risk premia in a commonly applied default intensity based model for pricing corporate bonds. Here, I refer to such models as completely affine defaultable dynamic term structure models (DDTSMs). There are two main contributions. First, I show that completely affine DDTSMs imply that the compensation for the risk associated with shocks to default intensities (the credit spread risk premium) is related to the volatility of default intensities. Second, I run regressions to show that this relationship holds in a set of corporate bond data.

Finally, Chapter 4 proposes a new dynamic model for default rates in large debt portfolios. The model is similar in principle to Duffie, Saita, and Wang (2007) and Duffie, Eckner, Horel, and Saita (2009) in that the default intensity depends on the observed macroeconomic state and unobserved frailty variables. However, the model is designed for use with more commonly available aggregate, rather than individual, default data. Fitting the model to aggregate charge-off rates in US corporate, real-estate and non-mortgage retail sectors, it is found that interest rates, industrial production and unemployment rates have quantitatively plausible effects on aggregate default rates.
## Contents

**List of Abbreviations** 8

1 Interest Rate and Credit Risk Modelling 11  
1.1 Pricing Treasury Bonds ................................. 11  
1.1.1 Why Study Treasury Bond Returns? ............... 11  
1.1.2 Pricing in the Cross Section ...................... 19  
1.2 Pricing Corporate Bonds ............................... 20  
1.2.1 Background ........................................... 20  
1.2.2 Why Study Corporate Bond Returns? .............. 21  
1.3 Pricing Portfolios of Defaultable Debt ............ 26  
1.3.1 Why Model Portfolios of Defaultable Debt? ...... 26  
1.3.2 Relating Portfolios of Defaultable Debt and Defaultable Debt ... 27  

2 A GDTSM with Risk Premia Linear in Forward Rates 29  
2.1 Introduction ............................................. 30  
2.1.1 Further Related Literature ......................... 33  
2.2 Affine Dynamic Term Structure Models ............. 34  
2.2.1 Affine Bond Pricing ................................ 34  
2.2.2 Expected Returns to Bonds ....................... 36  
2.2.3 Essentially Affine Market Prices of Risk .......... 36  
2.3 Model Specification .................................... 37  
2.3.1 Prices of Risk ...................................... 37  
2.3.2 Forward Rates in Terms of the State Vector .... 39  
2.3.3 Calculation of $F_{0,X}$ and $F_{1,X}$ .............. 40  
2.3.4 Parameter Vectors, Invariant Transformations and Notation .................. 40  
2.4 Estimation .............................................. 41  
2.4.1 Forward Rates in the Price of Risk ............... 41  
2.4.2 Canonical Representation ......................... 42  
2.4.3 Invariant Transformation to Observable States .. 43
List of Abbreviations

**CAPM** Capital Asset Pricing Model.

**CCAPM** Consumption Based Capital Asset Pricing Model.

**CDO** Collateralised Debt Obligation.

**CDS** Credit Default Swap.

**CPI** Consumer Price Index.

**DTSM** Dynamic Term Structure Model.

**DDTSM** Defaultable Dynamic Term Structure Model.

**GDP** Gross Domestic Product.

**HMI** Housing Market Index.

**ICAPM** Intertemporal Capital Asset Pricing Model.

**IID** Independently Identically Distributed.

**IP** Industrial Production.

**LEH** Local Expectations Hypothesis.

**ML** Maximum Likelihood.

**MLE** Maximum Likelihood Estimate.

**MPR** Market Price of Risk.

**MSE** Mean Squared (Forecast) Error.

**ODE** Ordinary Differential Equation.
**OLS** Ordinary Least Squares.

**SDF** Stochastic Discount Factor.

**VAR**\((n)\) Vector Autoregression of Order \(n\).

**VWRETD** Value Weighted Returns (including distributions) file distributed by CRSP.
Chapter 1

Interest Rate and Credit Risk Modelling

This thesis contributes to the literature on dynamic asset pricing applied to fixed income securities. The contributions are in three closely related areas: the pricing of Treasury bonds, the pricing of defaultable corporate bonds and the modelling of portfolios of defaultable debt. The purpose of this introductory chapter is to provide motivation and to describe background literature for the study of the above mentioned topics and to explain the close relationships that exist between them.

1.1 Pricing Treasury Bonds

1.1.1 Why Study Treasury Bond Returns?

One may ask, why should we study risk premia in Treasury bond returns? In this subsection, I provide several reasons organised according to their relevance to different participants in the economy: households, firms and government. This is not intended to be an exhaustive list, but it explains the substantial past and present research effort devoted to understanding the shape of the yield curve and its co-movement over time.

Households - Consumers and Investors

Treasury bonds, like stocks and derivatives, are assets that expose their holders to economic risks. From the perspective of an investor, the reason for studying bond prices is therefore the same as that for studying the pricing of any asset: the statistical properties of bond returns will influence the investor’s optimal consumption-portfolio decision. Moreover, since bond pricing is the basic building block of all asset pricing, the study
of bond returns is of particular importance. The remainder of this subsection explains these ideas.

As the availability of data on bond returns has increased, so has our understanding of their statistical properties. This has implications for optimal consumption-portfolio decisions, as I now describe.

The classic theory of bond returns is the expectations hypothesis. It says that expected excess returns to Treasury bonds of all maturities over bills equal a fixed constant. If the expectations hypothesis were true, then the conditional expected excess return to bonds at any time is equal to its unconditional counterpart. Under the portfolio rules of the one-period static Capital Asset Pricing Model (CAPM) (Sharpe (1964), Lintner (1965, 1965a), Mossin (1966)) or a standard continuous time optimal consumption-portfolio model (Merton (1969, 1971)), the portfolio weight assigned to a particular asset depends on its expected excess return (as well as other properties such as the covariance matrix of all assets and risk aversion parameters). The expectations hypothesis tells us that, provided that the other assumptions that underly these portfolio models are appropriate, unconditional bond returns may be used here to address the optimal consumption-portfolio problem.

However, it is well known that the expectations hypothesis is not supported by the data and expected excess returns to bonds are time-varying and forecastable using forward rates or yield spreads (Fama and Bliss (1987), Campbell and Shiller (1991)). More recently, Cochrane and Piazzesi (2005, 2008) have found empirical evidence that a single factor, a linear combination of five forward rates, is able to forecast expected excess returns to bonds of all maturities.

The recent empirical evidence suggests that the forward rates that form the above mentioned return forecasting factor are state variables that are required to describe the conditional distribution of bond returns. In other words, there are shifts in the investment opportunity set. This means that an Intertemporal CAPM (ICAPM) (Merton (1973a)) is required to address the optimal consumption-portfolio problem. The search for state variables is an unfinished task and it provides motivation to study bond prices

\footnote{To be precise, the expectations hypothesis says that expected log (rather than continuously compounded) holding period returns across maturities are equal to a constant. The difference is small. For example, if annual returns are distributed as $\ln R \sim N(\mu, \sigma^2)$ then $E[R] = e^{\mu + \frac{1}{2} \sigma^2}$. If $\mu = 0.1$ and $\sigma = 0.1$ then $\frac{1}{2} \sigma^2 = 0.005$ which is very small compared with $\mu$, but not zero.}

\footnote{I use the terms “risk premium” and “expected excess return” interchangeably.}
and, in particular, conditional bond returns. In Chapter 2 I construct a term structure model that is motivated by this recent empirical evidence.

More generally, since investors optimise their portfolios across all assets, a more fundamental reason to focus on Treasury bonds in particular is that bond pricing is the basic building block of all asset pricing. Any asset can be thought of as a bond plus cash flow risk. As a result, being able to price a bond is necessary to price any asset and an understanding of the statistical properties of bond returns is necessary to understand the statistical properties of returns to any asset. This point can be seen as follows.

The absence of arbitrage implies the existence of a stochastic discount factor (SDF), \( \pi \), that prices all traded assets (Harrison and Kreps (1979)). By the definition of the SDF, the price of any non-dividend paying asset at time \( t \), \( S_t \), is given by

\[
S_t = E_t \left[ \frac{\pi_T}{\pi_t} S_T \right] \quad \text{for all } T > t, \tag{1.1}
\]

where \( E_t \) denotes the time \( t \) conditional expectation. We can apply this equation to price bonds. A zero-coupon bond that matures at time \( T \) has, by definition, a price of 1 at time \( T \). Letting \( P_t(T) \) denote the time \( t \) price of a zero-coupon bond that matures at time \( T \), an application of (1.1) reveals that

\[
P_t(T) = E_t \left[\frac{\pi_T}{\pi_t}\right]. \tag{1.2}
\]

Next, by defining a random variable \( \epsilon_{t,T} \) by

\[
\epsilon_{t,T} \equiv S_T - E_t[S_T]
\]

and substituting this and the result in (1.2) into (1.1) we can rewrite the asset pricing equation as

\[
S_t = P_t(T)E_t[S_T] + E_t \left[ \frac{\pi_T}{\pi_t} \epsilon_{t,T} \right] \quad \text{for all } T > t. \tag{1.3}
\]

This equation says that the price of any asset is simply the price of \( E_t[S_T] \) zero-coupon bonds that mature at time \( T \) plus a cash flow risk term. The above arguments make clear that, of all assets, it is particularly important that we understand the pricing of bonds. A similar result holds for dividend paying assets.

The term structure literature is extensive. Most of this literature emphasises risk-adjusted probabilities and does not address the issue of specifying drifts and market
prices of risk. That is, authors have proposed only the risk-adjusted dynamics for the short rate and then calibrated the parameters to the cross section of bond prices. The one-factor models of Ho and Lee (1986), Black, Derman, and Toy (1990), Black and Karasinski (1991) and Hull and White (1990, 1993) are examples of such an approach, and numerous authors (too many to list here) have calibrated models in which the risk-adjusted dynamics of the short rate follow the stochastic process suggested in Vasicek (1977), Cox, Ingersoll, and Ross (1985) or a multifactor generalisation of these as described in Duffie and Kan (1996).

The approach of specifying only risk-adjusted probabilities only may be appropriate for applications such as pricing derivatives on bonds relative to the bond itself, where accurate volatility modelling is the most important concern. It is also suitable for constructing a smooth yield curve across all bond maturities to price bonds of non-traded maturities when the bond prices of only a subset of maturities are known (Section 1.1.2 below discusses cross sectional bond pricing further). However, this literature does not address empirical facts about expected excess returns to bonds and it is therefore not useful for studying the portfolio selection problems faced by households.

More recently, several authors have begun to specify models that attempt to fit both the cross section of bond prices and also the empirical facts about the temporal behavior of bond returns. Models that simultaneously attempt to understand the shape of the yield curve (the cross section of bond prices) and its co-movement over time are commonly known in the literature as dynamic term structure models (DTSMs). Prominent contributions in this area are Duffee (2002), Dai and Singleton (2002) and Duarte (2004).

DTSMs is an important field of current research. The research presented in Chapter 2 of this thesis is a contribution in this direction. Specifically, it takes the arbitrage free modelling framework proposed by Duffee (2002) and imposes restrictions on market prices of risk so that the model is able to generate return forecasting equations that are, in principle, consistent with the empirical results of Cochrane and Piazzesi (2005). Based on the recent insights presented in Joslin, Singleton, and Zhu (2010) relating to Gaussian affine DTSMs, the model is tested by comparing its out-of-sample yield forecasting ability with a more general nesting model.

**Firms**

The observed real yield curve and an understanding of conditional risk premia in bond returns can shed light on firms’ investment decisions in that, together, they can provide
conditional forecasts for future levels of consumption at multiple future times. If consumption is forecast to increase, then firms must be investing to increase their inventories in anticipation of higher future consumer spending at the time points suggested by the real yield curve. It is not surprising that asset prices can provide such forecasts. After all, it is through trading in financial markets that consumers are able to reflect their marginal rates of substitution across time. These ideas can be understood in the framework of a conditional macroeconomic asset pricing model/consumption based capital asset pricing model (CCAPM) as follows.

Suppose that at time $t$ the objective of a representative economic agent is to maximise his or her long term expected utility by choosing a number, $\xi$, of the $n$-period zero-coupon bond to purchase. Assuming a time separable form for the agent’s utility function, this problem can be written as

$$
\max_{\{\xi\}} \mathbb{E}_t \left[ U\left( \{c_{t+j}\}_{j=0}^{\infty} \right) \right] = \max_{\{\xi\}} \mathbb{E}_t \left[ \sum_{j=0}^{\infty} \beta^j u(c_{t+j}) \right], \text{ such that }
$$

$$
c_t = e_t - \xi P_t(t+n),
$$

$$
c_{t+j} = e_{t+j}, \quad \text{for } j = 1, \ldots, n-1, n+1, \ldots, \infty
$$

$$
c_{t+n} = e_{t+n} + \xi,
$$

where $c_{t+j}$ and $e_{t+j}$ are the time $t+j$ levels of consumption and endowment respectively, and $U(\{c_{t+j}\}_{j=0}^{\infty}) = \sum_{j=0}^{\infty} \beta^j u(c_{t+j})$ is the utility function for some impatience parameter $\beta < 1$. The first order conditions reveal that the $n$-period real bond yield, $R^n_t$, is given by

$$
R^n_t \equiv \frac{1}{P_t(t+n)} = \frac{1}{\mathbb{E}_t \left[ \beta \frac{u'(c_{t+n})}{u'(c_t)} \right].} \tag{1.4}
$$

This equation applies once households have implemented their consumption-portfolio choices. A high $n$-period real bond yield today suggests that, in expectation, $u'(c_{t+n})$ will be small relative to $u'(c_t)$. For a concave $u(\cdot)$ this means that $c_{t+n}$ is likely to be large relative to $c_t$ and consumption is forecast to rise. Firms may react to this forecast to ensure that they minimise unplanned changes to their inventories over time. An important question is, given the time $t$ real $n$-period yield, how much is consumption forecast to rise? The empirical evidence about bond returns suggests that the answer to this question may change over time; it is related to time $t$ conditional risk premia. This can be understood as follows.

Equation (1.4) shows that the change in yield associated with a change in consumption
forecasts depends on the form of $u'(\cdot)$. In turn, $u'(\cdot)$ depends on the representative agent’s level of risk aversion. For example, in the case of an agent with power utility, $u'(c) = c^{-\gamma}$, which depends on the risk aversion parameter $\gamma$. The complication is, however, that $\gamma$ may vary over time (perhaps $\gamma_t$ is more appropriate notation). Some evidence supporting the suggestion that $\gamma$ varies over time is given below. However, the main implication here is that at times when $\gamma$ is large, smaller increases in expectations of $c_{t+n}$ relative to $c_t$ can deliver higher $n$-period bond yields relative to times when $\gamma$ is small. In the power utility example, the quantity

$$\frac{u'(c_{t+n})}{u'(c_t)} = \left(\frac{c_{t+n}}{c_t}\right)^{-\gamma}$$

and so marginal utility growth is more sensitive to consumption growth at larger $\gamma$. Therefore, to use yields to forecast future consumption, since $\gamma$ is generally unobserved, one must be able to at least proxy for its temporal variation.

The empirical fact that excess returns to bonds are predictable was discussed in the previous subsection. Further, it is a standard result in a CCAPM with power utility and log normal consumption growth that expected excess returns depend on $\gamma$:

$$E_t \left[ \frac{P_{t+1}(t+n)}{P_t(t+n)} \right] - R^t_1 \approx \gamma \sigma_t \left( \frac{c_{t+1}}{c_t} \right) \sigma_t \left( \frac{P_{t+1}(t+n)}{P_t(t+n)} \right) \rho_t \left( \frac{c_{t+n}}{c_t} \right)^{-\gamma} \frac{P_{t+1}(t+n)}{P_t(t+n)}.$$

(1.5)

Here, the notation $\sigma_t(\cdot)$ denotes the conditional standard deviation and $\rho_t(\cdot, \cdot)$ denotes the conditional correlation.

Given the above equation, one way of reconciling the predictability of excess returns to bonds with the CCAPM is that $\gamma$ is time varying. If changing risk aversion is indeed the fundamental source of return predictability then variables such as forward rates, yield spreads and combinations of forward rates that have been used to forecast excess returns to bonds in Fama and Bliss (1987), Campbell and Shiller (1991) and Cochrane and Piazzesi (2005) may be good proxies for the temporal variation in $\gamma$. These proxies can be used to help make conditional forecasts of consumption. The research presented in Chapter 2 contributes to our understanding of the relationship between forward rates and the predictability of Treasury bond returns by reconciling this predictability with an arbitrage free (tractable) DTSM.

---

3The term “associated” is used to emphasise that the causality here is not clear; both consumption and yields are endogenous quantities that are related through [1.4].
As an aside, it is noteworthy that, since in principle one can use a CCAPM to price all assets, other asset prices can also be used to forecast consumption. However, this requires knowledge of the asset’s expected dividend (for stocks) or coupon payment (for defaultable bonds) at multiple future dates and the correlation of these dividends and default risky coupons with consumption. Amongst all assets, default-free zero-coupon Treasury bond prices provide perhaps the simplest data with which to forecast future consumption. Indeed, many articles have used the yield curve/bond prices to forecast future real economic activity (Harvey (1988), Estrella and Hardouvelis (1991), Hamilton and Kim (2002), Ang, Piazzesi, and Wei (2006) are some well known examples). Stock and Watson (2003) provide a detailed survey of the literature that has used bond prices as well as the prices of other assets (stocks, gold and others) to forecast economic output.

**Governments and Central Banks**

The DTSM in Chapter 2 attempts to quantify the size and variation over time of risk premia in Treasury bond returns. These quantities have implications for government debt policy relating to both the timing of debt issuance and to the management of its maturity structure. These ideas can be understood as follows.

Conditional risk premia are relevant for the timing of debt issuance for a government wishing to minimise its actuarial financing costs; the government may wish to avoid issuing debt when risk premia are high. This may be challenging. For instance, some governments use fiscal stimuli in efforts to counter recessions. They must therefore raise finance when recessions are forecast. However, this may be precisely the time when risk premia and therefore the cost of raising debt is highest. Indeed, there is empirical evidence that variables that forecast high excess returns also forecast recessions.

The term spread is a well known example of a variable that forecasts excess returns to stocks and bonds and also recessions (Fama and French (1989)). Another example is the index dividend to price ratio, which is found to forecast excess returns in equity indices (Cochrane (1992)). Since most expected cash flow variation in the cross section of firms is idiosyncratic, the variation in the index dividend to price ratio is mostly due to varying expected excess returns (Vuolteenaho (2002)).

The idea that the same variables that forecast excess returns also forecast recessions is intuitive. One may expect risk aversion and hence, by (1.5) above, risk premia to be highest when a recession is forecast; investors may require a larger risk premium to be induced into holding risky assets when their outside income becomes less certain.
Campbell and Cochrane (1999) present an economic model that captures the idea of a recession premium. A DTSM with market prices of risk specified with sufficient flexibility to capture the time variation in risk premia can allow one to calculate the conditional component of the cost of debt due risk premia and that due to expectations of future interest rates. The research in Chapter 2 is in the direction of finding such a specification.

Risk premia are also relevant to a government wishing to choose the maturity structure of its debt in a way that is optimal in terms of cost. The rationale for actively managing the maturity structure of public debt is not clear in frictionless markets with no distortionary taxes. The reason is that the Modigliani-Miller theorem (Modigliani and Miller (1958, 1961)) applies. Nevertheless, despite this, the actuarial expected cost to the government for financing its debt can depend on debt maturity. For example, if the term structure of the risk premium to bonds is downward sloping, then, after an adjustment for expectations of future short interest rates, financing using long bonds is actuarially less expensive than financing with short bonds. Campbell, Shiller, and Viceira (2009), for instance, find that long term inflation-indexed bonds are safe assets in that they command a low risk premium and argue that they are therefore a relatively cheap method of debt financing for governments.

Another reason to study risk premia in bond returns relates to stabilisation policy, a public policy often implemented by central banks aimed at manipulating aggregate demand to reduce the severity of short run economic fluctuations. The relevance of bond risk premia here can be understood as follows.

Loosely speaking, in a rational expectations framework, only movements in interest rates (through, for example, monetary action) that are unexpected by households can change their consumption-portfolio decisions. In order to manipulate aggregate demand, therefore, it is useful for the central bank to know households’ conditional expectations of future interest rates. Here, risk premia complicate matters. In general, yields on long maturity bonds are expected values of average future short yields only after adjustment for risk (the strategy of rolling over short bonds is “risky” compared with buying and holding a long bond). In other words, extracting conditional expectations of future interest rates from the yield curve requires a knowledge of conditional risk premia. Chapter 2 is a contribution in this direction.

The above discussion is mainly concerned with the time series behaviour of bond prices. It is important to note that studying the cross section of bond prices is also relevant to stabilisation policy. For example, the central bank is usually able to move the short
end of the yield curve. In the USA, for instance, the Federal Open Market Committee has in the past used and continues to employ the federal funds rate, the interest rate banks charge one another for overnight loans, as a policy instrument. This can move the short end of the yield curve. However, the main purpose is to shift the aggregate demand curve, but this depends on long term rather than short term yields. For instance, households base their decision on whether to buy or rent a house on long-term mortgage rates and not the federal funds rate. A no-arbitrage model of the yield curve can provide an understanding of the how short yields translate into long yields. This is the topic of the subsection below.

1.1.2 Pricing in the Cross Section

Studying the cross sectional relationships between bond prices is useful for at least three reasons. Two of these were mentioned above: to understand how movements in the short end of the yield curve translate into movements at the long end, and to price derivatives on bonds. A third reason is that the cross sectional relationships between bond prices can be used in separating risk premia from expectations about future short rates; Chapter 2 focusses on this topic. These reasons can be understood as follows.

A cross section of bonds of different maturities can be liquidly traded and so the assumption that there are no arbitrage opportunities is certainly reasonable. Further, if in addition, the short rate of interest, \( r_t \), is assumed to depend on a Markov state vector \( X_t \), then bond prices can be written as a function of \( X_t \):

\[
P_t(T) = \mathbb{E}_t^Q \left[ e^{-\int_t^T r_u du} \right] = f(X_t, t, T; \psi). \tag{1.6}
\]

Here, \( Q \) denotes the risk-adjusted probability measure and \( \psi \) denotes a vector of parameters that (i) enter the \( Q \) dynamics of the stochastic process followed by \( X_t \) and (ii) that determine the dependence of \( r_t \) on \( X_t \).

First, given the above equation, it is possible to understand how movements in different parts of the yield curve relate to each other. For instance, in the affine framework of Duffie and Kan (1996) each element of \( X_t \) is a bond yield of a different maturity. One can therefore see the affect on yields that are not in \( X_t \) of changing a yield that is in \( X_t \).

Next, given the cross section of bond prices at a given point in time, it is possible to calibrate the parameters \( \psi \) to fit, by some criterion, the observed prices. In other words, it is possible to extract information about the \( Q \) dynamics followed by \( X_t \) from the cross section of bond prices. Knowledge of \( \psi \) is then sufficient to price derivatives on Treasury
Finally, given the Q dynamics of $X_t$ and information about the time series behaviour of $X_t$ (the P dynamics), it is possible to calculate risk premia in bond returns. For example, if the cross section of bond prices reveals an upward sloping yield curve and simultaneously interest rates are forecast to fall then this suggests that risk premia on long maturity bonds are positive.

1.2 Pricing Corporate Bonds

1.2.1 Background

Given the seminal literature on and relating to the pricing of corporate bonds (Black and Scholes (1973), Merton (1973b, 1974)), one may suggest that the corporate bond pricing problem is redundant once the firm itself has been priced. The reason is that this literature models corporate debt as a contingent claim on the firm in a locally complete market and so once the firm has been priced, it is straightforward to price the bond relative to the firm by a no-arbitrage argument. However, under this model the predicted levels of corporate bond spreads are on average substantially lower than those observed empirically (Jones, Mason, and Rosenfeld (1984), Eom, Helwege, and Huang (2004)). This has lead to many extensions to make these frameworks more realistic in their modelling of a firm’s default process. The original models and the subsequent extensions are classified as structural models. Despite the increasing complexity of and increasing realism captured in structural models, the empirical support for them is mixed, as I discuss below.

Structural Models of Corporate Bonds Prices

Among the well known extensions to the above mentioned literature is Geske (1977) who uses compound option modelling methods to price debt at multiple maturities. Longstaff and Schwartz (1995) developed a pricing model with stochastic default-free interest rates and build on the idea first proposed in Black and Cox (1976) that default occurs when the value of assets first pass below an exogenously imposed default boundary. Collin-Durbesne and Goldstein (2001) reflect firms’ behavior by allowing for mean-reverting leverage ratios in addition to stochastic interest rates and an exogenously imposed default boundary. Leland and Toft (1996) extend previous models of optimal default timing and the valuation of debt with taxes and bankruptcy distress costs (Fisher, Heinkel, and Zechner (1989), Leland (1994)) to allow for coupon debt of finite maturity. However, Eom, Helwege, and Huang (2004) have found that, with the exception of Leland and
Toft (1996), all of the above mentioned extensions under predict level of corporate bond spreads for safer bonds and overpredict spreads on firms with high leverage or volatility. They also find that Leland and Toft (1996) overpredicts spreads on most bonds and particularly so on those with high coupons.

**Intensity Based Models of Corporate Bond Prices**

Given the above mentioned empirical difficulties associated with structural models, a large proportion of the recent corporate bond modelling literature and the closely related credit default swap (CDS) modelling literature has focused on another class of models: the intensity based default models that were introduced by Artzner and Delbaen (1990, 1995), Jarrow and Turnbull (1995) and Lando (1998). These are models in which default is said to occur at the first jump time of a conditional Poisson process. That is, loosely speaking, conditional on no default occurring until time $t$, the probability that the corporate will default in the next small amount of time $\Delta t$ is approximately $\lambda_t \Delta t$. $\lambda_t$ is called the default intensity and it can be shown that it is also closely related to the short corporate bond spread. $\lambda_t$ can be allowed to follow a non-negative stochastic process.

Intensity based models have gained popularity for their simplicity, flexibility and analytical tractability. In particular, Duffie and Singleton (1999) provide a framework in which the short rate of interest is “default adjusted” so that corporate bond pricing problems simplify to Treasury bond pricing problems. Under the restriction that the default adjusted short rate of interest is an affine function of a state vector that follows an affine diffusion under the risk adjusted probability measure, the full machinery of Duffie and Kan (1996) that was originally designed for the relative of pricing Treasury bonds can also be used for the relative pricing of corporate bonds. Duffie (2005) provides a detailed summary of the intensity based approach to credit default modelling and Duffie and Singleton (2003), Schonbucher (2003) and Lando (2004), amongst others, have provided textbook treatments of this topic.

**1.2.2 Why Study Corporate Bond Returns?**

Chapter 3 studies conditional risk premia in corporate bond returns. Given the above discussion, there are several reasons why the study of risk premia is important. Here, I describe several of these, organised again according to their relevance to households, firms and government.
Households

Perhaps motivated by the empirical difficulties faced by several structural models, researchers have focussed on intensity models that essentially ignore capital structure considerations and treat corporate bonds as separate assets in their own right. One notable exception is Duffie and Lando (2001) who show that intensity and structural models are not necessarily distinct from each other. These authors construct a structural model with incomplete accounting information that is able to endogenously generate a default intensity. However, in general, in the intensity framework, corporate bonds should not be considered redundant assets once their issuing firm has been priced. The study of corporate bond prices is therefore important for households’ consumption-portfolio decisions for the reason that corporate bonds may provide new investment opportunities (in the sense that they make markets more complete) and an investor must separately consider the statistical properties of their returns to make an optimal consumption-portfolio decision.

Further, compared to the research effort devoted to understanding conditional risk premia in Treasury bond and stock returns (some of which was discussed in the previous subsection), there is relatively little research into conditional risk premia in corporate bond returns. As discussed in Section 1.1.1, understanding conditional risk premia is important for studying households’ optimal consumption-portfolio decisions; variables that forecast excess returns become state variables in an ICAPM. One may suspect that variables that forecast stock and Treasury bond returns, and others, may also forecast corporate bond returns. When constructing a defaultable dynamic term structure model (DDTSM) - a model that describes the co-movement of corporate bond yields over time - one must ensure that it is sufficiently flexible to capture the time variation in risk premia.

Several authors have estimated intensity based DDTSMs. Duffee (1999) was an early (and, to my knowledge, the first) paper to fully test an intensity based specification for defaultable bonds and estimate prices of risk. More recently, Driessen (2005) has extended this work to include multiple risk factors, a liquidity risk factor and, most significantly, an estimate of a default event risk premium. However, both of these models implicitly impose strong restrictions on the statistical properties of corporate bond returns; Duffee (1999) imposes that conditional expected excess returns are related to the bond’s contemporaneous volatility and both of these authors impose that compensation for the risk of changes in default intensities is related to the volatility of default intensities.
The restrictions on risk premia can be understood in the context of Duffee (2002) which shows explicitly that the subset of affine Treasury bond pricing models, the so-called *completely* affine class, impose that conditional risk premia depend only on contemporaneous bond volatility. In these models, the state vector follows an affine diffusion, and the variance of the SDF is affine in the state vector. Given the similarities between Treasury bond modelling and intensity based defaultable bond modelling, analogous restrictions carry over to DDTSMs in which the state vector follows an affine diffusion and the variance of the part of the SDF that correlates with default intensities is affine in the state vector. Extending the nomenclature of Duffee (2002), I refer to such models as completely affine DDTSMs. These ideas are made clear and the result is shown formally in Chapter 3.

Given the above discussion, Chapter 3 tests the validity of the completely affine restriction in a set of corporate bond data. This research can be considered a contribution to the extension to corporate bonds of the work that was cited in Section 1.1.1 on constructing DTSMs that are consistent with empirical findings in both the cross section and the time series of Treasury bond prices.

Further, Chapter 3 attempts to separate the prices of risk associated with changes in default intensities from those associated with changes in default free interest rates. Clearly, the risks to default intensities rather than interest rates are the risks that provide households with new investment opportunities over Treasury bonds, and their changing conditional prices provide shifts in the investment opportunity set. The research here contributes towards understanding these.

**Firms**

There is empirical evidence to suggest that risk premia, rather than Treasury interest rates, are the dominant factor to consider when studying firms’ investment decisions. For example, the average real return to stocks in the USA in the closed period from 1953 to 2009 was 8.1%\(^4\) (the turmoil of 2008 was offset by the recovery of 2009). However, over the corresponding time period, average real Treasury interest rates were 1.7%. Further, since risk premia are large compared with interest rates and the volatility of interest rates is relatively low (2.5% in the sample), one may suspect that most of the variation

\(^4\)The calculations here use annual (December 31) data from the CRSP Value-Weighted Return (including distributions) (VWRETD) file for stocks and the Fama-Bliss Discount Bond file for 1 year bonds. Inflation is calculated using the annual (December 31) Consumer Price Index (CPI) levels distributed by the Bureau of Labour Statistics.
over time in the cost of capital for firms comes from the variation in risk premia rather than the variation in interest rate.

Given the above, one may ask, do these large risk premia/costs of capital in stock returns also exist in corporate bond returns? This question is particularly relevant to a firm wishing to choose its capital structure in a way that is optimal in terms of cost.

The structural framework of Merton (1974) provides a direct answer to the above question: yes. The reason is that the no-arbitrage condition implies that excess returns to stocks and debt per unit of volatility are equal (recall that the Black-Scholes-Merton partial differential equation whose solutions, subject to appropriate boundary conditions, price both debt and equity is derived by equating each of their market prices of risk with that of the firm and hence each other). Indeed, Merton (1977) proved the Modigliani-Miller theorem in the context of a structural credit risk model, thus suggesting that effort spent on designing capital structure is wasted.

However, the empirical difficulties associated with structural models could mean that the Modigliani-Miller result does not hold empirically and the costs of capital associated with corporate bonds and equities are different, making one form of financing cheaper than the other. As a result, quantifying risk premia in corporate bond returns is important for firms.

A more subtle point is that it is not just risk premia in corporate bonds, but also their time variation that is important for firms. There are at least three reasons.

First, conditional risk premia in corporate bond returns are relevant for the timing of debt issuance. Corporates, like governments, may attempt to time their new debt issuances to times when risk premia are low, and thus minimise their actuarial financing costs.

Second, conditional risk premia are relevant for firms’ investment decisions. For example, firms may be able to profitably invest in projects with lower internal rates of return when

---

5Although the discussion here is focussed on large risk premia, I note that the empirical evidence is not conclusive. Given the level of US stock volatility ($\sigma = 18\%$ in the $T = 57$ year VWRETD sample), the standard error of the sample mean real excess return is also large ($\hat{\sigma} = 2.4\%$). As a result, it may be that such high excess returns are luck. Further, there is economic rationale behind the luck argument; to reconcile these premia with the CCAPM in (1.5), an unreasonably high risk aversion of $\gamma \geq 36$ is required. To calculate this, I assume that $\sigma_t(c_{t+1}/c_t) = 1\%$, equal to that observed in the postwar US data sample. This is part of the well known equity premium puzzle.
risk premia are low.

Third, if conditional risk premia to corporate bonds and stocks do not, for some reason, move in lock step, then the relative costs of financing with debt and equity change over time. This may influence the financing choice made at any particular time.

The research presented in Chapter 3 contributes to understanding conditional risk premia in corporate bond returns in intensity models (independent of capital structure considerations).

Governments and Central Banks

The size and variation over time of conditional risk premia in corporate bond returns (studied in Chapter 3) have implications for governments and central banks. For example, in the USA, the Mission of the Federal Reserve tasks it with “conducting the nation’s monetary policy by influencing the monetary and credit conditions in the economy in pursuit of maximum employment...”. An important question is, how effective can monetary policy be here? The answer may, in some part, depend on risk premia in corporate bond returns. This idea can be understood as follows.

Given the above discussion that the size of risk premia is large compared with interest rates, the level of federal funds rate may have only a small affect on the cost of capital. Further, given the evidence discussed in Subsection 1.1.1 that variables that forecast excess returns also forecast recessions, it may be the case that the risk premium component of firms’ costs of capital is greatest at those times when the Federal Reserve most wishes to support employment via support for investment. In other words, monetary actions to move the federal funds rate may be relatively less effective in improving credit conditions and stimulating investment at those times when it is desired to be more effective. If risk premia are high, then firms cannot invest in projects with low internal rates of return even if the federal funds rate is low.

The topics discussed here can be studied using a DDTSM with sufficient flexibility in the specification of market prices of risk to capture the time variation in risk premia to corporate bonds. Chapter 3 contributes in this direction.
1.3 Pricing Portfolios of Defaultable Debt

1.3.1 Why Model Portfolios of Defaultable Debt?

As discussed, the focus in Chapters 2 and 3 on Treasury and corporate bonds respectively is on conditional risk premia and why an understanding of conditional risk premia is important to different participants in the economy. The focus in Chapter 4 however, is on calculating expected cash flows to portfolios of defaultable debt. The reasons for this more conservative approach are given below.

In asset pricing research in which expected cash flows are modelled explicitly, it is appropriate to focus on risk premia only once one is confident that the expected cash flows are reliably calculated. The reason is that, in the absence of arbitrage, being able to calculate expected cash flows is, in general, necessary but insufficient to calculate asset prices. This is can be seen immediately from the asset pricing equation, Equation (1.3), given above. The first term in this equation is the expected asset payoff discounted by the risk free bond. The second term reflects the covariance of the asset’s payoff with the SDF and adjusts the asset’s price to account for risk.

In the case of Treasury zero-coupon bonds, calculating the expected payoff is trivial; it is 1 at maturity (assuming away sovereign credit risk). For a corporate zero-coupon bond, this problem is more difficult. The payoff is 1 at maturity only if there is no default. The structural and default intensity based methods that are commonly used to model corporate bond cash flows were discussed above. Calculating the expected cash flows associated with a portfolio of credit risks, however, is substantially more difficult. The modeller must now consider the entire conditional joint distribution of defaults among the obligors. Although there is a growing literature on modelling correlated defaults (Vasicek (1987, 1991), Li (2000), Duffie and Garleanu (2001), Davis and Lo (2001), amongst others), this research area remains fertile ground; some of the papers cited above and a substantial proportion of the literature has studied static, one-period models and the study of dynamic models is an emerging field. This point it discussed further in Chapter 4.

Furthermore, even in asset pricing research in which expected cash flows are not explicitly modeled, such as when regressing excess returns to a security on the SDF (or, perhaps more commonly, portfolios mimicking the SDF), or variables that proxy for the conditional risk premium, one assumes that the market participants that set prices have done so with rational expectations. By this, I mean that they have, at a minimum,
used the data generating process to calculate expected cash flows. However, in recent times very high default rates have been observed in the loan pools that underly financial securities such as collateralised debt obligations (CDOs). Further, these securities had often been issued credit ratings that reflect a low default probability and priced accordingly. Although it is not necessarily the case, it is certainly conceivable that the market participants themselves had miscalculated the expected cash flows associated with large loan pools. For example, Adelino (2009) finds that in the mortgage-backed securities market the prices of triple-A rated securities did not have predictive power for future performance.

The above discussion suggests that, in the case of portfolio credit risk, it is possible to proceed cautiously; rather than jump to the task of studying risk premia on CDOs, there is a substantial enough gap to first contribute to the development of better models of correlated default and the calculation of expected cash flows. A dynamic model is also able to calculate conditional expected cash flows. Chapter 4 is work in this direction.

The model developed in Chapter 4 calculates the distribution of losses on a portfolio conditional on a set of macroeconomic observable and unobservable state variables. In principle, one could separately calculate the prices of the relevant observable macroeconomic risks by studying the pricing of other securities that depend on these same variables. Although this task is left to further research, it is one direction in which the model presented in Chapter 4 has the potential to become a dynamic pricing model for securities such as CDOs.

Another reason for studying conditional expected cash flows and the conditional distribution of losses to a portfolio of credit risks is to understand the credit cycle. In particular, one may wish to relate the distribution of losses to a loan portfolio to the state of the macroeconomy. This has economic implications for wider economic performance and growth since credit downturns affect the capital adequacy requirements of banks and hence their willingness to extend new credit. The model presented in Chapter 4 is a contribution in this direction.

1.3.2 Relating Portfolios of Defaultable Debt and Defaultable Debt

Of course, a model of credit risk is a special case of a portfolio credit risk model where the size of the portfolio is one. Indeed, the portfolio credit risk model that is developed in Chapter 4 simplifies to a form that is very similar to the models that were used to...
study common economic factors that drive individual defaults in Duffie, Saita, and Wang (2007) and Duffie, Eckner, Horel, and Saita (2009). This is a desirable feature. However there is the key difference that idiosyncratic/obligor specific factors that drive defaults are not included. This feature was sacrificed so that model could remain econometrically tractable and implementable given data on aggregate default rates only. This type of data is more commonly available than data on individual defaults. Nevertheless, this issue and extensions of the modelling framework to cases in which obligor specific factors can be included are discussed in Chapter 4.
Chapter 2

A GDTSM with Risk Premia Linear in Forward Rates

Chapter Summary

Recent empirical evidence suggests that excess returns to bonds are forecastable by a linear combination of forward rates (Cochrane and Piazzesi (2005)). Here, I construct a tractable arbitrage free dynamic term structure model (DTSM) with prices of risk that incorporate this predictability. Using a novel method to initialise the maximum likelihood procedure, the estimated model improves on the out-of-sample forecasting of the most important principle component of bond yields, the level component, compared to an unrestricted VAR(1). Given recent insight on the relationship between Gaussian DTSMs and the VAR(1) on principle components (Joslin, Singleton, and Zhu (2010)), the forecasting improvements provide some support for the restrictions on prices of risk imposed here.
2.1 Introduction

Bond yields are often studied using dynamic term structure models (DTSMs) - models that describe the co-movement over time of the entire yield curve. The Gaussian affine DTSM (GDTSM henceforth) is commonly applied in such studies. These are models in which the short rate, \( r_t \), is an affine function of a state vector \( X_t \): 
\[
    r_t = \delta_{0,X} + \delta_T^T X_t,
\]
and \( X_t \) follows a Gaussian affine diffusion under both the risk-adjusted (Q) and objective (P) probability measures. Amongst several others, Duffee (2002), Ang and Piazzesi (2003) and Chernov and Mueller (2008) have explored the forecasting performance of GDTSMs.

Recently, Joslin, Singleton, and Zhu (2010) (JSZ henceforth) have provided new insights into unrestricted GDTSMs (GDTSMs in which market prices of risk are left as free parameters). Loosely speaking, JSZ show that, regardless of the constraints imposed on the Q distribution of \( X_t \), the forecasts implied by a GDTSM (estimated by maximum likelihood (ML)) and an unrestricted first order vector autoregression (VAR(1)) (estimated by ordinary least squares (OLS)) are identical. This means that, once given data on \( n \) accurately measured yields, then data on the shape of the remainder of the yield curve cannot in any way contribute to the model’s forecasting ability. Simply put, imposing no-arbitrage alone does not help in forecasting.

For researchers aiming to forecast future yields using a GDTSM, the above mentioned findings pose the important question, can the no-arbitrage restriction be used at all to improve yield forecasts? JSZ show formally that restrictions on market prices of risk (or, equivalently, on the P dynamics of \( X_t \)) are able to increase the efficiency of yield forecasts compared with the VAR(1). The intuition behind their result can be understood as follows.

Consider, as an illustrative example, an economy in which interest rates are described by a GDTSM in which the local expectations hypothesis (LEH) holds. Under the LEH, tight restrictions are imposed on market prices of risk; they are zero. Investors are neutral to interest rate risk and so risk premia to long bonds and short bonds are both zero. An upward (downward) sloping yield curve must forecast rises (falls) in future short yields because that is the only way that high (low) yielding long bonds can have expected returns equal to those of a low (high) yielding short bonds over a given holding period. The key point to note is that, due to the restrictions on market prices of risk of the LEH, the arbitrage free cross section of yields has provided information about the time series of yields and hence we have the result mentioned above.
Although, in principle, imposing the LEH on a GDTSM allows the researcher to use no-arbitrage for forecasting, it has long been established that the restrictions on market prices of risk that are imposed by the LEH are empirically rejected (Fama and Bliss (1987), Campbell and Shiller (1991)). Such a GDTSM therefore provides poor yield forecasts. The actual forecasting efficiency of a GDTSM can only be increased if the restrictions on market prices of risk are consistent with the empirical facts relating to bond excess returns. The main purpose in this chapter is to construct such a model. I now introduce some of this empirical evidence.

The classic regressions of Fama and Bliss (1987) provide evidence that the one year excess return to the $n$-year bond can be forecast using the spread between the $n$-year forward rate and the one year yield with an $R^2$ of 18%. Campbell and Shiller (1991) find related results when forecasting yield changes with yield spreads. More recently, Cochrane and Piazzesi (2005) (CP henceforth) were able to considerably improve on this predictability. They find that a single factor, a tent-shaped linear combination of forward rates, forecasts excess returns to bonds of one to five year maturity with an $R^2$ of 44%.

The GDTSM in this chapter is motivated by the regressions of CP; risk premia to all bonds are determined by a linear combination of forward rates. The reason that it is possible to accommodate this predictability into a GDTSM can be understood as follows.

It is easily shown that forward rates in a GDTSM are affine in the state vector. Further, a Gaussian essentially affine DTSM (first described by Duffee (2002)) allows market prices of risk to be affine in the state vector. Comparing the forward rate equation and the form of market prices of risk allows one to choose parameters such that market prices of risk (and hence risk premia) depend on forward rates. These ideas are made formal in Section 2.2 below.

Imposing the restrictions on market prices of risk just described is intended to improve the GDTSM’s forecasting. Given the results of JSZ, it is therefore natural to take the forecasts of a VAR(1) estimated by OLS as a benchmark for comparison. I find that the restrictions imposed here can improve out-of-sample forecasts.

The discussion so far has centered on how, via restrictions on market prices of risk in a GDTSM, the yield curve may improve yield forecasting. The other side of the same coin is that such restrictions can improve our understanding of the yield curve by making its decomposition into expectations and risk premium components more efficient.
Restrictions on market prices of risk link the time series (P dynamics) and cross section (Q dynamics) of bond yields and so the time series of bond yields helps estimate the Q parameters. Indeed, in estimating a GDTSM with restrictions on prices of risk by ML, the JSZ separation of the likelihood function into two parts that depend on only either the Q or P parameters is no longer generally possible; all parameters are estimated simultaneously.

Given the above, another aim in this chapter is to understand risk premia in bond returns. The approach taken is to apply appropriate invariant transformations to the GDTSM’s parameters (as described in Dai and Singleton (2000)) so that each element in the state vector is a principle component of bond yields. The first three principle components (in order of decreasing variance/eigenvalues) are commonly labelled level, slope and curvature for the respective effects of shocks to them on the shape of the yield curve. The GDTSM is then estimated by ML. The approach of transforming to a GDTSM with principle components as the state vector has the following two main advantages.

One immediate advantage is in economic interpretation. It is both intuitive and now commonplace in the literature to decompose the variation in the yield curve into level, slope and curvature components; thinking of movements in the entire yield curve as being driven mainly by a small number of factors derives from the classic analyses of Knez, Litterman, and Scheinkman (1994) and Litterman and Scheinkman (1991). If, for example, risk to the level factor carries a high price, then this is perhaps more easily interpreted than risk to a latent factor (or even a particular yield) carrying a price. This is due to both the easy interpretation of the level factor, and the fact that this factor is orthogonal to other factors, meaning that the contributions to risk premia due to the different priced risks are neatly separated.

A second advantage of using the principle components (or, indeed, any n observable linearly independent combinations of bond yields) as the state vector is in econometric implementation. Despite the restrictions that are imposed on market prices of risk in the GDTSM estimated here, there are still a relatively large number of free parameters to optimise over in the ML procedure (17 in a 3 factor model). This can make it difficult to ensure that estimates resulting from the ML procedure are globally optimal.

One method by which researchers increase the chance that their ML procedure converges to the global optimum is to initialise the optimisation at many different locations and generate parameter estimates at each convergence. They then take as their ML estimates
those that correspond to the largest likelihood. This approach may be improved if there is some guidance as to suitable initialisation points. If the state vector is observable, then the parameters of an unrestricted VAR(1), quickly and consistently estimated by OLS, may provide a good initialisation point for some of the parameters in the model (means and mean reversion rates). Indeed, in the case of an unrestricted GDTSM, JSZ show that the OLS estimates are globally optimal. Subsection 2.4.6 devotes considerable attention to the initialisation problem. I show how one can choose suitable initialisation points for all the parameters in the model and simultaneously, via invariant transformations, maintain econometric identification.

More generally, imposing restrictions on market prices of risk carries the two econometric advantages that (i) there are fewer parameters to estimate and (ii) some of the parameters that are most difficult to estimate (those that affect drifts) have been removed. If the GDTSM is correctly specified then restrictions on prices of risk increase the efficiency of parameter estimates.

The remainder of this chapter proceeds as follows. I begin with a discussion of the related literature. I continue in Section 2.2 with a brief recapitulation of the bond pricing equation for affine GDTSMs provided by Duffie and Kan (1996) and I discuss the essentially affine form for market prices of Duffee (2002) that lie central to the GDTSM of this chapter. Section 2.3 specifies prices of risk such that risk premia to all bonds depend on a linear combination of forward rates. Next, Section 2.4 discusses the estimation problem, including transformations to an observable state vector and the ML procedure. Section 2.5 presents parameter estimates and measures of the GDTSM’s forecasting. Finally, Section 2.6 concludes.

2.1.1 Further Related Literature

The model presented in this chapter is most closely related to Cochrane and Piazzesi (2008), who also construct a GDTSM motivated by their earlier findings in CP. The main difference between their approach and that taken in this chapter is in implementation. These authors take the level, slope and curvature principle components as elements of the state vector, but also take a fourth factor: the linear combination of forward rates that drive market prices of risk. Their SDF allows only the level principle component to be priced and their model is chosen to exactly match the regressions in CP.

Duffee (2002) and Kim and Wright (2005) have estimated 3-factor unrestricted GDTSMs. Their models therefore nest the 3-factor GDTSM used here. One task in this chapter is
to compare the forecasting of the GDTSM with its unrestricted counterpart.

More generally, any DTSM consists of three components: (i) the dependence of the short rate of interest on the state vector, (ii) the dynamics of the state vector under the Q probability measure and (iii) a specification of market prices of risk (the SDF).

Points (i) and (ii) have been well studied (Subsection 1.1.1 discussed the relevant literature). Focussing on (i) and (ii) is suitable for pricing derivatives on bonds where modelling bond volatility rather than expected return is of most importance. It is also useful for drawing smooth yield curves across maturities to either price bonds of non-traded maturities or, if the modeller believes strongly enough in the model, to arbitrage away any bonds that do not lie on the curve drawn. However, this approach is not suitable for bond portfolio analysis. For this, point (iii) must be addressed.

The more recent literature has focussed on point (iii). Duffee (2002), Dai and Singleton (2002) and Duarte (2004) are notable early contributions towards understanding bond prices in both the time series and cross section. The GDTSM presented in this chapter is also a contribution in this direction.

2.2 Affine Dynamic Term Structure Models

This section provides a brief recapitulation of the key results relating to the prices of bonds in the cross section and time series implied by GDTSMs. These results are subsequently used in Section 2.3 to construct the model central to this chapter: a GDTSM where risk premia to bonds depend on a linear combination of forward rates.

2.2.1 Affine Bond Pricing

Risk is generated in the economy by \( n \) Q-Brownian motions \( W^Q_t \equiv (W^Q_{t,1}, \ldots, W^Q_{t,n})^T \).

The \( n \)-dimensional state vector is denoted \( X_t \equiv (X_{t,1}, \ldots, X_{t,n})^T \) and is driven by \( W^Q_t \).

The short rate of interest, \( r_t \), is an affine function of the state vector:

\[
r_t = \delta_{0,X} + \delta_X^T X_t.
\]

(2.1)

Here \( \delta_{0,X} \) is a scalar and \( \delta_X \) is an \( n \)-vector.

Throughout this chapter, the \( X \) subscript to parameters denotes that they enter the model in the form given in (2.1) (and below in (2.2) and (2.21)) when the state vector is \( X_t \). This notation is useful later when I make invariant transformations to different state
vectors. For example, if $Y_t$ is used as the state vector then I write $r_t = \delta_{0,Y} + \delta^Q Y_t$ and, of course, $\delta_{0,Y}$ and $\delta_Y$ are related to $\delta_{0,X}$ and $\delta_X$ in a way that depends on the relation between $Y_t$ and $X_t$. I return to this topic in Subsection 2.3.4. A similar approach to notation is taken with functions and matrices that depend on parameters.

GDTSMs impose that the Q evolution of the state vector is described by

$$dX_t = K^Q_X \left( \theta^Q_X - X_t \right) dt + \Sigma_X dW^Q_t,$$

(2.2)

where $K^Q_X$ and $\Sigma_X$ are $n \times n$ matrices and $\theta^Q_X$ is an $n$-vector.

Duffie and Kan (1996) show that, in this framework, the time $t$ price of the time $T$ maturity zero-coupon bond is given by

$$P_t(u) = P_X(X_t, u) \equiv e^{A_X(u) - B_X(u)^T X_t}.$$

(2.3)

Here, $u \equiv T - t$, $A_X(u)$ is a scalar function and $B_X(u)$ is an $n$-valued function. These are the solutions to the system of Riccati ordinary differential equations (ODEs), provided in Subsection 2.7.1 in the appendix to this chapter. Given (2.3), the bond yield defined by

$$y_t(u) \equiv \frac{-1}{u} \ln P_t(u)$$

(2.4)

is given by

$$y_t(u) = \frac{1}{u} (-A_X(u) + B_X(u)^T X_t).$$

(2.5)

Finally, letting $f_t(u)$ denote the time $t$ instantaneous forward rate that prevails at time $t + u$, it is a standard no-arbitrage result that

$$f_t(u) = -\frac{\partial}{\partial u} \log P_X(X_t, u) = -\frac{\partial A_X(u)}{\partial u} + \frac{\partial B_X(u)^T}{\partial u} X_t.$$

(2.6)

[2.3] provides the time $t$ conditional prices of the cross section of bond maturities. However, we are yet to make a statement about conditional expected returns to bonds in the time series. For this we must specify the market prices of risk.
2.2.2 Expected Returns to Bonds

The SDF, denoted by \( \pi_t \), follows the process

\[
\frac{d\pi_t}{\pi_t} = -r_t dt - \Lambda_t^T dW^P_t,
\]

where \( W^P_t \equiv (W^P_{t,1}, \ldots, W^P_{t,n})^T \) is an \( n \)-vector of P-Brownian motions and \( \Lambda_t \) is a column \( n \)-vector specifying market prices of risk. Applying Girsanov’s theorem to (2.2) gives the dynamics of the state vector under the P measure as

\[
dX_t = K_X^Q \left( \theta_Q^X - X_t \right) dt + \Sigma_X \Lambda_t dt + \Sigma_X dW^P_t.
\] (2.7)

The bond pricing function solves

\[
\frac{dP_X(X_t, u)}{P_X(X_t, u)} = (r_t + e_t(u))dt + v_t(u)^T dW^P_t,
\] (2.8)

where \( e_t(u) \) is a scalar that denotes the time \( t \) instantaneous risk premium to holding a bond that matures at time \( t + u \). \( v_t(u) \) is an \( n \)-vector that determines the volatility of this bond. An application of Itô’s lemma to (2.3) and using (2.7) reveals that these two quantities are given by

\[
e_t(u) \equiv -B_X(u)^T \Sigma_X \Lambda_t \quad \text{and} \quad v_t(u)^T \equiv -B_X(u)^T \Sigma_X.
\] (2.9)

The GDTSM is not yet fully specified; (2.9) shows that we must choose a \( \Lambda_t \) to determine risk premia. The choice of form for \( \Lambda_t \) is delicate. It must be sufficiently flexible to ensure that the model is able to capture the key feature of the data that we wish to expose (that expected excess returns to bonds are linear in forward rates) but also restricted enough so that the model remains econometrically tractable. An essentially affine specification, given in the next subsection, is suitable for this task.

2.2.3 Essentially Affine Market Prices of Risk

When the state vector follows a Gaussian affine process, the essentially affine specification for the market price of risk vector is

\[
\Lambda_t \equiv \lambda_{0,X} + \lambda_{1,X} X_t,
\] (2.10)

where \( \lambda_{0,X} \) and \( \lambda_{1,X} \) are an \( n \) dimensional column vector and \( n \times n \) matrix of parameters respectively. The \( X \) subscript is used because, in the GDTSM here, these quantities
depend on the parameters that govern $X_t$. This will become clear below.

The equations for the short rate, the Q-dynamics and the market prices of risk in terms of the state vector, given in (2.1), (2.2) and (2.10) respectively, completely specify the GDTSM in the sense that the statistical properties of bond prices in both the time series and in the cross section of bond maturities are pinned down. The aim is to make risk premia depend on a linear combination of forward rates. By inspecting (2.10) and the equation for forward rates in (2.6) it is immediately apparent that, with appropriate choices of $\lambda_{0,X}$ and $\lambda_{1,X}$, it is possible to ensure this aim. The next step is to formalise this observation.

2.3 Model Specification

CP provide motivation for choosing $\Lambda_t$ such that risk premia to all bonds depend on a linear combination of forward rates. The main purpose in this section is to choose such a $\Lambda_t$.

2.3.1 Prices of Risk

Recall that there are $n$ state variables. Define the following $n$-vector

$$
e_t \equiv (e_t(u_1), \ldots, e_t(u_n))^T. \tag{2.11}
$$

Here, $e_t$ denotes the risk premia to $n$ bonds with times to maturity of $u_1, \ldots, u_n$ respectively. The meaning of $e_t(u_i)$ can be understood from (2.8).

In the model, risk premia to the set of $n$ bonds that enter $e_t$ (those with times to maturity of $u_1, \ldots, u_n$) and also all other bonds will depend on a linear combination of forward rates. However, the $n$ bonds that enter $e_t$ are special in the sense that their risk premia can be specified independently of one another and explicitly in terms of free parameters that enter the GDTSM. Given the risk premia to the bonds in $e_t$, risk premia to all other bonds are then determined by the no-arbitrage relationship in (2.9). This idea is made formal below, but the intuition as to why the risk premia to exactly $n$ bonds can be specified explicitly and independently of one another in a model with $n$ state variables can be understood as follows.

Without loss of generality, one can let the $n$ state variables in the GDTSM be $n$ orthogonal observable combinations of yields (principle components); based on Dai and Singleton (2000), Subsection 2.4.3 makes transformations that allow one to change from
latent to observable combinations of yields as state variables whilst leaving the short rate and hence bond prices unchanged. Since the principle components are orthogonal to one another, there is no reason that the risk premium associated with shocks to each of them must be related to each other. In other words, we must be able to specify the risk premia to these \( n \) combinations of bonds separately of one another.

The next step is to write \( e_t \) in terms of forward rates. Let \( f_t \) denote an \((m + 1)\)-vector that consists of a constant and the \( m \) forward rates that will drive risk premia. Recalling from (2.6) that \( f_t(v) \) denotes the time \( t \) instantaneous forward rate that prevails at time \( t + v \), I define

\[
f_t \equiv [1 \ f_t(v_1) \ \ldots \ f_t(v_m)]^T.
\]

Note that \( m \) is allowed to be different from \( n \). Making risk premia to the \( n \) bonds that enter \( e_t \) linear in forward rates requires that

\[
e_t = b\gamma^T f_t.
\]  

(2.12)

Here \( b \) and \( \gamma \) are \( n \times 1 \) and \((m + 1) \times 1\) vectors of parameters respectively. Clearly, for the purposes of model estimation, the \( b \) and \( \gamma \) are not separately identified and so appropriate normalisations are imposed (see Section 2.4). The scalar, \( \gamma^T f_t \), drives risk premia and the \( i \)-th entry in the vector \( b \), \( b_i \), gives the loadings of the risk premium of the \( u_i \) maturity bond on this factor. Therefore, as discussed above, the risk premia to the \( n \) bonds in \( e_t \) can be determined separately to each other. However, risk premia to bonds of all maturities and not just the \( u_1, \ldots, u_n \) maturity bonds have been determined via (2.9); a fully specified DTSM describes the co-movement over time of the entire yield curve.

The parameters \( b \) and \( \gamma \) do not have \( X \) subscripts because these parameter do not change under invariant transformations of the GDTSM’s state vector. The reason is clear; expected returns to bonds should be the same in all observationally equivalent models, regardless of the choice of state vector.

Simple algebraic manipulation of (2.9), (2.11) and (2.12) then gives the market price of risk vector as

\[
\Lambda_t = -\Sigma^{-1}_X B^{-1}_X b\gamma^T f_t,
\]

(2.13)

where

\[
B_X \equiv (B_X(u_1), \ldots, B_X(u_n))^T.
\]  

(2.14)
Equation (2.13) assumes that $B_X$ is invertible. The invertibility of $B_X$ is guaranteed by assuming that no two bonds yields in the set of bonds with maturities $u_1, \ldots, u_n$ are proportional to each other. This is clear from (2.5); if no two yields are proportional to each other then the rows of $B_X$ must be linearly independent.

2.3.2 Forward Rates in Terms of the State Vector

The elements of vector $f_t$ are all affine in the state variables. This can be seen from the standard no-arbitrage result that

$$f_t(u) = -\frac{\partial}{\partial u} \log P_t(u)$$  \hspace{1cm} (2.15)

and an application of (1.2) to give

$$f_t(u) = -\frac{\partial A_X(u)}{\partial u} + \frac{\partial B_X(u)^T}{\partial u} X_t.$$  \hspace{1cm} (2.16)

We can then define an $(m + 1) \times 1$ vector $F_{0,X}$ and an $(m + 1) \times n$ matrix $F_{1,X}$ such that

$$f_t = F_{0,X} + F_{1,X} X_t.$$  \hspace{1cm} (2.17)

In the above equation,

$$F_{0,X} \equiv -\left[ -1 \left| \frac{\partial A_X(s)}{\partial s} \right|_{s=v_1} \cdots \left| \frac{\partial A_X(s)}{\partial s} \right|_{s=v_m} \right]^T,$$  \hspace{1cm} (2.18)

and

$$F_{1,X} \equiv \begin{pmatrix} 0_n^T \\ \frac{\partial B_X(s)^T}{\partial u} \bigg|_{s=v_1} \\ \vdots \\ \frac{\partial B_X(s)^T}{\partial u} \bigg|_{s=v_m} \end{pmatrix}.$$  \hspace{1cm} (2.19)

Here, $0_n$ is an $n$-vector of zeros.

Given (2.17), the price of risk vector in (2.13) can then be written in the essentially affine form given in (2.10). The quantities $\lambda_{0,X}$ and $\lambda_{1,X}$ are

$$\lambda_{0,X} = -\Sigma^{-1}_X B^{-1}_X b \gamma^T F_{0,X},$$  \hspace{1cm} (2.20)

$$\lambda_{1,X} = -\Sigma^{-1}_X B^{-1}_X b \gamma^T F_{1,X}.$$  \hspace{1cm}

The GDTSM is now completely specified. The short rate of interest and Q-dynamics of
the state vector were given in (2.1) and (2.2) respectively. The state vector’s P-dynamics follow from (2.7), (2.10) and (2.20):

\[ dX_t = K^P_X (\theta^P_X - X_t) \, dt + \Sigma_X dW^P_t, \]

(2.21)

where

\[ K^P_X \equiv K^Q_X - \Sigma_X \lambda_{1,X}, \text{ and} \]

\[ \theta^P_X \equiv (K_X^P)^{-1}(K_X^Q \theta_X^Q + \Sigma_X \lambda_{0,X}). \]

(2.22)

(2.23)

2.3.3 Calculation of \( F_{0,X} \) and \( F_{1,X} \)

An important practical point to note is that the quantities \( F_{0,X} \) and \( F_{1,X} \) (and hence also \( \lambda_{0,X} \) and \( \lambda_{1,X} \)) are particularly easy to calculate. Once the functions \( A_X(u) \) and \( B_X(u) \) that determine the yield curve are known at \( v_1, \ldots, v_m \), then the ODEs in Equations (2.33) and (2.34) (see Subsection 2.7.1 in the appendix) can be used to immediately provide the derivatives that enter \( F_{0,X} \) and \( F_{1,X} \). Numerical differentiation is not required.

2.3.4 Parameter Vectors, Invariant Transformations and Notation

The final task in this section is to characterise GDTSMs by a parameter vector and to describe invariant transformations of these parameters to observationally equivalent GDTSMs. These are important for the estimation methods that follow in Section 2.4 below.

A GDTSM can be completely characterised (in that the relative prices of all bonds in the cross section and their statistical properties in time series are determined) by its parameter vector. I use the notation

\[ \psi_X = (\delta_{0,X}, \delta_X, K^Q_X, \theta^Q_X, \Sigma_X, K^P_X, \theta^P_X) \]

to characterise a GDTSM with state vector \( X_t \). (2.1), (2.2) and (2.21) showed how these parameters enter the model. One can characterise another GDTSM with state vector \( Y_t \) with parameter vector \( \psi_Y = (\delta_{0,Y}, \delta_Y, K^Q_Y, \theta^Q_Y, \Sigma_Y, K^P_Y, \theta^P_Y) \) and equations analogous to (2.1), (2.2) and (2.21), but with parameters as given in \( \psi_Y \). If \( Y_t = C + DX_t \) then, for the GDTSMs characterised by \( \psi_Y \) and \( \psi_X \) to be observationally equivalent, the elements of \( \psi_X \) and \( \psi_Y \) must be related by an invariant transformation. The invariant transformation is provided in Subsection 2.7.2 in the appendix.
Finally, functions that depend explicitly on $\psi_X$ are subscript by $X$. For example, if $\psi_X$ and $\psi_Y$ are related by an invariant transformation then

$$y_t(u) = \frac{1}{u}(-A_X(u) + B_X(u)^T X_t) = \frac{1}{u}(-A_Y(u) + B_Y(u)^T Y_t).$$

Here, $A_X(u)$ and $B_X(u)$ depend explicitly on parameters in $\psi_X$ and $A_Y(u)$ and $B_Y(u)$ depend explicitly on parameters in $\psi_Y$. These dependencies are made clear in Subsection 2.7.1 in the appendix.

### 2.4 Estimation

#### 2.4.1 Forward Rates in the Price of Risk

Before estimating the model, we must choose which forward rates drive risk premia. That is, we must choose $v_1, \ldots, v_m$. I set $m = n$ and for $i = 1, \ldots, n$, $v_i = u_i$. Recalling that the $u_i$ are the bond maturities that enter $e_t$ (see Subsection 2.3.1), the forward rates that drive risk premia correspond to the bond maturities in $e_t$. There are three further points to note here.

First, the case $m > n$ has been excluded. Although, in principle, we can use $m > n$, it should be noted that in an $n$ state variable model, once we have $n$ different forward rates, the entire state of the economy is known. Additional forward rates, therefore, do not carry any additional information. It therefore seems reasonable to restrict risk premia to depend on a maximum of $n$ forward rates. The real restriction is not that only $n$ forward rates are used; it is that risk premia are forced to depend on a linear function of these forward rates.

Second, the case $m < n$ has not been excluded. Such models are nested by the GDTSM estimated here (simply set elements of $\gamma$ that correspond to the forward rates to be excluded to zero). If certain forward rates are irrelevant for calculating risk premia, we expect that their coefficient in $\gamma$ we be insignificantly different from zero.

Third, the imposition that the forward rates that drive risk premia correspond to the bond maturities in $e_t$ is not necessary. However, this choice is motivated by the fact that CP’s results were obtained taking a similar approach, with the main difference being that they use one year rather than instantaneous forward rates.
2.4.2 Canonical Representation

The first step is to apply appropriate normalisations so that the GDTSM specified in the previous section is econometrically identified. Dai and Singleton (2000) provide “canonical” normalisations for affine DTSMs. These are normalisations to the Q-parameters that ensure econometric identification whilst remaining “maximally flexible” in the sense that the minimal known sufficient conditions for econometric identification are imposed. In their notation, the GDTSM in this chapter is $A_0(n)$.

The $A_0(n)$ canonical representation requires $K^Q_X$ to be a lower (or upper) triangular matrix, $\Sigma_X = O$ and $\theta^Q_X = 0_n$. Here, $0_n$ is an $n$-vector of zeros and $O$ denotes any $n \times n$ orthonormal matrix ($O O^T = I_n$). The $A_0(n)$ GDTSM is characterised by

\[
\psi_X = (\delta_{0,X}, \delta_X, K^Q_X, 0_n, O, K^P_X(b, \gamma), \theta^P_X(b, \gamma)) \quad \text{ (2.24)}
\]

The notation $K^P_X(b, \gamma)$ and $\theta^P_X(b, \gamma)$ is to emphasise that $K^P_X$ and $\theta^P_X$ are both determined once the vectors $b$ and $\gamma$ (and, of course, $K^Q_X$) are known. This is clear from (2.21); $\lambda_{0,X}$ and $\lambda_{1,X}$ depend on $b$ and $\gamma$. Further, to ensure that $b$ and $\gamma$ are separately identified I impose that $\frac{1}{n} \sum_i b_i = 1$.

The total number of free parameters is $2 + 3n + \frac{1}{2} n(n + 1)$; 17 if $n = 3$. This compares with 22 free parameters in the unrestricted counterpart GDTSM. The 5 fewer parameters relative to the unrestricted GDTSM are all from the price of risk vector and are therefore parameters that are otherwise difficult to estimate precisely in the sense that the likelihood function is relatively insensitive to changes in their values. The free parameters are estimated using time series data on the cross section of yields (the data is described in Subsection 2.4.7 below) and an ML approach.

It is straightforward to write the likelihood function when the state vector in the model is a set of yields or portfolios of yields. Also, an observable state vector provides clues as to an appropriate initialisation point for the ML procedure (see Subsection 2.4.6). For these reasons (and others relating to the ease of economic interpretation that were discussed in Section 2.1) it is useful to transform the model so that the state vector is an observable portfolio of yields itself, rather than latent factors. This is the topic of the next subsection.
2.4.3 Invariant Transformation to Observable States

There are \( p \) bonds in the data set. We can stack yields to these \( p \) bonds into the \( p \)-vector \( y_t \):

\[
y_t \equiv [y_t(w_1), \ldots, y_t(w_p)]^T.
\]

Here, the \( w_1, \ldots, w_p \) refer to the bond maturities in the data set (see Subsection 2.4.7).

For any full rank \( n \times p \) portfolio matrix \( W \), let \( P_t \equiv Wy_t \) denote the associated \( n \)-dimensional set of portfolios of yields. Here the \( i \)-th portfolio puts weight \( W_{i,j} \) on the \( w_j \) maturity yield. Below, \( W \) is chosen so that \( P_t \) is a vector of principle components of bond yields but the results in this subsection apply for any \( W \).

The main aim here is to apply invariant transformations to \( \psi_X \) to obtain an observationally equivalent GDTSM in which the observable vector, \( P_t \), replaces the latent vector, \( X_t \), as the state. That is, we wish to find \( C \) and \( D \) such that \( P_t = C + DX_t \), and then calculate

\[
\psi_P = (\delta_0, \delta_P, K^Q_P, \theta_P^Q, \Sigma_P, K^P_P, \theta_P^P)
\]

via invariant transformations. However, we must preserve econometric identification. The difficulty is that the normalisations imposed in the canonical form above (Equation (2.24)) cannot, in general, be imposed when the state vector is observable. The approach taken is therefore to (i) calculate a \( \psi_P \) in terms of \( \psi_X \) and (ii) search over \( \psi_X \) in the ML procedure. The remainder of this subsection is devoted to point (i). I follow Dai and Singleton (2000) in this task. Point (ii) is discussed in more detail in Subsection 2.4.5 below.

The yield, \( y_t(u) \), was given in terms of the state vector in (2.5). Applying this equation, \( y_t \) can be written in terms of the state vector as

\[
y_t = -A + BX_t.
\]

Here, \( A \equiv \left( \frac{A_X(w_1)}{w_1}, \ldots, \frac{A_X(w_p)}{w_p} \right)^T \) and \( B \equiv \left( \frac{B_X(w_1)}{w_1}, \ldots, \frac{B_X(w_p)}{w_p} \right)^T \). Pre-multiplying the above equation by the portfolio matrix \( W \) gives

\[
P_t = -WA + WBX_t, \tag{2.25}
\]

Given \( A, B, W \) and \( \psi_X \), the results in Subsection 2.7.2 in the appendix are used to calculate \( \psi_P \). Further, since \( A \) and \( B \) are calculated using only elements in \( \psi_X \) and the exogenously chosen matrix \( W \), if we are given \( \psi_X \), we can calculate \( \psi_P \). We now have
an observationally equivalent GDTSM with $P_t$ as the state vector, as required.

Given $\psi_X$ (and hence $\psi_P$), model implied yields can then be written as

$$y_t(u) = \frac{1}{u}(-A_P(u) + B_P(u)^TP_t).$$  \hspace{1cm} (2.26)

Here, $A_P(u)$ and $B_P(u)$ are scalar and $n$-vector functions that are the solutions to the Ricatti ODEs given in Subsection 2.7.1 in the appendix, calculated under $\psi_P$.

The next task is to choose the portfolio matrix, $W$.

### 2.4.4 Choice of $W$ and Stochastic Singularity

We wish to choose $W$ so that $P_t$ is a vector of principle components of bond yields. However, the choice of $W$ is related to the stochastic singularity difficulty that arises in estimating DTSMs. This can be understood as follows.

Given $n$ portfolios of precisely measured yields at any time, $P_t$, the GDTSM provides the entire yield curve. Any additional yields are predicted by the model with an $R^2$ of 1. The model is therefore rejectable by any precisely measured additional yield that does not lie on the $P_t$ conditional yield curve. In the model, such a yield would represent an arbitrage opportunity. This is the stochastic singularity problem (see chapter 12 of Ait-Sahalia and Hansen (2009) for a fuller discussion).

In light of this, the technique used here is to follow Duffee (2002) and others in assuming that $n$ (of the $p > n$) yields are measured without error and the remaining yields are measured with error. Given the use of $P_t$ as the state variable, we would like $P_t$ to be measured without error. This is possible if $W$ is chosen only to depend on the $n$ yields that are measured without error. I make such a choice. Further, I assume that all additional yields (not lying on the $P_t$ conditional yield curve) are measured with Gaussian and serially uncorrelated errors. The remainder of this subsection formalises these ideas and discusses choices for $W$.

I assume that at each time $t$ the observed data is a $p$-vector $y_t^o$ for which the following holds:

$$y_t^o = y_t + \left( \begin{array}{c} 0_n \\ \epsilon_t \end{array} \right)$$  \hspace{1cm} (2.27)

Here, $\epsilon_t$ is a $(p - n)$-vector with distribution $\epsilon_t \sim \mathcal{N}(0, \sigma^2 I_{p-n})$ and the property that
\[ E[\epsilon_t \epsilon_{t+j}^T] = 0_{(p-n) \times (p-n)} \text{ for } j \neq 0. \]  
\( \alpha^2 \) is a parameter that is estimated jointly with the other model parameters. The \( p \)-yields in the data are arranged so that the \( n \) that are measured without error are placed in the first \( n \) positions in \( y_t \) and in order of maturity, with the shortest maturity first (there is no loss of generality here).

The next step is to choose a \( W \) such that \( P_t \) depends only on the first \( n \) yields. That is, \( W_{i,j} = 0 \) for \( j > n \). There are, of course, many possible choices for \( W \). The idea, therefore, is to choose \( W \) so that the elements of the state vector \( P_t \) have an intuitive economic interpretation.

Perhaps the simplest option is to choose \( W \) such that \( P_t \) is a vector of the first \( n \) yields (\( W_{ij} = 1 \) if \( i = j \), and \( W_{ij} = 0 \) otherwise). However, it has now become standard fare in the term structure literature to think about the majority of the variation in the yield curve as being driven by three principle components: level, slope and curvature. Here, I follow papers such as JSZ and take \( P_t \) to be \( n \) principle components. I calculate \( W \) as follows.

Letting \( V_Y \) denote the \( n \times n \) sample covariance matrix of the first \( n \) yields in \( y_t^o \) (those that are measured without error), its eigenvalue decomposition is

\[ V_Y = Q_Y D_Y Q_Y^T, \]

where \( Q_Y \) is an \( n \times n \) orthonormal matrix with columns that are the eigenvectors of \( V_Y \), and \( D_Y \) is a matrix of eigenvectors. Setting

\[ W \equiv \begin{bmatrix} Q_Y^T & 0_{n \times (p-n)} \end{bmatrix} \quad (2.28) \]

means that \( P_t \) consists of the \( n \) principle components of the \( n \) well measured yields. The calculated \( W \) used in this chapter is given in the Section 2.7.4 in the appendix. The next step is to write the likelihood function.

**2.4.5 Conditional Likelihood Function and ML Procedure**

The conditional likelihood function is given by

\[ f(y_t^o|y_{t-\Delta}^o; \psi_P, \sigma^2) = f(y_t^o|P_t; K_P^Q, \theta_P^Q, \Sigma_P, \delta_0, \delta_P, \sigma^2) \times f(P_t|P_{t-\Delta}; K_P, \theta_P^P, \Sigma_P), \quad (2.29) \]

where \( \Delta \) denotes the temporal spacing of the data. The forms of the two functions in the above equation are particularly simple.
The first term relates to the $P_t$ conditional cross section of bond yields. Maximising this term on its own relates to minimising the sum of squared measurement errors. This is made clear in Subsection 2.7.3 in the appendix. The second term relates to the time series of $P_t$ and it is given in (2.30) below. The two terms are connected; market prices of risk (that are restricted to be linear in forward rates) link $K_Q^P$ and $\theta_Q^P$ with $K_P^P$ and $\theta_P^P$ (equations analogous to (2.22) and (2.23) but with parameters subscript by $P$ apply).

I ensure econometric identification by searching over estimates of $\psi_X$ (denoted by $\hat{\psi}_X$) instead of estimates of $\psi_P$ (denoted by $\hat{\psi}_P$) in the ML procedure. Recalling from Subsection 2.4.3 that $\psi_P$ can be calculated via invariant transformations when given $\psi_X$, the approach taken is as follows:

1. Choose an initial $\hat{\psi}_X$ and denote this choice by $\hat{\psi}_X^0$. This choice is based on (among other regressions) an OLS regression on $P_t$. The procedure for choosing $\hat{\psi}_X^0$ is described in detail in the next subsection.

2. Starting from $\hat{\psi}_X^0$, estimate the $\hat{\psi}_X$ corresponding to the maximum likelihood using the Nelder-Mead simplex scheme (the $fminsearch$ function in MATLAB\(^1\)). The likelihood is evaluated by calculating the $\hat{\psi}_P$ that corresponds to the $\hat{\psi}_X$ and substituting this into (2.29). The uniqueness of $\hat{\psi}_P$ for a given $\hat{\psi}_X$ was proven by JSZ (see their Theorem 1).

3. Repeat steps 1 to 2 many times, but each time using different starting values of $\hat{\psi}_X$ that are draws from a multivariate normal distribution with mean given by $\hat{\psi}_X^0$.

4. Take as the ML estimate the $\hat{\psi}_X$ that corresponds to the highest value of the likelihood function. Call this $\hat{\psi}_X^{MLE}$, and apply invariant transforms to calculate the corresponding $\hat{\psi}_P$.

The remaining task is to choose an initial $\hat{\psi}_X^0$ in Step 1 above to begin the ML procedure.

2.4.6 ML Initialisation; Choice of $\hat{\psi}_X^0$

The choice of $\hat{\psi}_X^0$ is important because it determines whether or not the ML procedure converges to the global maximum of the likelihood function. The task of choosing $\hat{\psi}_X^0$ is complicated by the large number of local optima associated with the likelihood function in DTSMs (see, for example, Duffee (2009) for further discussion of this issue). However,

\(^1\)Duffee (2009) advises against the use of derivative based methods in the estimation of DTSMs.
the fact that $P_t$ is observable, the form of the second term in the likelihood function and
the invariant transformation methods for GDTSMs (given in Subsection (2.7.2) in the
appendix) can provide clues as to a sensible choice.

The intuition behind the approach taken to choosing $\hat{\psi}_X^0$ in this subsection is that, since
$P_t$ is observable, OLS regressions can be used to calculate consistent estimates of the
elements in $\psi_P$, with the exception of $K_P^Q$ and $\theta_P^Q$. Next, regressions similar to those
used by CP can be used to estimate market prices of risk and hence, via (2.22) and
(2.23), estimates of $K_P^Q$ and $\theta_P^Q$. I denote the vector of parameters estimated in this
way by $\hat{\psi}_P^0$. We can then set $\hat{\psi}_X^0$ to be an invariant transform of $\hat{\psi}_P^0$ such that $\hat{\psi}_X^0$ takes
canonical form in (2.24). The remainder of this subsection formalises these ideas.

**Initial Estimates of $K_P^P$, $\theta_P^P$ and $\Sigma_P$**

I use the second term in the likelihood function in (2.29)) to form initial estimates of
$K_P^P$ and $\theta_P^P$. Since $P_t$ follows a Gaussian affine process, this term is given by

$$f(P_t|P_{t-\Delta};K_P^P,\theta_P^P,\Sigma_P) = (2\pi)^{\frac{n}{2}}|V_P|^{-\frac{1}{2}} \exp \left( \frac{1}{2} \|V_P^{-\frac{1}{2}}(P_t - E_{t-\Delta}[P_t])\|^2 \right),$$

where, for a vector $x$, $\|x\|^2$ denotes the Euclidean norm squared: $x^T x$. Given $\psi_P$, it is
a standard result that

$$E_{t-\Delta}[P_t] = (I_n - e^{-K_P^P \Delta})\theta_P^P + e^{-K_P^P \Delta} P_{t-\Delta},$$

where $e^{K_P^P \Delta}$ is the fundamental matrix associated with $-K_P^P \Delta$. $V_P$ is the covariance ma-
trix of $P_t$ (its dependence on $\Sigma_P$ is provided in (2.37) in the appendix). The parameters
that maximise (2.30), namely

$$(\hat{K}_P^P, \hat{\theta}_P^P) = \arg\max_{K_P^P, \theta_P^P} \prod_{j=1}^T f(P_{\Delta_j}|P_{\Delta_{j-1}};K_P^P,\theta_P^P,\Sigma_P)$$

$$= \arg\min_{K_P^P, \theta_P^P} \sum_{j=1}^T \|V_P^{-\frac{1}{2}}(P_{\Delta_j} - E_{\Delta_{j-1}}[P_{\Delta_j}])\|^2$$

(2.31)

can be calculated using OLS, independent of $V_P$ (Zellner (1962)). JSZ show that, in an
unrestricted GDTSM, the $\hat{K}_P^P$ and $\hat{\theta}_P^P$ calculated in this way correspond to the global
optimum of the likelihood function. Here, this is not generally the case. The reason is
as follows.
$K_P$ and $\theta_P$ are connected via restrictions on market prices of risk to the terms $K_P^Q$ and $\theta_P^Q$ that appear in the first term in (2.29). Indeed, this must be the case; a main aim in this chapter is to use the cross section of yields (which are related the first term in (2.29)) to improve the efficiency of our estimates of $K_P$ and $\theta_P$ and hence our forecasts of future yields. Nevertheless, given JSZ’s above mentioned result and the fact that OLS provides consistent estimates for $K_P$ and $\theta_P$, I use the OLS estimates $\hat{K}_P$ and $\hat{\theta}_P$ in (2.31) in $\psi_0$.

Next, $\hat{K}_P$ and $\hat{\theta}_P$ can be used to calculate an estimate for $V_P$, denoted by $\hat{V}_P$. I take $\hat{V}_P$ to be the sample conditional variance of $P_t$, calculated as if $\hat{K}_P$ and $\hat{\theta}_P$ were the data generating parameters. Equation (2.37) in the appendix and some simple algebra can then be used to calculate an estimate for $\Sigma_P$, denoted by $\hat{\Sigma}_P$.

### Initial Estimates of $\delta_{0,P}$ and $\delta_P$

The initial estimates of $\delta_{0,P}$ and $\delta_P$ are be calculated based on the relation $r_t = \delta_{0,P} + \delta_P P_t$. I proxy for $r_t$ using the shortest maturity well measured yield (the 3 month yield in this chapter). Letting $y_t^n$ denote an $n$-vector consisting of the first $n$ elements of $y_t$, I make the approximation that $y_{t,1}^n \approx \delta_{0,P} + \delta_P P_t$.

Here, the left hand variable denotes the first element of $y_t^n$. Next, using (2.28) to write the relation $P_t = Q_T^T y_t^n$ and using the above approximation reveals the initial estimates,

$$\hat{\delta}_{0,P} = 0,$$

$$\hat{\delta}_P = Q_T^T [1 0_{n-1}]^T.$$ 

### Initial Estimates of $K_P^Q$ and $\theta_P^Q$

The approach taken to estimate $K_P^Q$ and $\theta_P^Q$ is to run regressions similar to CP to estimate $b$ and $\gamma$ and then use (2.22) and (2.23) and $\hat{K}_P$ and $\hat{\theta}_P$ calculated above to find $\hat{K}_P^Q$ and $\hat{\theta}_P^Q$.

Recalling (2.12) and the normalisation $\frac{1}{n} \sum_{i=1}^n b_i = 1$, I calculate $\hat{\gamma}$ by estimating the following regression by OLS:

$$\sum_{i=1}^n \frac{P_t(u_i - \Delta) - P_t(u_i)}{P_t(u_i)} = \hat{\gamma} f^{\text{discrete}}_t + \bar{\eta}_t.$$ 

\[2\text{In fact, (2.37) can be used to estimate } \Sigma_P \Sigma_P^T. \hat{\Sigma}_P \text{ is the Cholesky decomposition of this estimate.} \]
The first term above proxies for $\sum_{i=1}^{n} \varepsilon_{t,i}$. Here, I choose $u_i = w_i$ for $i = 1, \ldots, n$ (the $u_i$ correspond to the well-measured bond yields in the data set). The vector $f_t^{\text{discrete}}$ proxies for $f_t$. The superscript in $f_t^{\text{discrete}}$ denotes that discrete, rather than instantaneous, forward rates are used. The reason is that the data set consists of bond prices at discrete maturities. I use

$$f_t^{\text{discrete}} = \begin{bmatrix}
\frac{1}{u_2-u_1} \ln \frac{P_t(u_1)}{P_t(u_2)} \\
\vdots \\
\frac{1}{u_n-u_{n-1}} \ln \frac{P_t(u_{n-1})}{P_t(u_n)}
\end{bmatrix}.$$

Next, I calculate $\hat{b}$ by estimating by OLS the following $n$ bond regressions,

$$\frac{P_t(u_i - \Delta) - P_t(u_i)}{P_t(u_i)} = b_i (\hat{\gamma}^T f_t^{\text{discrete}}) + \eta_t$$

for $i = 1, \ldots, n$ (i.e., for the individual bonds $u_1, \ldots, u_n$).

Given $\hat{b}$ and $\hat{\gamma}$, the next task is to calculate $B_P$ ($B_P$ is analogous to $B_X$ in (2.20)). First, using (2.26) we can write

$$\begin{bmatrix}
y_t(u_1) \\
\vdots \\
y_t(u_n)
\end{bmatrix} = -A_P + B_P P_t,$$

(3.32)

Here, $A_P \equiv \begin{pmatrix} A_P(u_1) & \cdots & A_P(u_n) \end{pmatrix}^T$ and $B_P \equiv \begin{pmatrix} B_P(u_1) & \cdots & B_P(u_n) \end{pmatrix}^T$. Further, using $P_t \equiv W y_t$, (2.28) and (3.32) we have that

$$Q_Y^T A_P = 0_n,$$

$$Q_Y^T B_P = I_n.$$

Since $Q_Y$ is orthonormal (and hence invertible), the above equations imply that $A_P = 0_n$ and $B_P = (Q_Y^T)^{-1} = Q_Y$.

The final task is to estimate $F_{0,P}$ and $F_{1,P}$. Given the relation $f_t = F_{0,P} + F_{1,P}$ in (2.17), I proxy for $f_t$ with $f_t^{\text{discrete}}$ and calculate $\hat{F}_{0,P}$ and $\hat{F}_{1,P}$ by running the vector regression

$$f_t^{\text{discrete}} = F_{0,P} + F_{1,P} P_t + \varepsilon_t.$$
We now have estimates of all the quantities that are required to calculate $\hat{K}_P^Q$ and $\hat{\theta}_P^Q$ via Equations (2.22) and (2.23) respectively, and hence also of the parameter vector $\hat{\psi}_P^0$. The final task is to transform these estimates to calculate the ML procedure’s starting vector $\hat{\psi}_X^0$.

**Invariant Transformation from $\hat{\psi}_P^0$ to $\hat{\psi}_X^0$**

Letting $X_t = C + DP_t$, the purpose in this subsection is to find $C$ and $D$ such that $\hat{\psi}_X^0$ is an invariant transformation of $\hat{\psi}_P^0$ and satisfies the normalisations imposed by the $A_0(n)$ canonical form in (2.24).

Let

\[
D = Q_K^T \hat{\Sigma}_P^{-1} \quad \text{and} \quad C = -D\hat{\theta}_P^Q.
\]

Here, $Q_K$ is an orthonormal matrix; the $n(n - 1)/2$ free elements in this matrix are chosen so that

\[
Q_K^T (\hat{\Sigma}_P^{-1} \hat{K}_P^Q \hat{\Sigma}_P) Q_K
\]

is a lower triangular matrix. In other words, the $n(n - 1)/2$ free elements are simply chosen to “zero-out” $n(n - 1)/2$ elements in $\hat{\Sigma}_P^{-1} \hat{K}_P^Q \hat{\Sigma}_P$. The Schur decomposition is used here.

Substituting the above choices of $C$ and $D$ into the invariant transformation equations given in Subsection 2.7.2 in the appendix shows that the forms given above are satisfy the requirements of the $A_0(n)$ canonical form that was described in Subsection 2.4.2.

These choices of $C$ and $D$ are then used to make the invariant transformation from $\hat{\psi}_P^0$ to the ML initialisation point $\hat{\psi}_X^0$.

### 2.4.7 Data

The data set consists of $p = 7$ yields. I take the 3 and 6 month zero coupon bill yields and 12, 24, 36 and 60 month zero coupon bond yields from the Fama-Bliss CRSP Treasury bill and bond files respectively. The full data set includes 679 time series observations of continuously compounded yields for each maturity at monthly frequency covering the period June 1952 to December 2008. I follow Fama and Bliss (1987) and CP and omit all data in the closed period 1952 to 1963. The reason is that this data behaves very differently from the rest of the set in terms of, for example, predictability regressions. This leaves 540 time series observations. I use the observations from 1964 to 2003 ($T =$
Table 2.1: GDTSM ML Parameter Estimates.
The table shows the elements of $\hat{\psi}_{X}^{\text{MLE}}$. The standard errors are shown in the parentheses and are calculated using the second derivative estimate of the information matrix.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Index No.(i)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta_{X,0}$</td>
<td>0.025</td>
</tr>
<tr>
<td>$\delta_{X,i}$</td>
<td>0.26</td>
</tr>
<tr>
<td>$K_{X,Q}^{1,i}$</td>
<td>1.62</td>
</tr>
<tr>
<td>$K_{X,Q}^{2,i}$</td>
<td>0.00</td>
</tr>
<tr>
<td>$K_{X,Q}^{3,i}$</td>
<td>0.00</td>
</tr>
<tr>
<td>$b_i$</td>
<td>0.15</td>
</tr>
<tr>
<td>$\gamma_i$</td>
<td>0.040</td>
</tr>
</tbody>
</table>

480 time series observations) for the purpose of model estimation, and the observations from 2004 to 2008 (60 time series observations) for out-of-sample forecasting.

I estimate an $n = 3$ model. I choose $w_1 = u_1 = 3$ mth, $w_2 = u_2 = 2$ yr and $w_3 = u_3 = 4$ yr to be well measured yields. Repeating the estimation with a different choice of well measured yields ($w_1 = 1$ yr, $w_2 = 3$ yr and $w_3 = 5$ yr) does not yield substantial enough differences in the estimate $\hat{\psi}_{X}^{\text{MLE}}$ to change the conclusions of this chapter.

2.5 Results

2.5.1 Parameter Estimates

The estimate $\hat{\psi}_{X}^{\text{MLE}}$ is given in Table 2.1. However, since the vector $P_t$ is easier to interpret than the latent state vector $X_t$, the invariant transform of $\hat{\psi}_{X}^{\text{MLE}}$, denoted by $\hat{\psi}_{P}^{\text{MLE}}$ is provided in Table 2.2.
Table 2.2: Observationally Equivalent GDTSM with state $P_t$.
The table shows the invariant transform to $\hat{\psi}^\text{MLE}_P$ of $\hat{\psi}^\text{MLE}_X$. Recall that $P_t$ is a vector of principle components of bond yields related to $X_t$ via Equation (2.25).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Index No.(i)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta_{P,0}$</td>
<td>-0.0027</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Index No.(i)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta_{P,i}$</td>
<td>0.60 -1.07 -0.72</td>
</tr>
<tr>
<td>$K^Q_{P,1,i}$</td>
<td>0.045 -2.31 -1.62</td>
</tr>
<tr>
<td>$K^Q_{P,2,i}$</td>
<td>0.027 1.19 3.41</td>
</tr>
<tr>
<td>$K^Q_{P,3,i}$</td>
<td>-0.013 0.28 -0.56</td>
</tr>
<tr>
<td>$K^P_{P,1,i}$</td>
<td>0.64 -0.09 -0.02</td>
</tr>
<tr>
<td>$K^P_{P,2,i}$</td>
<td>-0.0042 1.12 0.34</td>
</tr>
<tr>
<td>$K^P_{P,3,i}$</td>
<td>-0.17 -0.09 0.65</td>
</tr>
<tr>
<td>$(\theta^Q_P)^T_{1,i}$</td>
<td>0.0038 -0.0348 0.017</td>
</tr>
<tr>
<td>$(\theta^P_P)^T_{1,i}$</td>
<td>0.048 -0.0045 0.0066</td>
</tr>
<tr>
<td>$(\Sigma_P\Sigma_P^P)_{1,i}$</td>
<td>0.083 -0.02 -0.01</td>
</tr>
<tr>
<td>$(\Sigma_P\Sigma_P^P)_{2,i}$</td>
<td>-0.02 0.018 -0.00</td>
</tr>
<tr>
<td>$(\Sigma_P\Sigma_P^P)_{3,i}$</td>
<td>-0.0061 -0.00 0.0020</td>
</tr>
</tbody>
</table>
There are several points to note here. First, Table 2.1 shows that the elements in $b$ corresponding to bonds of longer maturities are larger than those corresponding to shorter maturities. This means that risk premia in the returns of long bonds are larger (in an absolute sense, since risk premia are allowed to be negative in this model) than those to bonds of shorter maturities.

A second point to note is that the elements in $\gamma$ form a tent shape. That is, the 1 year forward rate is more important in determining risk premia than the 3 month or 4 year forward rates. The results here are not unexpected. Indeed, they are along similar lines to those found by and discussed in CP.

Next, Table 2.2 shows that, of the three principle components, the first, the level component, is the most volatile. The reason is that $\langle \Sigma_P \Sigma_P^T \rangle_{1,1}$ is substantially larger than the other elements in $\langle \Sigma_P \Sigma_P^T \rangle$ and the rate of mean reversion $K_P^1$ is smaller in magnitude than the reversion rates among the other components. Further, shocks to this principle component carry a large price; its mean level under $P$, $\langle \theta_P^Q \rangle_1$ is substantially larger than its mean level under $Q$, $\langle \theta_Q^P \rangle_1$. Given the result here that the level principle component drives most of the variation in bond yields and also carries that largest premium, it is clear that the most important part of forecasting future bond yields is forecasting this principle component. This is the topic of the subsection below.

### 2.5.2 Forecasts

Table 2.3 provides the MSEs of forecasts of the level, slope and curvature principle components of bond yields implied by the GDTSM, taken at the parameter estimates given in Table 2.1. These forecast errors are compared with those obtained from a VAR(1) on the principle components of yields, estimated by OLS. The reason behind this comparison is as follows.

The finding of JSZ, that an unrestricted GDTSM estimated by MLE forecasts exactly as well out of sample as a VAR(1) on principle components of yields estimated by OLS, was discussed in Section 2.1. Therefore, if one wishes to improve forecasts, one appropriate approach is to improve the efficiency of the estimates of the P dynamics in the model.

In this chapter, I have argued that, for a given data set, one way of further increasing the efficiency of the parameter estimates is to impose restrictions on model that correspond to the empirical evidence; I do this by restricting risk premia in bond returns to be linear in forward rates.

53
Table 2.3: Out-of-Sample MSE of Forecasts of Level, Slope and Curvature Principle Components, January 2004 to December 2008.

The out-of-sample level, slope and curvature principle components are forecast using a VAR(1) and the GDTSM at 1, 3, 6 and 12 month horizons and the MSEs are calculated. There are 60 monthly time series observations in the out-of-sample data. Standard errors shown in parentheses are calculated using Newey-West with 6 lags.

<table>
<thead>
<tr>
<th>Forecast</th>
<th>1 month</th>
<th>3 months</th>
<th>6 months</th>
<th>12 months</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Level</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VAR(1)</td>
<td>0.0249 (0.006)</td>
<td>0.124 (0.02)</td>
<td>0.286 (0.04)</td>
<td>0.800 (0.11)</td>
</tr>
<tr>
<td>Restricted GDTSM</td>
<td>0.0249 (0.003)</td>
<td>0.117 (0.01)</td>
<td>0.236 (0.02)</td>
<td>0.492 (0.06)</td>
</tr>
<tr>
<td><strong>Slope</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VAR(1)</td>
<td>0.0044 (0.0016)</td>
<td>0.0087 (0.0032)</td>
<td>0.0127 (0.0056)</td>
<td>0.0254 (0.010)</td>
</tr>
<tr>
<td>Restricted GDTSM</td>
<td>0.0042 (0.0018)</td>
<td>0.0117 (0.0060)</td>
<td>0.0213 (0.0083)</td>
<td>0.0622 (0.0086)</td>
</tr>
<tr>
<td><strong>Curvature</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VAR(1)</td>
<td>0.0006 (0.0004)</td>
<td>0.0007 (0.0002)</td>
<td>0.0010 (0.0008)</td>
<td>0.0014 (0.0012)</td>
</tr>
<tr>
<td>Restricted GDTSM</td>
<td>0.0006 (0.0003)</td>
<td>0.0026 (0.0006)</td>
<td>0.0040 (0.0015)</td>
<td>0.0089 (0.0014)</td>
</tr>
</tbody>
</table>

The economics behind the improvements in estimation efficiency here are that, since, in the presence of such restrictions, the yield curve can be used to forecast future yields, the cross section of yields and the time series are now jointly used in estimation. That is, the Q and P dynamics are estimated jointly. From a purely econometric perspective, the number of parameters to estimate has been reduced and, furthermore, the parameters that have been restricted are difficult to estimate in the sense that the likelihood function is not very sensitive to them. The improvements in estimation efficiency result in better forecasts than a VAR(1) estimated by OLS in expectation.

Table 2.3 shows that the forecasts of the level, slope and curvature principle components of bond yields result in smaller MSEs than a VAR(1) estimated by OLS at the one month horizon. For the level component, this result holds at the four horizons reported. Further, since, as discussed above, the level component is the most important principle component accounting for most of the variation (96%) in bond yields, this result is particularly important. The improvements in forecasting lend support to the idea that the parameters may have been estimated more efficiently than by an OLS on a VAR(1).
The exceptions to the above mentioned forecasting improvements are that the slope and curvature principle components are forecast less well than the VAR(1) at longer horizons (3, 6 and 12 months). The difference, however, is small.

### 2.6 Concluding Comments

This chapter has developed an arbitrage free, analytically and econometrically tractable GDTSM that is able to capture the empirical relation found by CP that a linear combination of forward rates forecasts excess returns to Treasury bonds over bills. Implementing the model, it is found that its forecasting is better than a VAR(1) for the level principle component of bond yields. This is the principle component with the largest variance, accounting for approximately 96.5% of the variation in yields. Following the insights of JSZ relating to GDTSMs, this result suggest that the restrictions imposed on prices of risk here may have improved the efficiency of the estimates of the model parameters and these results therefore provide support for the empirical findings of CP.

Finally, building on the invariant transformation methods developed by Dai and Singleton (2000), a new technique is used to aid the econometrician in choosing an initialisation point for the ML estimation of the model. Given the large number of local maxima associated with the likelihood functions of DTSMs, the methods discussed here help in this otherwise difficult task. Although JSZ provided a canonical form that greatly simplifies the estimation of unrestricted GDTSMs, their results do not carry forward to restricted GDTSMs. The methods used here are a contribution towards simplifying the estimation problem when prices of risk in a GDTSM are restricted.
2.7 Appendix

2.7.1 Gaussian Affine Bond Pricing

An $n$ factor Gaussian DTSM with state vector $X_t$ has a parameter vector

$$\psi_X = (\delta_{0,X}, \delta_X, K_Q^X, \theta_Q^X, \Sigma_X, K_P^X, \theta_P^X).$$

Here $\delta_{0,X}$ is a scalar, $\delta_X$, $\theta_Q^X$ and $\theta_P^X$ are $n$-vectors, and $K_Q^X$, $K_P^X$ and $\Sigma_X$ are $n \times n$ matrices. The no-arbitrage price at time $t$ of a zero-coupon bond that matures at time $T$ is

$$P_t(u) = E^Q_t [e^{-\int_t^T r_u du}] = P_X(X_t, u) \equiv e^{A_X(u)-B_X(u)^T X_t}.$$

Here, $u \equiv T - t$ and $A_X(u)$ and $B_X(u)$ solve the ODEs

$$\begin{align*}
\frac{\partial B_X(s)}{\partial s} &= -(K_Q^X)^T B_X(s) + \delta_X, \\
\frac{\partial A_X(s)}{\partial s} &= \sum_{i=1}^n \frac{1}{2} [\Sigma_X^T B_X(s)]_i^2 - B_X(s)^T K_Q^X \theta_Q^X - \delta_{0,X},
\end{align*}$$

subject to the initial conditions that $A_X(0) = 0$ and $B_X(0) = 0$. The notation $[\Sigma_X^T B_X(s)]_i$ denotes the $i$-th element of the vector $\Sigma_X^T B_X(s)$. Clearly, this price does not depend on either $K_P^X$ or $\theta_P^X$.

2.7.2 Invariant Transformations

This subsection restates the results of Dai and Singleton (2000). Consider an $n$ factor Gaussian DTSM with state vector $X_t$ and parameter vector

$$\psi_X = (\delta_{0,X}, \delta_X, K_Q^X, \theta_Q^X, \Sigma_X, K_P^X, \theta_P^X).$$

If $Y_t = C + DX_t$ then we may apply invariant transformations to yield an observationally equivalent Gaussian DTSM with state $Y_t$ and parameters

$$\psi_Y = (\delta_{0,Y}, \delta_Y, K_Q^Y, \theta_Q^Y, \Sigma_Y, K_P^Y, \theta_P^Y).$$
Here, $D$ is an $n \times n$ nonsingular matrix, $C$ is an $n$-vector and

\[
\begin{align*}
\delta_{0,Y} &= \delta_{0,X} - D^T X^{-1} C, \\
\delta_Y &= (D^T)^{-1} \delta_X, \\
K_Y^Q &= DK_X Q D^{-1}, \\
\theta_Y^Q &= C + D \theta_X^Q, \\
\Sigma_Y &= D \Sigma_X, \\
K_P^Y &= DK_X Q D^{-1} \\
\theta_P^Y &= C + D \theta_X^Q.
\end{align*}
\]

### 2.7.3 Conditional Likelihood Function

The first term in the conditional likelihood function in Equation (2.29) is given by

\[
f(y_t^c|P_t; K_P^Q, \theta_P^Q, \Sigma_P, \delta_{0,P}, \delta_P, \sigma^2) = (2\pi\sigma^2)^{-\frac{p-n}{2}} \exp\left(\frac{\|\epsilon_t\|^2}{2\sigma^2}\right),
\]

where $\epsilon_t$ is the $(p - n)$-vector given in equation (2.27), and for a vector $x$, $\|x\|^2$ denotes the Euclidean norm squared: $x^T x$. The dependence of $\epsilon_t$ on the parameters and $P_t$ can be made more explicit: for $i = 1, \ldots, p - n$, the $i$-th element of $\epsilon$ is given by

\[
\epsilon_{t,i} = y_{t,i+n} - y_{t,i+n} = y_{t,i+n} - \frac{1}{w_{t,i+n}} (A_P(w_{i+n}) - B_P(w_{i+n})^T X).
\]

The second equality and the dependence of $A_P(w_{i+n})$ and $B_P(w_{i+n})$ on the parameters follow from equation (2.26).

The quantity $V_P$ in Equation (2.30) is the covariance matrix of $P_t$ and it is given by

\[
V_P = Q_P \Sigma_P Q_P^T.
\]

Here, $Q_P$ is the matrix of eigenvectors of $K_P^Q$,

\[
K_P^Q = Q_P D_P Q_P^{-1},
\]

and the $(i,j)$-th element of $\Sigma_P$ is

\[
\Sigma_{P,ij} = \frac{1 - e^{-(D_{P,ii} + D_{P,jj})\Delta}}{D_{P,ij} + D_{P,jj}} (Q_P^{-1} \Sigma_P (Q_P^{-1} \Sigma_P)^T)_{ij}
\]

57
The above result is a specialisation to the Gaussian case of those in the appendix of Duffee (2002).

2.7.4 Portfolio Matrix $W$

The portfolio matrix $W$ defined in (2.28) is given in Table 2.4.

Table 2.4: The portfolio matrix $W$ is calculated using the eigenvectors of $V_Y$

<table>
<thead>
<tr>
<th></th>
<th>Portfolio Matrix $W$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level</td>
<td>0.60 0.59 0.54 0 0 0 0</td>
</tr>
<tr>
<td>Slope</td>
<td>-0.76 0.20 0.62 0 0 0 0</td>
</tr>
<tr>
<td>Curvature</td>
<td>-0.26 0.79 -0.56 0 0 0 0</td>
</tr>
</tbody>
</table>
Chapter 3

Risk Premia in Affine Corporate Bond Pricing Models

Chapter Summary

Important recent research into risk premia in corporate bond returns uses intensity based models of default. Perhaps to maintain tractability, a common approach has been to assume that the prices of risk associated with shocks to default intensities/credit spreads are affine functions of the state vector. Here, I show that such specifications implicitly force a strong link between the risk premium component of corporate bond spreads and their contemporaneous volatility. However, it is not clear that such a relation should be consistent with empirical fact. Using data on corporate bond indices, I find evidence that such specifications, though restrictive, may indeed be able to capture much of the time variation in risk premia in corporate bond spreads.
3.1 Introduction

How should we construct a defaultable dynamic term structure model (DDTSM) - a model that describes the co-movement of corporate bond yields over time? No-arbitrage reduced form default intensity models, originally introduced by Artzner and Delbaen (1990, 1995), Jarrow and Turnbull (1995) and Lando (1998), have now become standard fare in defaultable bond pricing. These are models in which corporate default is said to occur at the first jump time of a conditional Poisson process. Perhaps due to their tractability and flexibility, so-called affine formulations of intensity models have become particularly popular. These are models in which corporate bond yields are affine functions of the state vector and the state vector follows an affine diffusion under the risk adjusted probability measure (Q) (Duffie (2005) provides an overview of such models). However, many such models do not reference drifts and market prices of risk; they tell us about the relative prices of corporate bonds in the cross section, but not about their expected returns in the time series.

Models that do not specify market prices of risk can be used in derivative pricing, where modelling volatilities is the main concern. Another application is in pricing corporate bonds of untraded maturities relative to those that are traded. This may be useful in valuing new bond issuances relative to bonds that are currently outstanding. However, if our main interest is bond portfolio selection, credit spread forecasting or decomposing the defaultable yield curve into expectations and risk premium components to study, for instance, firms’ costs of debt, then it is necessary that prices of risk are specified and that they are specified such that the implied defaultable yield curve dynamics are consistent with the empirical facts.

One well known approach to modelling corporate bond returns has been to construct an affine DDTSM and choose prices of risk so that the variance of the SDF is affine in the state vector (see, for example, the seminal contribution of Duffee (1999)). Such an approach offers tractability; this is one way in which the state vector can be made to follow an affine diffusion under both the objective probability measure, P, and under Q. These DDTSMs are relatively straight forward to estimate (Singleton (2006) provides a textbook discussion on estimating affine models). However, this specification imposes restrictions that force a strong link between the conditional risk premia in, and the contemporaneous volatility of, corporate bond returns.

The tension between risk premia and volatility discussed here was first noted by Duffee (2002) and Dai and Singleton (2002) in the context of affine dynamic term structure
models (DTSMs) for pricing Treasury bonds. Such Treasury bond pricing models are commonly called completely affine DTSMs. Given the close relationship between affine default free and default intensity based term structure modelling, I show in this chapter that this tension carries over to risk premia in corporate bond returns. A main aim then is to examine whether this relation between risk premia and volatility holds in corporate bond data.

I regress monthly excess returns to corporate bond indices on an estimate of the previous month’s variance of bond returns and other factors that are known to proxy for risk premia in the returns of other assets. I find that bond return variance does not drive these other factors out of the regression and so, indeed, such specifications for DDTSMSs are unable to capture the time variation in risk premia in corporate bond returns.

The above mentioned result is, perhaps, unsurprising. The reason is that DDTSMSs in which the variance of the SDF is affine in the state vector are just completely affine DTSMs, plus default risk. Given the well documented empirical failures of completely affine DTSMs for modelling fluctuations in the Treasury yield curve over time (see Duffee (2002)), these failures carry over to DDTSMSs. For this reason, in the next part of this chapter, I use a simple method to separate the components of risk premia in corporate bond returns due to shocks in the default free interest rate and shocks in default intensities. I then study whether allowing only the variance of only that part of the SDF that correlates with default intensities to be affine in the state vector is able to capture the time variation in risk premia associated with shocks to default intensities. This specification maintains tractability; it is similar to that used by Driessen (2005) to estimate risk premia in corporate bond returns. Extending the nomenclature of Duffee (2002), I refer to such models as completely affine DDTSMSs. Here, the results provide support for such specifications.

Simply put, the findings in this chapter suggest that completely affine DDTSMSs may be suitable for modelling conditional risk premia in corporate bonds returns. That is, the risk premia associated with shocks to default intensities/credit spreads are closely related to the volatility of default intensities. Whilst completely affine DTSMs may not adequately capture the time variation in risk premia to Treasury bonds, their analogous counterparts for pricing shocks in default intensities may indeed capture the time variation in default related risk premia. The results here also suggest that any completely affine DDTSM should allow flexible specifications for the default free interest rates; the model provided in Chapter 2 is one suggestion.
The results from the analysis here also suggests another interesting result; the risk premium component of credit spreads decreases with increasing volatility of credit spreads. If the volatility of credit spreads is high when credit spreads are high, as one may expect, then the risk premium fraction of the credit spread in investment grade bonds is greater than that in high yield bonds. In other words, the default rates of investment grade debt may be substantially lower than their credit spreads imply under risk neutrality. However, this effect is less pronounced for high yield debt.

The remainder of this chapter proceeds as follows. Section 3.2 recapitulates the key equations relating to affine credit risk models and discusses important empirical literature on credit risk premia with which these models must be consistent. Section 3.3 discusses the restrictions that affine DDTSMs place on risk premia in bond returns and proceeds to test these restrictions against corporate bond index data. Section 3.4 concludes.

3.2 Affine Reduced Form Pricing Models

3.2.1 Analytical Bond Pricing

This subsection derives the bond pricing function for affine DDTSMs under the recovery of market value assumption of Duffie and Singleton (1999). The results in this section hold for all affine DDTSMs, regardless of whether or not they are completely affine. The meaning of completely affine is made formal later (see Section 3.3).

At the default time, the value of the defaultable bond falls to a constant fraction, \((1 - L)\), of its value just before default. In this framework, in the absence of arbitrage, the time \(t\) conditional price of a time \(T\) maturity defaultable zero-coupon bond that has not defaulted by time \(t\) is

\[
\tilde{P}_t(u) = \mathbb{E}_t^Q \left[ \exp \left( - \int_t^T (r_v + s_v^Q) dv \right) \right].
\]  

(3.1)

Here, \(u \equiv T - t\), \(r_v\) is the short interest rate and \(s_v^Q\) is the short (instantaneous) credit spread level under the Q measure. \(s_v^Q \equiv \lambda_v^Q L\), where \(\lambda_v^Q\) is the default intensity under Q. The notation \(\tilde{P}_t(u)\) distinguishes the corporate bond price from that of the Treasury bond, which I denote by \(P_t(u)\).

The above equation shows that it suffices to model \(s_v^Q\) to price corporate bonds; the assumption of a constant \(L\) is innocuous.
Risk is generated in the economy by \( n \) \( \mathbb{Q} \)-Brownian motions \( W^Q_t \equiv (W^Q_{t,1}, \ldots, W^Q_{t,n})^T \). The \( n \)-dimensional state vector is denoted \( X_t \equiv (X_{t,1}, \ldots, X_{t,n})^T \) and is driven by \( W^Q_t \). The short rate and \( \mathbb{Q} \) short credit spread are specified by

\[
\begin{align*}
    r_t &= \delta_0 + \delta^T X_t, \\
    s^Q_t &= \gamma_0 + \gamma^T X_t,
\end{align*}
\]

where \( \delta_0 \) and \( \gamma_0 \) are scalars and \( \delta \) and \( \gamma \) are \( n \)-vectors. The \( \mathbb{Q} \) evolution of the state vector is assumed to be

\[
dX_t = K^Q \left[ \theta^Q - X_t \right] dt + \Sigma S_t dW^Q_t,
\]

where \( K^Q \) and \( \Sigma \) are \( n \times n \) matrices and \( \theta^Q \) is an \( n \)-vector. The matrix \( S_t \) is diagonal with elements

\[
S_{t(ii)} = \sqrt{\alpha_i + \beta_i^T X_t}.
\]

for \( i = 1, \ldots, n \). Here, the \( \beta_i \) are \( n \)-vectors and the \( \alpha_i \) are scalars. I assume that the dynamics in (3.4) are defined so that \( \alpha_i + \beta_i^T X_t \) is nonnegative for all \( i \) and all possible \( X_t \). We will see below that the matrix \( S_t \) that governs the volatility of \( X \) (and hence also of bond returns) also enters the conditional risk premium in the DDTSMs studied here.

The approach of Duffie and Kan (1996) to price Treasury bonds can be applied to calculate the price of a defaultable bond in this framework. Conditional on no default by time \( t \), the price is

\[
\bar{P}_t(u) = \bar{P}(X_t, u) \equiv e^{A(u) - B(u)^T X_t}
\]

where \( u \equiv T - t \), \( A(u) \) is a scalar function and \( B(u) \) is an \( n \)-valued function. These functions are calculated by solving a system of Riccati ODEs (see Subsection 3.5.1 in the appendix). Finally, given (3.6) the defaultable bond yield is

\[
\bar{y}_t(u) \equiv -\frac{1}{u} \ln \bar{P}_t(u) = \frac{1}{u} (-A(u) + B(u)^T X_t).
\]

1 Sufficient conditions for existence of a solution to (3.4) are

1. For all \( X \) such that \( S_{ii} = 0 \), \( \beta_i^T K^Q (\theta^Q - X) > \frac{1}{2} \beta_i^T \Sigma \Sigma^T \beta_i \).

2. For all \( j \), if \((\beta_i^T \Sigma)_{ij} \neq 0 \), then \( S_{ii} \) and \( S_{jj} \) are proportional.

This multidimensional extension of the Feller condition for correlated affine diffusions was provided by Duffie and Kan (1996).
3.2.2 Risk Premia

Given the bond pricing function in (3.6), I now turn to the problem of specifying the SDF that will determine the risk premia associated with holding corporate bonds. I begin by specifying the SDF up to the market prices of risk only and then analyse the form of the risk premium.

The SDF, denoted by $\pi_t$, follows the process

$$\frac{d\pi_t}{\pi_t} = -r_t dt - \Lambda_t^T dW_t^P - \Gamma(dZ_t - \lambda_t^P dt). \quad (3.8)$$

Here, $\Lambda_t$ is an $n$-vector specifying market prices of risk and $Z_t$ is a counting (jump) process with associated state-dependent intensity $\lambda_t^P$. $\Gamma$ governs the size of the default event risk premium, assumed here to be constant. The role of $\Gamma$ will become clear below.

A standard change of probability measure reveals that

$$dX_t = K_Q^t[\theta^Q - X_t] dt + \Sigma S_t \Lambda_t dt + \Sigma S_t dW_t^P,$$

and

$$\lambda_t^Q = (1 - \Gamma)\lambda_t^P. \quad (3.10)$$

We can now see that $\Gamma$ is the price of default event risk. If $\Gamma$ is different to 0 then the default intensity level that is used in pricing, $\lambda_t^Q$, is different to the objective default intensity, $\lambda_t^P$. This means that the investor can earn ($\Gamma < 0$) or pay ($\Gamma > 0$) a default event risk premium. This is in addition to the premium associated with priced changes in default intensities that are determined by $\Lambda_t$. This point was noted by Jarrow, Lando, and Yu (2005), who also proved that default event risk cannot be priced ($\Gamma = 0$) if default jumps are conditionally independent across an infinite number of corporates. Driessen (2005) provides a test of the hypothesis that $\Gamma = 0$.

Conditional on no default by time $t$, the pricing function of the bond solves

$$d\bar{P}(X_t, u) = \mu_t(u) dt + v_t(u)^T dW_t^P - L dZ_t,$$

where $\mu_t(u)$ denotes the instantaneous expected excess return conditional on no default occurring in the next instant of time and $v_t(u)$ is an $n$-dimensional volatility vector. An application of Itô’s lemma to the function in (3.6) and using (3.9) reveals the form of
\( \mu_t(u) \) and \( v_t(u) \). We have,

\[
\mu_t(u) \equiv \frac{1}{dt} \mathbb{E}^P \left[ \frac{d\overline{P}(X_t, u)}{P(X_t, u)} \right]_{\text{no default}} = (r_t + s_t^Q - B(u)^T S_t \Lambda_t) \tag{3.12}
\]

and

\[
v_t(u)^T = -B(u)^T S_t. \tag{3.13}
\]

We may now calculate the instantaneous expected return, denoted by \( e_t(u) \). Yu (2002) shows that the unconditional expected return in this framework is

\[
\mathbb{E}^P \left[ \frac{d\overline{P}(X_t, u)}{P(X_t, u)} \right] = (1 - \lambda_t^d dt) \mu_t(u) dt + \lambda_t^d dt (-L). \tag{3.14}
\]

This equation and (3.12) together imply that the conditional risk premium, denoted by \( e_t(u) \) is given by

\[
e_t(u) \equiv \frac{1}{dt} \mathbb{E}^P \left[ \frac{d\overline{P}(X_t, u)}{P(X_t, u)} \right] - r_t = s_t^Q - B(u)^T S_t \Lambda_t - L \lambda_t^d = -B(u)^T S_t \Lambda_t - \Gamma s_t^P, \tag{3.15}
\]

where, in the last line, I use (3.10) and \( s_t^P \equiv L \lambda_t^d \).

This equation says that risk premia in affine DDTSMs are, in general, time varying; they vary with the volatility matrix \( S_t \), the price of risk vector \( \Lambda_t \) and with \( s_t^P \) although, of course, the value of \( s_t^P \) is closely related to \( S_t \) because they both depend on the state \( X_t \). The model is not yet complete. We must specify the price of risk vector \( \Lambda_t \).

At this stage \( \Lambda_t \) is the only remaining flexibility with which to ensure that the DDTSM is consistent with the empirical facts about risk premiums in corporate bond returns. I will review some of these facts in Subsection 3.2.4 before specifying, and testing against the data, forms for \( \Lambda_t \) that have been used in past studies. Before this, I discuss some results that will be useful when discussing the determinants of risk premia and credit spreads.

**3.2.3 Corporate Bond Credit Spreads**

I begin by splitting corporate bond yields into Treasury yields and credit spreads. This allows us to examine the determinants of the credit spread component of corporate bond yields separately to the Treasury yield component. I focus on the credit spread
Let $P_t(u)$ denote the time $t$ price of a zero-coupon Treasury bond that matures in a time period of length $u$. Then the Treasury yield for this bond, denoted by $y_t(u)$, is defined in

$$P_t(u) \equiv e^{-y_t(u)u}.$$ (3.17)

The credit spread at time $t$ for a corporate zero-coupon bond maturing in time $u$, denoted by $s_t(u)$, is then defined in

$$P_t(u) \equiv P_t(u)e^{-s_t(u)u}.$$ (3.18)

An important question is, what factors determine credit spreads? To address this question I write the no-arbitrage result that the price of a zero-coupon Treasury bond is given by

$$P_t(u) = \mathbb{E}_t^P \left[ \frac{\pi_{t+u}}{\pi_t} \right],$$ (3.19)

and the price of a zero-coupon defaultable bond with zero recovery is given by

$$\bar{P}_t(u) = 1_{\tau>t} \mathbb{E}_t^P \left[ \frac{\pi_{t+u}}{\pi_t} 1_{\tau>t+u} \right] = 1_{\tau>t} \mathbb{P}_t(\tau > t + u) \mathbb{E}_t^P \left[ \frac{\pi_{t+u}}{\pi_t} \bigg| \tau > t + u \right].$$ (3.19)

Here, $\tau$ denotes the random default time of the bond and $1_{\tau>t}$ indicates that the bond has not defaulted by time $t + u$. The subscript in $\mathbb{P}_t$ denotes that the probability is time $t$ conditional. The second equality follows from the Law of Iterated Expectations.

There are two main points to note here. First, (3.19) tells us that spreads can change due to a change in default probabilities. This is the risk neutral component of credit spread changes. I mean this in the sense that it would exist even in an economy in which agents are risk neutral.

Second, spreads can change due to a change in the non-default conditional expected growth in the SDF. If the expected growth of the SDF is held fixed (so that, by (3.18), Treasury bond prices are fixed) then any changes in this factor will be reflected in spreads but not in Treasury yields. This is the risk premium component of credit spread changes.

I conclude from this simple analysis that the factors that are used in explaining the levels of credit spreads should proxy for two sources of changes in spread levels: changes in default probabilities and changes in risk premia. It will be shown below that completely affine DDTSMs assume that the risk premium part of the credit spread is captured in...
the volatility of short credit spreads.

Finally, as an aside, the factors used in explaining credit spread levels should also proxy for the sources of variation in credit spreads that I have not discussed here. Some examples are bond illiquidity, loss given default and taxes. Increases in these variables will have a positive impact on spread levels and the risks to changes in these factors may be priced.

3.2.4 Related Empirical Work on Defaultable Bond Pricing

I show in Section 3.3 that completely affine DDTSMs make the strong statement that the level of volatility of short credit spreads effects risk premia. Here, I discuss some of the recent empirical work in corporate bond pricing that relates to this statement and to the results that were discussed in the subsection above.

Corporate Bond Spread Levels

Campbell and Taksler (2003) conducted an econometric analysis that showed that realised equity volatility is a significant positive factor in explaining credit spread levels. Cremers, Driessen, Maenhout, and Weinbaum (2008) showed a similar result using stock option implied volatilities. Although intensity based corporate bond pricing models do not, in general, explicitly make a statement about the relationship between equity and bond volatility, one may expect that the two are closely positively related. If this is true, then the results in these papers can be implied by affine DDTSMs whether or not they are completely affine. However, if the data generating model were a completely affine DDTSM, then these results may be strengthened or weakened, depending on the sign of the relation between risk premia and volatilities. These points can be understood as follows.

First, consider a general affine DDTSM (that is not necessarily completely affine). This is any DDTSM in which the state vector $X_t$ follows an affine diffusion under the Q probability measure and corporate bond yields are affine functions of $X_t$. Regardless of the specification of the risk premium, (3.13) and (3.15) together show that the bond variance $v_t^T(u)v_t(u)$ is linear in $X_t$. It is straightforward to construct a model in this framework in which spreads increase with volatility. For example, in the case of scalar $X_t$, choosing prices of risk so as to allow the P default intensity (and hence short spread

---

2Capital structure based corporate bond pricing models such as Merton (1974) and it’s extensions (see Subsection 1.2.1 for references) provide an economic rationale for a relationship between equity and corporate bond volatility.
\( s_t^p = \lambda^p L \) to follow a square root process \( \lambda_t = X_t \), \( dX_t = \kappa(\theta - X_t)dt + \sigma \sqrt{X_t}dW_t^p \) would yield this result.

Second, if the risk premium is specified so that the DDTSM is completely affine, then this may strengthen or weaken the dependence of spreads on volatilities. This is because, in such models, high volatilities lead not only to high short spreads, but also to a large risk premium component to overall spreads. If this risk premium is positive (negative), this will lead to an increase (decrease) in spreads. This will be made clear in Section 3.3 below. The results in the above mentioned papers do not necessarily provide conclusive evidence against completely affine DDTSMs.

Other related empirical literature on the determinants of credit spreads is Collin-Dufresne, Goldstein, and Spencer-Martin (2001) who find that a single unobserved factor common to all corporate bonds drives most of the variation in credit spreads across firms. Since these authors already control for market wide volatility, their empirical results have implications for affine DDTSM specifications, regardless of whether or not they are completely affine. This can be understood as follows.

Since the unobserved factor drives the variation in spreads across all firms, by (3.3) this factor must enter \( X_t \) and its coefficient in \( \gamma \) must be non-zero for every firm. If this factor also enters \( S_t \) (its \( \beta_i \) coefficient is non zero for at least one \( i \)), then this factor will affect market wide bond volatility. However, since the authors have controlled for market wide bond volatility, this cannot be the case. In other words, to reconcile the results in Collin-Dufresne, Goldstein, and Spencer-Martin (2001) with affine DDTSMs it must be imposed that there is at least one element in \( X_t \) that is Gaussian (it does not enter \( S_t \)) and so does not affect market wide volatility, but does affect market wide spread levels.

**Corporate Bond Returns**

Elton, Gruber, Agrawal, and Mann (2001) is a well known contribution to the study of corporate bond returns. These authors carry out tax-adjusted Fama and French (1993) style regressions and find that the Fama-French factors are priced in corporate bonds. It is important to note that these authors answer a different question to that asked in this chapter. Here, I study the time variation in expected excess returns to corporate bonds, and ask if this time variation is related to the volatility of short credit spreads, as completely affine DDTSMs imply. I do not study bond pricing factors. In effect, if it is true that corporate bonds returns correlate with the Fama-French factors, then I
ask if the time variation in the price of the Fama-French factors is proxied for by the volatility of default intensities.

More generally, there is a large literature relating to the predictability of excess returns to risky assets. For example, dividend yields were used to forecast excess returns in stock indices by Cochrane (1992) and various term structure related variables (such as forward rates and yield spreads) have been used to forecast excess returns to Treasury bonds over bills by Fama and Bliss (1987), Campbell and Shiller (1991) and Cochrane and Piazzesi (2005). One way in which this predictability can be understood is that the forecasting variables proxy for risk aversion. When risk aversion is high (such as preceding a recession), greater risk premiums may be required to induce investors into holding risky assets. Indeed, the same variables that forecast excess returns to assets can also forecast recessions (see, for example, Fama and French (1989)). It is shown in the next section that completely affine DDTSMs impose that the time variation in risk premia to corporate bonds is included in the volatility of default intensities.

3.3 Completely Affine Risk Premia for Defaultable Bonds

3.3.1 Risk Premia in Individual Corporate Bond Returns

Perhaps the simplest specification of risk premia to ensure that the state vector follows an affine diffusion under P is if the price of risk vector is given by

$$\Lambda_t = S_t \xi,$$  \hspace{1cm} (3.20)

where $\xi$ is a constant $n$-vector. The model in Duffee (1999) is a special case of (3.20).

It is clear that the variance of the SDF, $\Lambda_t^T \Lambda_t = \xi^T S_t^2 \xi$, is affine in $X_t$ (recall that $S_t^2$ is affine in $X_t$). Here, the prices of both risks associated with shocks to interest rates and short credit spreads are affine in the state vector. For this reason, this choice of $\Lambda_t$ can be thought of as a completely affine DDTSM for default intensities plus a completely affine DTSM for Treasury yields. The remainder of this subsection examines this choice of $\Lambda_t$. Subsection 3.3.5 below discusses a specification in which the volatility of only the part of the SDF that correlated with shocks to credit spread is affine in $X_t$.

Substituting (3.20) into (3.16) reveals that the expected excess return is given by

$$e_t(u) = -B(u)^T \Sigma S_t^2 \xi - \Gamma s_t^P = e_0(u) + e_1(u)^T X_t.$$  \hspace{1cm} (3.21)

Here, $e_0(u)$ is a scalar and $e_1(u)$ is an $n$-vector.
We see from the first equality in the above equation that variations in the expected excess return to the corporate bond are driven by the variation in $S_t^2$ and $s_t^P$. However, $s_t^P$ is an affine function of $X_t$ and so is each element in $S_t^2$. This fact allows us to write the second equality.

Next, we can use (3.13) to write the variance of bond returns as

$$v_t(u)^T v_t(u) = B(u)^T \Sigma S_t^2 \Sigma^T B(u) = v_0(u) + v_1(u)^T X_t,$$

(3.22)

where $v_0(u)$ is a scalar and $v_1(u)$ is an $n$-vector. The relation between risk premia in (3.21) and return volatility (or, rather, return variance) in (3.22) is now clear; they are both affine functions of the state vector. There are three further important points relating to these equations to note.

First, consider the case where $S_t$ is specified so that the state vector is positive. Perhaps the simplest example of such a specification is where the elements of $X_t$ each follow independent square root diffusions (for $i = 1, \ldots, n$, $\alpha_i = 0$ and the $i$-th element of the vector $\beta_i = 1$). Choices of $S_t$ that yield positive $X_t$ are useful because they can be used to construct models in which $r_t$ and $s_t^Q$ and hence also $s_t^P$ are positive (ensuring that, in addition, the parameters $\delta_0, \delta, \gamma_0, \gamma$ are all positive is a sufficient condition). In this case, apart from for some special choices of parameters, all the elements of $e_1(u)$ and $v_1(u)$ are non-zero. This means that the same state variables that drive the variance of bond returns also drive risk premia.

Second, suppose that $X_t$ is not required to be positive. Then, if the conditional diversification hypothesis of Jarrow, Lando, and Yu (2005) holds so that default event risk is not priced ($\Gamma = 0$), the only elements of $e_1(u)$ that can be non-zero are those that are coefficients of elements of $X_t$ that enter into $S_t$. That is, the $j$-th element of $e_1(u)$ can only be non-zero if, for some $i$, the $j$-th element of $\beta_i$ is non-zero. The result is that variations in expected excess returns to bonds can only driven by the factors that affect the volatility of $X_t$ and hence also the variance of corporate bond returns.

Third, if $X_t$ is not required to be positive and $\Gamma \neq 0$ then, in general, apart from some special choices of parameters, there are elements in $X_t$ that drive risk premia but not the return variance. Unlike above, the condition $e_1(u)_j = 0$ does not imply that $v_1(u)_j = 0$. Nevertheless, given the discussion, it certainly seems reasonable to suspect a correlation between $e_1(u)$ and $v_1(u)^T v_1(u)$ if the model described here generated the data.
The above results can be extended to a portfolio context.

### 3.3.2 Risk Premia in Corporate Bond Portfolio Returns

Let $N$ denote the number of corporate bonds in a portfolio. Denote the time $t$ pricing function of the $i$-th bond with time $u_i$ to maturity by $\bar{P}_i(X_t, u_i)$. These functions solve stochastic differential equations of the form (3.11). I denote the instantaneous expected excess return and volatility vectors of bond $i$ by $e_t(u_i)_i$ and $v_t(u_i)_i$ respectively; these quantities take the forms given in (3.21) and (3.22). Next, letting $w_{i,t}$ be the time $t$ portfolio weight on the $i$-th bond with the property that $\sum_{i=1}^{N} w_{i,t} = 1$, the time $t$ value of the bond portfolio, denoted by $\bar{V}_t$, and return are given by

$$\bar{V}_t = \bar{V}(X_t, t) \equiv \sum_{i}^{N} w_{i,t} \bar{P}_i(X_t, u_i)$$

(3.23)

and

$$\frac{d\bar{V}(X_t, t)}{\bar{V}(X_t, t)} = \sum_{i}^{N} w_{i,t} \frac{d\bar{P}_i(X_t, u_i)}{\bar{P}_i(X_t, u_i)}$$

(3.24)

respectively. Taking conditional expectations in (3.24) and using (3.15), the instantaneous expected excess return to the portfolio is given by

$$e_t \equiv \frac{1}{dE_t^P} \left[ \frac{d\bar{V}(X_t, t)}{\bar{V}(X_t, t)} \right] - r_t = \sum_{i=1}^{N} w_{i,t} e_t(u_i)_i.$$

Next, applying Itô’s lemma to (3.23) reveals that the volatility vector (the coefficient of the $dW^P_t$ term) in the stochastic differential equation of the portfolio value function $\bar{V}(X_t, t)$ is

$$v_t \equiv \sum_{i=1}^{N} w_{i,t} v_t(u_i)_i.$$

The key point here is that both the instantaneous expected excess return to the portfolio $e_t$ and the variance of the portfolio return, $v_t^T v_t$, remain affine functions of vector $X_t$ and so the close relationship between excess returns and bond volatility described in Subsection 3.3.1 above carries through to bond portfolios.

In the case of an equally weighted portfolio ($w_{i,t} = 1/N$), the coefficients of $X_t$ in the equations for $e_t$ and $v_t^T v_t$ are functions of the $u_i$ only, and not of time. If the portfolio weights are time-varying then the relation between $e_t$ and $v_t^T v_t$ is more complicated.
Further, although the above result is derived for zero-coupon defaultable bonds, since a coupon-bearing bond can be considered a portfolio of zero-coupon bonds with different maturity dates, the result holds for both coupon bearing bonds and portfolios of coupon bearing bonds.

The next task is to study the relation between risk premiums in returns and volatility in the data.

3.3.3 Data

I take daily data on the Dow Jones Corporate Bond Total Return Indices distributed by Reuters. There are four indices and five maturity classifications in each index, leading to twenty price time series. The Composite All Maturity Index consists of 96 bonds at all times, 32 in each of the industry sectors (financial, industrial and utilities/telecom) and 8 bonds in each maturity cell (2, 5, 10 and 30 years). Maturity 2 year includes bonds with maturities between 1.50 and 3.49 years, 5 year includes bonds with maturities between 3.50 and 7.49 years, 10 year includes bonds with maturities between 7.50 and 17.49 years and 30 year includes bonds with maturities over 17.50 years. The sectoral (financial, industrial and utilities/telecom) indices each consist of 32 bonds with 8 from each maturity cell. All of the indices are equally weighted. The data span the 12 year period of January 1997 to December 2008 inclusive.

The data on Treasury bond yields is taken from the Federal Reserve H15 report, distributed through WRDS.

3.3.4 Empirical Evidence

Table 3.1 shows the results of regressions of net excess returns to the Dow Jones Composite Corporate Bond Index over the one month Treasury bill return. The first regressor is a measure of volatility, calculated as the sum of squared daily log returns to the index over the previous month, divided by the number of days in the month.

\[ \text{It is important that net, rather than log, returns are used on the left hand side of these regressions. The reason is that volatility enters the mean of the log return (applying Ito's Lemma to (3.11) reveals that } d \log P(X_t, u) = (\mu(u) - \frac{1}{2} v_t(u)^2 v_t(u)) dt + v_t(u)^2 dW_t - \log(1 - L) dZ_t) \text{. If log returns were used, it is then confusing as to whether the regressors are forecasting returns, or forecasting volatility. This problem is particularly relevant given the well known persistence of volatility in financial data and the use of realised volatility as a regressor here.}\]
whether the measure of volatility is able to drive these regressors out of the regression, as the DDTSM described above suggests, or if their inclusion improves our conditional forecasting of excess returns to corporate bonds.

The second regressor is the return forecasting factor from Cochrane and Piazzesi (2005) (CP factor henceforth). This factor is a tent-shaped linear combination of forward rates, and it was used by these authors to forecast excess returns to Treasury bonds. I update their regressions to include data to the end of 2008. The reason for the inclusion of this regressor is that a factor that forecasts Treasury bond returns must forecast returns to all assets; recall from Equation (1.3) in Subsection 1.1.1 that any asset can be thought of as a number of Treasury bonds plus cash flow risk.

Next, since the index dividend yield has been used to forecast returns to stocks (see, for example, Cochrane (1992)) and since most of the expected cash flow variation in the cross section of firms is idiosyncratic, the variation in the index dividend to price ratio is mostly due to varying risk premia (see Vuolteenaho (2002)). For this reason, the index dividend yield calculated using the CRSP Value Weighted Return Indices (NASDAQ, AMEX and NYSE) including and excluding distributions was used but then dropped due to a lack of significance.

As described, if a DDTSM specified by (3.20) generated the data on corporate bond returns one may suspect that the coefficients in the regressions in Table 3.1, other than that of bond return variance (or variables highly correlated with bond variance), should appear insignificant. This is not the case. Other than for the 30 year maturity group, the t-statistics of the coefficient on variance are very small. Further, the largest magnitude sample correlation between the measure of volatility and the CP factor in any of the above regressions is -0.18 and despite this, the CP factor coefficient consistently appears with larger (although still less than the 5% significance level) t-statistics than the variance coefficient. Again, the 30 year maturity group is an exception. These regressions do not, therefore, provide support for this specification of prices of risk.

Next, I re-run the above regressions, but split the corporate bond portfolios by their industry classification. The coefficients and t-statistics are provided in Table 3.3 at the back of this chapter. For Industrial Firms the results tell a similar story to that given above. For Financial Firms and Utility and Telecoms Firms, the coefficient on variance appears in several cases with high t-statistics. However, the CP factor is still not driven out of the regressions. Again, these regressions do not provide support for the specification of prices of risk in (3.3).

73
Table 3.1: Regressions of Excess Returns to Corporate Bonds, January 1997 to December 2008.

Monthly net excess returns to the Dow Jones Composite Corporate Bond Index are regressed on the previous month’s CP factor and an estimate of the previous month’s variance of bond portfolio returns. Monthly variance is measured by the average of squared daily log returns to the portfolio. Asymptotic t-statistics adjusted for heteroskedasticity are shown in the parentheses. There are 144 monthly observations.

<table>
<thead>
<tr>
<th>Maturity (years)</th>
<th>Annualised Mean Excess Return (%)</th>
<th>Coefficient on Variance</th>
<th>Coefficient on CP Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>2.99</td>
<td>1.49</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>(0.44)</td>
<td>(1.55)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2.30</td>
<td>-0.26</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>(-0.16)</td>
<td>(1.61)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2.50</td>
<td>1.00</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>(0.34)</td>
<td>(1.83)</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>3.28</td>
<td>-0.22</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>(-0.11)</td>
<td>(1.37)</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>3.89</td>
<td>3.19</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>(1.67)</td>
<td>(1.30)</td>
<td></td>
</tr>
</tbody>
</table>

3.3.5 Risk Premia for Shocks in Credit Spreads

Given the empirical evidence provided by Duffee (2002) against completely affine DTSMs for pricing Treasury bonds and the results of Cochrane and Piazzesi (2005) and others relating to factors that forecast excess returns to Treasury bonds, one may find the regression results above unsurprising. The reasons are as follows.

First, one should expect corporate bond returns to exhibit similar behaviour to Treasury bonds, and increasingly so as their default risk decreases. A simple Law of One Price argument implies that the price of a corporate bond with no default risk equals that of a Treasury bond. More generally, one may expect corporate bonds and Treasury bonds to have similar dependencies on movements in the short interest rate. Since the prices of interest rate risks in (3.20) are completely affine, the empirical result that conditional risk premia in Treasury bond returns do not depend on the variance of Treasury bond returns therefore leads to the analogous result for corporate bonds.

Second, since the CP factor (and other quantities such as forward spreads (Fama and Bliss (1987)) and yield spreads (Campbell and Shiller (1991))) can forecast excess returns
to Treasury bonds, and both Treasury bonds and corporate bonds both have similar exposure to interest rate risk, this factor must also forecast excess returns to corporate bonds. However, this factor is not correlated with the variance of corporate bond returns and so it will not get driven out of regressions upon the inclusion of an estimate of corporate bond return variance.

The two points above can be summarised as follows. Since the DDTSM tested here is simply a completely affine DTSM plus default risk, the failures of completely affine DTSMs carry over to DDTSMs. Given this discussion, an important question is, how are the risks associated with shocks to a bond’s default intensity only reflected in risk premia? Can market price of risk specifications of the form (3.20) remain useful here, even if such specifications are not able to describe the market prices of shocks to the default free interest rate? Perhaps a clearer way of stating the points raised by these questions is as follows.

Consider a hypothetical economy in which nominal default-free interest rates are deterministic. Here, the holder of a corporate bond in an intensity based framework can earn a risk premium in two ways. First, shocks in default intensities can be priced and second, the default event itself can be priced. In a DDTSM such as Duffee (1999) or Driessen (2005), Subsection 3.3.1 showed that these risk premiums are directly related to the variance of bond returns. The task then is to see if this feature is supported empirically. In reality, however, since default free interest rates are stochastic and the risks associated with them priced, one must separate this component of the risk premium from that associated with stochastic default intensities. This is the topic of the remainder of this subsection.

I make the simplifying assumption that the short credit spread $s_t^Q$ and short rate of interest $r_t$ are independent. Several papers support a negative correlation between spreads and interest rates (Duffie, Saita, and Wang (2007), Longstaff and Schwartz (1995), Duffee (1999) and Driessen (2005) are examples). However, the modelling framework most closely related to that studied here is that in Driessen (2005), and this author notes that the dependence is very small.

Using the above assumption and (3.1) we have that, regardless of the dynamics of $r_t$, conditional on no default by time $t$,

$$\frac{P_t(u)}{P_t(u)} = \mathbb{E}^Q_t \left[ \exp \left( - \int_t^T s_v^Q dv \right) \right].$$

(3.25)
Here, the no-arbitrage result that $P_t(u) = \mathbb{E}_t^Q \left[ \exp \left( -\int_t^T r_v dv \right) \right]$ has been used. Next, assuming, as before, the forms for $s_t^Q$ and $X_t$ in (3.3) and (3.4) respectively, we have that

$$\frac{\dot{P}_t(u)}{P_t(u)} = f(X_t, u) \equiv e^{AP(u) - BP(u)^T X_t}.$$  

(3.26)

The function $A_P(u)$ and the $n$-vector valued function $B_P(u)$ are solutions to the Ricatti ODEs given in Subsection 3.5.1 in the appendix, but with $\delta_0$ and $\delta$ set to 0 and the $n$-vector of zeros, $0_n$, respectively. This is because (3.25) is the same as (3.30) in the appendix, but without the short rate term.

The next step is to apply analysis to $f(X_t, u)$ similar to that that was applied to $P(X_t, u)$. I split the market price of risk vector into two parts; $\Lambda_t = (\Lambda_r^t \Lambda_s^Q)^T$. Here, the elements in the vectors $\Lambda_r^t$ and $\Lambda_s^Q$ form the coefficients of the Brownian motions in the SDF (Equation (3.8)) that drive $r_t$ and $s_t^Q$ respectively. The independence assumption allows this separation. I then impose that

$$\Lambda_s^Q = S_t \xi.$$  

(3.27)

This is the completely affine restriction. The model here is now a completely affine DDTSM. I impose no restrictions on $\Lambda_r^t$ here.

Using analysis similar to that applied to the function $P(X_t, u)$ in Subsection 3.2.2 and using (3.27) it follows that

$$\frac{1}{dt} \mathbb{E}_t^P \left[ \frac{df(X_t, u)}{f(X_t, u)} \right] = -B_P(u)^T \Sigma S_t \Lambda_t - \Gamma s_t^P$$

(3.28)

and

$$\frac{1}{dt} \mathbb{E}_t^P \left[ \left( \frac{df(X_t, u)}{f(X_t, u)} \right)^2 \right] = B(u)^T \Sigma S_t^2 \Sigma^T B(u) = v_{0f}(u) + v_{1f}(u)^T X_t.$$  

(3.29)

Here, $e_{0f}(u)$ and $v_{0f}(u)$ are scalar functions and $e_{1f}(u)$ and $v_{1f}(u)$ are $n$-vector valued functions respectively.

With some abuse of nomenclature, I refer to the quantity $df(X_t, u)/f(X_t, u)$ in (3.28) as a return henceforth. The main point to note above is that the relationship between
expected returns to \( f(X_t, u) \) and the variance of these returns in (3.29) is very similar to that between expected returns to corporate bonds and the variance of these returns that was studied in the previous subsection. However, importantly, in deriving this, I have made no assumptions about the stochastic process followed by the short rate \( r_t \) and the market prices associated with shocks to it, other than that \( r_t \) is independent of the short spread \( s_t^Q \). These equations therefore provide a method to test whether a completely affine specification for the market prices of risk associated with shocks to \( s_t^Q \) is supported empirically, without worrying about assumptions about market prices of interest rate risk, or the empirical failures associated with completely affine DTSMs for Treasury bonds. The approach taken is as follows.

I take corporate bond price data from the 2, 5, and 10 year cells described in Subsection 3.3.3 and data on 2, 5, and 10 year Treasury bond prices. I calculate daily values of \( f(X_t, 2), f(X_t, 5), f(X_t, 10), f(X_t, 30) \) according to (3.26); I divide the corporate bond price by the Treasury bond price of the corresponding maturity.

Next, I regress monthly net returns to \( f(X_t, u) \) on an estimate on the previous month’s volatility calculated, as before, by taking the average of squared daily log returns. The results from these regressions are presented in Table 3.2. Further, the results of analogous regressions split by industry are presented in Table 3.4 at the end of this chapter.

The table shows that the coefficient on variance is statistically significant, except for the 2 year maturity. The results split by industry sector are very similar, implying some robustness to these findings. The only exception is that, for financial firms, the coefficient on variance is statistically significant at every maturity, including the 2 year. These results therefore provide support to completely affine specifications of DDTSMS; the variance is significant in explaining the time variation in the expected return to \( f(X_t, u) \).

Another important result to note is that, in every case that the coefficient on volatility is statistically significant, it is also negative. That is, if the volatility of \( f(X_t, u) \) is high, then the risk premium associated with shocks to default intensities/credit spreads is smaller. This result has the implication that the fraction of the credit spread due to risk premium may be greater for investment grade bonds than for high yield bonds.

---

4The Federal Reserve H15 file discounted the 30 year Treasury constant maturity series in February 2002 and reintroduced it in February 2006. As a result, there is substantial missing data in our period of interest relating to the 30 year yield and so the 30 year maturity cell is dropped from the analysis here.
Table 3.2: Regressions of Returns to $f(X_t, u)$, January 1997 to December 2008.

Monthly net returns to the ratio of the Dow Jones Composite Corporate Bond Indices and the maturity matched Treasury bond are regressed on the previous month’s CP factor and an estimate of the previous month’s variance of returns. Monthly variance is measured by the average of squared daily log returns. Asymptotic t-statistics adjusted for heteroskedasticity are shown in the parentheses. There are 144 monthly observations.

<table>
<thead>
<tr>
<th>Maturity (years)</th>
<th>Corporate Composite</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Constant</td>
</tr>
<tr>
<td>2</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>(1.37)</td>
</tr>
<tr>
<td>5</td>
<td>0.45</td>
</tr>
<tr>
<td></td>
<td>(4.63)</td>
</tr>
<tr>
<td>10</td>
<td>0.47</td>
</tr>
<tr>
<td></td>
<td>(2.42)</td>
</tr>
</tbody>
</table>

can be understood as follows.

Credit spreads tend to be more volatile when their levels are high. Indeed, the fact that short spreads cannot become negative (a negative short spread implies a negative default probability) means that, in any model, the volatility of low short spreads must be small. Given this, one may suspect that the volatility of the credit spreads of high yield bonds is generally greater than that of investment grade bonds. The results here then suggest that the component of the credit spread attributable to risk premia in corporate bonds with lower spread levels and lower spread volatility, namely investment grade bonds, is greater than that in high yield bonds that have high spread levels and volatility. This result has implications in applications such as default prediction and bond portfolio analysis.

Credit spreads provide risk-adjusted default probabilities (after, of course, an adjustment for recovery rates). However, if one’s purpose is default prediction or bond portfolio analysis then knowledge of objective default probabilities is required. Here, an understanding of risk premia is required to adjust credit spreads into objective default probabilities. The results here suggest that this adjustment is smaller for high yield bonds than for investment grade corporate bonds.

The economics underlying the affect described here are not immediately clear. For ex-
ample, in a CCAPM, one normally expects risk premia to increase with return volatility (credit spread volatility contributes to return volatility). To reconcile the results here with a CCAPM, therefore, one must assume that shocks in credit spreads are less correlated with consumption shocks among high yield bonds than among investment grade bonds.

An possible alternative explanation is market segmentation. One may suggest that the holders of investment grade bonds have longer holding periods or are more risk averse than investors in high grade debt. A risk averse investor requires a higher risk premia. A pension fund is one such example. Further exploration of this topic is left to further research.

3.4 Concluding Comments

This chapter has studied an important model that has been used in the recent literature to study risk premia in corporate bond returns, the completely affine DDTSM. Such models provide analytical and econometric tractability. Both corporate bond prices and the probability distributions of the state vector are expressed in closed form (up to a system of ODEs). However, these models also impose restrictions on the form of the time variation in risk premia.

Here, it is shown that such models force a close relationship between risk premia and volatilities. In a completely affine DDTSM, the risk premium associated with shocks to short credit spreads is related to the volatility of short credit spreads. Further, when a completely affine DDTSM is used in conjunction with a completely affine DTSM to price Treasury bonds, risk premia in corporate bond returns are closely related to the volatility of returns.

The empirical analysis provides evidence that completely affine DDTSMs may indeed capture much of the time variation in corporate bond returns. A further interesting result is that the risk premium component of credit spreads in lower quality debt is smaller than that in high quality debt. This suggests that the credit spreads of lower quality debt are better measures of objective default probabilities than in higher quality debt.
3.5 Appendix

3.5.1 Affine Corporate Bond Pricing

The no-arbitrage time \( t \) conditional price of a zero-coupon corporate bond that has not defaulted by time \( t \) and matures at time \( T \) in a recovery of market value framework is

\[
\tilde{P}_t(u) = \mathbb{E}_t^Q \left[ \exp \left( - \int_t^T (r_u + s_u^Q) du \right) \right].
\]  

(3.30)

Under (3.2), (3.3) and (3.4), the price is given by

\[
\tilde{P}_t(u) = \bar{P}(X_t, t) = e^{A(u) - B(u)^T X_t}.
\]

Here, \( u \equiv T - t \) and \( A(u) \) and \( B(u) \) solve the ODEs

\[
\frac{\partial B(s)}{\partial s} = -(K^Q)^T B(s) - \sum_{i=1}^N \frac{1}{2} \left( \left[ \Sigma^T B(s) \right]_i \right)^2 \beta_i + \delta + \gamma, 
\]

(3.31)

\[
\frac{\partial A(s)}{\partial s} = \sum_{i=1}^N \left( \frac{1}{2} \left[ \Sigma^T B(s) \right]_i \right)^2 \alpha_i - B(s)^T K^Q g^Q - \delta_0 - \gamma_0 
\]

(3.32)

subject to the initial conditions that \( A(0) = 0 \) and \( B(0) = 0 \). The notation \( \left[ \Sigma^T B(s) \right]_i \) denotes the \( i \)-th element of the vector \( \Sigma^T B(s) \).
Table 3.3: Regressions of Excess Returns to Corporate Bonds by Industry Sector, January 1997 to December 2008.

Monthly net excess returns to the Dow Jones Corporate Bond Indices by industry sector are regressed on the previous month’s CP factor and an estimate of the previous month’s variance of bond portfolio returns. Monthly variance is measured by the average of squared daily log returns. Asymptotic t-statistics adjusted for heteroskedasticity are shown in the parentheses.

<table>
<thead>
<tr>
<th>Maturity (years)</th>
<th>Annualised Mean Excess Return (%)</th>
<th>Coefficient on Variance</th>
<th>Coefficient on CP Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Financial Firms</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>3.26</td>
<td>0.65</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>(1.34)</td>
<td>(1.37)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2.49</td>
<td>0.54</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>(6.86)</td>
<td>(1.41)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2.51</td>
<td>0.88</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>(3.73)</td>
<td>(1.68)</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>2.86</td>
<td>0.16</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>(1.07)</td>
<td>(1.24)</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>5.09</td>
<td>0.96</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>(0.90)</td>
<td>(1.19)</td>
<td></td>
</tr>
<tr>
<td><strong>Industrial Firms</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>3.11</td>
<td>2.01</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>(0.73)</td>
<td>(1.64)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1.95</td>
<td>0.23</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(1.65)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2.84</td>
<td>-0.37</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>(-0.11)</td>
<td>(1.37)</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>3.35</td>
<td>1.11</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>(0.61)</td>
<td>(1.49)</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>4.31</td>
<td>2.34</td>
<td>0.34</td>
</tr>
<tr>
<td></td>
<td>(1.80)</td>
<td>(1.67)</td>
<td></td>
</tr>
<tr>
<td><strong>Utility and Telecoms Firms</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>2.37</td>
<td>2.43</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>(1.88)</td>
<td>(1.42)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2.21</td>
<td>1.14</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>(1.65)</td>
<td>(2.27)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2.12</td>
<td>2.30</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>(2.09)</td>
<td>(1.75)</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>3.37</td>
<td>1.21</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>(0.87)</td>
<td>(1.58)</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>1.63</td>
<td>1.16</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>(2.38)</td>
<td>(0.73)</td>
<td></td>
</tr>
</tbody>
</table>
Table 3.4: Regressions of Returns to \( f(X_t, u) \) by Industry Sector, January 1997 to December 2008.

Monthly net returns to the ratio of the Dow Jones Composite Corporate Bond Indices by industry sector and the maturity matched Treasury bond are regressed on the previous month’s CP factor and an estimate of the previous month’s variance of bond portfolio returns. Monthly variance is measured by the average of squared daily log returns to the portfolio. Asymptotic t-statistics adjusted for heteroskedasticity are shown in the parentheses. There are 144 monthly observations.

<table>
<thead>
<tr>
<th>Maturity (years)</th>
<th>Coefficient on Constant</th>
<th>Variance</th>
<th>CP Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Financial Firms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.54</td>
<td>-0.52</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>(3.48)</td>
<td>(-2.46)</td>
<td>(1.14)</td>
</tr>
<tr>
<td>5</td>
<td>0.36</td>
<td>-0.32</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>(3.04)</td>
<td>(-5.68)</td>
<td>(0.59)</td>
</tr>
<tr>
<td>10</td>
<td>0.29</td>
<td>-0.37</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>(1.47)</td>
<td>(-3.82)</td>
<td>(0.98)</td>
</tr>
<tr>
<td>Industrial Firms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.54</td>
<td>-1.28</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>(2.22)</td>
<td>(-0.54)</td>
<td>(1.22)</td>
</tr>
<tr>
<td>5</td>
<td>0.55</td>
<td>-2.87</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>(3.89)</td>
<td>(-2.00)</td>
<td>(0.07)</td>
</tr>
<tr>
<td>10</td>
<td>0.65</td>
<td>-1.51</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>(2.85)</td>
<td>(-1.90)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Utility and Telecoms Firms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.09</td>
<td>1.80</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>(0.18)</td>
<td>(1.38)</td>
<td>(0.13)</td>
</tr>
<tr>
<td>5</td>
<td>0.39</td>
<td>-0.76</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>(3.52)</td>
<td>(-1.14)</td>
<td>(1.41)</td>
</tr>
<tr>
<td>10</td>
<td>0.18</td>
<td>-0.13</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>(0.54)</td>
<td>(-0.13)</td>
<td>(0.97)</td>
</tr>
</tbody>
</table>
Chapter 4

Portfolio Credit Risk: A Frailty Approach to Credit Cycles

Chapter Summary

We model aggregate loss rates on credit portfolios dynamically using a default intensity approach. The default intensity we employ is allowed to depend on both observable macroeconomic variables and unobserved frailties. We use the model to extract measures of the credit cycle from US bank charge-off rates and find that unemployment, industrial production (IP) and interest rates have significant and qualitatively plausible effects on aggregate defaults. Using smoothed estimates for the unobserved frailty factors, we characterize the credit cycles driving defaults in the corporate, real-estate and non-mortgage retail sectors.
4.1 Introduction

Understanding how defaults cluster in time (the credit cycle) is an important objective for policy-makers, financial firms and academic researchers. Such cycles have implications for wider economic performance and growth since credit downturns affect the capital adequacy of banks and hence their willingness to extend new credit. They are also important for the pricing and risk analysis of financial securities that depend on the performance of pools of credit exposures such as Collateralized Debt Obligations (CDOs) and other credit-related structured products.

Building on recent research that has studied the relationship between defaults and the macro-economy, we develop a model that can be used to understand credit cycles. We follow an intensity based approach to modeling default that is broadly similar to that used in Duffie, Saita, and Wang (2007) in that the default intensity of individual loans is driven by a set of macroeconomic factors. Building on the analysis of Duffie, Eckner, Horel, and Saita (2009), we suppose that default intensities also depend on unobservable, so-called frailty random processes.

The frailties capture common latent drivers of loan default intensities that cannot be directly associated with observable macro-economic variables. By including frailties, we are able to capture additional co-movement in loan default intensities and thereby to explain additional default clustering. Some justification for the inclusion of a frailty variable comes from Das, Duffie, Kapadia, and Saita (2007) who find that a version of their intensity based model that omits a frailty is unable to explain the observed level of temporal default clustering observed in corporate default data.

Our approach differs from of the above mentioned studies in that our model is designed for use with aggregate default rate data rather than with individual obligor or exposure level data. The approach we follow is particularly well-suited for analyzing the performance of, for example, bank-level data or performance data on structured product pools. We implement our approach using data on charge-off rates from US-banks provided by the United States Federal Reserve Board. Aggregate data has been used for dynamic loan loss distribution modeling in, for example, Lamb and Perraudin (2008).

A feature of our model is that it imposes dynamics on both the observable and frailty covariates. This allows the model to produce estimates of the loan loss distribution over multiple future periods conditional on the observed state of the macro-economy and the history of observed aggregate default rates.
The facts that (i) the frailty process is unobserved and (ii) the data are observed discretely in time whilst the model is formulated in continuous time affects the econometric implementation of the model in that the estimation requires a filtering approach. Specifically, we express the model in state-space form and then apply standard ML techniques based on the Kalman filter for parameter estimation.

One of the main objectives of our research is to determine which macroeconomic variables drive default probabilities. The observable macroeconomic quantities that are used include unemployment, short term interest rates, IP, consumer debt as a ratio to GDP and the housing market index (HMI). For different categories of loans, we find that each of these series has statistical significance in driving default intensities. In almost all cases, the sign of the effect and precisely which variables are significant for a given loan type are intuitively reasonable.

4.1.1 Further Related Literature

Past literature relevant for our study includes: (i) default prediction studies, (ii) analysis of portfolio loss distributions and (iii) studies of how defaults and credit ratings are affected by macroeconomic variables.

On (i) i.e., analysis of the predictability of defaults, early studies include Altman (1968) in which a discriminant analysis approach is used to predict corporate default. Beaver (1968) examined how equity returns and financial ratios together may be used to predict default. Ohlson (1980) employs logit regressions with financial ratio explanatory variables. Lane, Looney, and Wansley (1986) employ a Cox proportional hazards model to predict bank failure. Lee and Urrutia (1996) compare logit and hazard rate approaches to predicting default among property-liability insurers.

the latter performs best.

On (ii), i.e., portfolio loss modeling, an influential practitioner contribution is Vasicek (1987) which derives the limiting loan loss distribution of a large homogeneous pool of correlated credit risks. Vasicek’s approach may be thought of as a multi-firm extension of the seminal corporate bond pricing model of Merton (1974) in which firms’ cash flows are correlated and their default probabilities are all equal. However, as noted by Lamb and Perraudin (2008), the static nature of the Vasicek model means that it is unable to capture the observed predictability in default rates from period to period. In contrast, the model in this chapter is dynamic.

Loss distribution models like that of Vasicek have been widely employed for pricing and rating structured products. An early contribution using copulas to generate correlated default times is Li (2000). Schonbucher and Schubert (2001) looks at other copulas and models with infectious defaults where a default increases default hazards for other borrowers. Giesecke and Goldberg (2005) also look at so-called self-exciting processes where intensities respond to default events. This approach was originally suggested by Davis and Lo (2001).

The Vasicek model and simple copula based models are static in the sense that they do not consistently describe the evolution of information through a well-specified filtration. An early contribution that is fully dynamic is Duffie and Garleanu (2001). More recently, Chapovsky, Rennie, and Tavares (2006) propose a similar model in which individual defaults are driven by a hazard rate equal to the sum of a common random process with known dynamics, such as a CIR process, and a deterministic function calibrated to individual names. Giesecke and Goldberg (2005) develop an intensity based approach to modeling total portfolio losses, inferring single name default processes using ‘thinning’ techniques. Hull and White (2007) develop a model in which each obligor’s hazard rate follows a deterministic process that is subject to periodic shocks.

Recently, Sidenius, Piterbarg, and Anderson (2006) and Schonbucher (2006) employ a framework akin to the Heath-Jarrow-Morton term structure model to describe the full forward distribution of the loss process. Brigo, Pallavicini, and Torresetti (2007) assumes the loss process is a sum of independent Poisson processes that incorporates correlation into the model.

On (iii), i.e., credit market developments and the macro-economy, there has been particular focus on the dependence of default and rating transition rates on the business
cycle. Notable contributions in this area are Nickell, Perraudin, and Varotto (2000), Koopman and Lucas (2005) and Koopman, Lucas, and Klaassen (2005) which examine how transition probabilities and default rates are linked to macro-economic drivers. McNeil and Wendin (2006) and McNeil and Wendin (2007) study corporate defaults and ratings transitions using latent-variable models in which the underlying factors are serially correlated and estimation is performed using Bayesian methods based on the Gibbs sampler and then relate the results to the evolution of the business cycle.

The remainder of this chapter proceeds as follows. Section 4.2 describes the model and the econometric approach. Section 4.3 reports the parameter estimates of the analysis of US bank charge-off rates. Finally, Section 4.5 concludes.

4.2 Model

4.2.1 Model Framework

In a default intensity model, default occurs at the first jump time of a conditional Poisson process. Loosely speaking, for a small time period $\Delta t$, conditional on loan obligor $i$ having not defaulted by time $t$, the probability that default will occur before time $t + \Delta t$ is approximately given by $\lambda_i^t \Delta t$. Here, $\lambda_i^t$ is called the default intensity of obligor $i$. The Poisson process is said to be doubly-stochastic if $\lambda_i^t$ follows a stochastic process. The term doubly-stochastic comes from the fact that both the default time and future values of $\lambda_i^t$ are stochastic. Duffie (2005) provides a detailed description of the intensity approach to modeling default.

In the doubly-stochastic Poisson process for default that we will employ, $\lambda_i^t$ depends on a vector-valued Markov state process, $X_t$. The state vector, $X_t$, consists of both observable macroeconomic variables and unobserved frailty variables. This makes the framework used here similar in principle to Duffie, Eckner, Horel, and Saita (2009).

Frailties induce additional correlation between individual defaults and hence introduce additional variability in aggregate default rates. Alternative ways of introducing default correlation and hence aggregate default rate volatility include contagion, whereby the default of one obligor triggers an increase in the default probability of other borrowers (Davis and Lo (2001)). In our context, contagion does not seem a sensible modeling approach. We apply our model to bank charge-off data for such exposures as credit

cards and consumer loans for which direct default contagion is implausible.

We assume that the portfolio under consideration is homogeneous in that obligors have identical default intensities. That is, \( \lambda_i^t = \Lambda(X_t, t) \) for all \( i \). Of course this does not imply that obligors have the same default times. It implies that any two obligors who have not yet defaulted have the same probability of defaulting by any given time in the future. The homogeneity assumption is not unusual. For instance, the well known and commonly used static loan loss distribution model of Vasicek (1987) makes this assumption. We discuss the case of heterogeneous portfolios in Section 4.4 and show that the homogeneous model derived here can be reinterpreted to account for heterogeneities across obligors.

In doubly-stochastic intensity based default models such as Duffie, Saita, and Wang (2007) it is assumed that, conditional on the path of a fully observable state vector, the random default times of the different obligors are independent of each other. We also make this assumption.

Given sufficient data it is possible in this framework to estimate a rich model with multiple frailty factors and more complex stochastic processes for the frailties to follow. For simplicity and due to limitations on the amount of available data, we restrict attention to the case in which the state process \( X_t \) consists of a single frailty (the first element of \( X_t \), denoted \( X_{1,t} \)) and multiple observable macroeconomic risk factors \( (X_{2,t}, \ldots, X_{n,t}) \).

The next step is to specify (i) the functional form \( \Lambda(X_t, t) \) and (ii) the stochastic process followed by the state variables. Fully specifying the stochastic processes for the state variables is necessary if one is to make forward-looking statements such as calculating the term structure of the loan loss distribution or forecasting future default probabilities. The chosen specifications also have implications for the econometric estimation of the model. This will become clear in the next subsection. We adopt the following assumption.

**Assumption** The default intensity of obligor \( i \) at time \( t \) depends linearly on an \( n \)-dimensional vector \( X_t \) and a time trend:

\[
\lambda_i^t = \Lambda(X_t) \equiv \mu^t X_t + \gamma t, \quad (4.1)
\]

where \( X_t \) follows the multivariate Ornstein-Uhlenbeck process

\[
dX_t = K(\theta - X_t)dt + \Sigma dW_t \quad (4.2)
\]
Here, $W_t = (W_{t,1}, \ldots, W_{t,n})^T$ is an $n$-dimensional vector of independent Brownian motions. $\mu$, $\gamma$, $K$, $\Sigma$ and $\theta$ are parameters to be estimated. $\mu$ is an $n$-vector, $\gamma$ is a scalar, $K$ is a diagonal $n \times n$ matrix, $\Sigma$ is an $n \times n$ matrix and $\theta$ is an $n$-vector. Furthermore, suppose that the frailty variable, $X_{1,t}$, is orthogonal to the remaining state variables ($X_{2,t}, \ldots, X_{n,t}$). That is: $\Sigma_{1,j} = 0$ for $j = 2, \ldots, n$. Given that $X_{1,t}$ is latent, for econometric identification $\mu_1$ is set to 1.

An immediate disadvantage of the above specification where the state variables follow an Ornstein-Uhlenbeck process is that there is a non-zero probability that the intensity $\lambda^i_t$ becomes negative over a discrete time horizon. This could result in negative probabilities of default. However, once the model has been implemented, it will become clear that, for the parameter estimates found, the probability that $\lambda^1_t$ becomes negative is extremely low.

One might also question the Ornstein-Uhlenbeck distributional restrictions imposed on the macroeconomic variables. In principle, the autoregressive dynamics of the processes could be generalized without difficulty and one may employ transformations of the macro variables so that innovations have the properties well approximated by Gaussian distributions.\footnote{Note that it is necessary to specify an explicit process for the macroeconomic variables even though we will estimate the model conditional on these variables. The reason is that since the model is formulated in continuous time yet the data is time-aggregated and discrete time, estimation involves a filtering problem and this requires a full specification of the processes.}

More generally, non-Gaussian processes may also be modeled statistically or more complex functional forms for $\Lambda(\cdot)$ can be used. However, implementing such approaches requires computationally intensive non-linear filtering techniques. In the current study, we follow a simpler approach so that the model may be implemented using a linear Kalman filter, and defer an investigation into non-linear models to future research.

The Ornstein-Uhlenbeck specification for the frailty variable adopted here is common to Duffie, Eckner, Horel, and Saita (2009), although these authors use a proportional hazards form for $\Lambda(\cdot)$. They also provide some justification for Ornstein-Uhlenbeck specification in the context of corporate default. In the context of credit cards and consumer loans this variable could capture, for example, shocks to the costs associated with default. Such shocks may decay over time leading to mean reversion. The mean-reversion parameter $K_{11}$ captures the expected rate of decay of such shocks.
The model also allows for a time trend in default intensities. This feature captures the possibility that a gradual loosening of lending criteria has made the underlying obligor pool more risky. The time trend may also capture the possibility that the social costs associated with default may have decreased over time.

4.2.2 Econometric Implementation

The aim is to use data on aggregate default rates to estimate the model. The purpose of this section is to restate the model so that it can be used for this task. This will require the additional assumption that the pool of obligors is well approximated by its asymptotic limit. The data that is used to estimate the model is from large but finite pools.

It is a standard result relating to doubly stochastic Poisson processes that, conditional on the macroeconomic state \( X_t \) and having not defaulted by time \( t \), the probability that the default time of an obligor \( i \), \( \tau^i \), will be within the next period of time \( \delta \) is

\[
P(\tau^i \leq t + \delta | \tau > t, X_t) = 1 - E \left[ e^{-\int_t^{t+\delta} \Lambda(X_u) du} | X_t \right].
\]

The homogeneity assumption means that the right hand side of the above equation is independent of the obligor \( i \). Conditional on the entire path followed by \( X \), we then have that

\[
P(\tau^i \leq t + \delta | \tau > t, X) = 1 - e^{-\int_t^{t+\delta} \Lambda(X_u) du}.
\] (4.3)

Next, let \( D_{t,t+\delta} \) denote the fraction of the undefaulted loan pool of obligors at \( t \) that default before time \( t+\delta \). Recalling the assumption that, conditional on the path followed by \( X \), defaults are independent, then conditional on the path followed by \( X \), the Law of Large numbers tells us that as the obligor pool approaches its asymptotic limit,

\[
D_{t,t+\delta} = P(\tau^i \leq t + \delta | \tau > t, X)
= 1 - e^{-\int_t^{t+\delta} \Lambda(X_u) du},
\]

where the second line follows from (4.3). If in the data set, aggregate default rates are observed, for example, every quarter then measuring time in years means that \( \delta = 0.25 \) and we observe realizations of the set of random variables \( D_{t,t+0.25} \) for \( t = 0, 0.25, 0.5, \ldots, T - 0.25 \), where \( T \) is the time at which the data set finishes. More detail on the data sets is provided in Subsection 4.3.1 below.
Next, we make a simple transformation and define a more convenient random variable $d_{t,t+\delta}$ by

$$d_{t,t+\delta} = -\log(1 - D_{t,t+\delta}) = \int_t^{t+\delta} \Lambda(X_u) \, du.$$ 

Substituting (4.1) into the equation above gives

$$d_{t,t+\delta} = \int_t^{t+\delta} \mu' X_u + \gamma u \, du.$$ 

A standard integration of the Ornstein-Uhlenbeck state vector gives

$$X_{t+\delta} = (I_n - e^{-K\delta})\theta + e^{K\delta} X_t + \epsilon_{t+\delta}, \tag{4.4}$$

where $e^{-K\delta}$ is the fundamental matrix associated with $-K\delta$, $I_n$ is the $n \times n$ identity matrix and $\epsilon_{t+\delta}$ is a zero mean $n$-dimensional Gaussian shock with $i$-th element

$$\epsilon_{t+\delta,i} = \int_t^{t+\delta} e^{-K_{ii}(t+\delta-u)} \sum_{j=1}^n \Sigma_{ij} \, dW_{u,j}. \tag{4.5}$$

Next, a simple double integration of an Ornstein-Uhlenbeck process means that the time-integrated state variables may be expressed as

$$\int_t^{t+\delta} X_{u,i} \, du = \theta_i \left( \delta + \frac{1}{K_{ii}} \left( e^{-K_{ii}\delta} - 1 \right) \right) + \frac{1}{K_{ii}} \left( 1 - e^{-K_{ii}\delta} \right) X_{t,i} + \eta_{t+\delta,i},$$

where $\eta_{t+\delta,i}$ is a zero mean Gaussian shock,

$$\eta_{t+\delta,i} = \int_t^{t+\delta} \left( 1 - e^{K_{ii}(t+\delta-u)} \right) \sum_{j=1}^n \Sigma_{ij} \, dW_{u,j}. \tag{4.6}$$

Hence, by simple manipulation, we may write the transformed loss rate, $d_{t,t+\delta}$, as

$$d_{t,t+\delta} = \mu' g + \gamma t + \mu' AX_t + \mu' \eta_{t+\delta}. \tag{4.7}$$

Here, $g$ is an $n$-dimensional vector with $i$-th element

$$g_i \equiv \theta_i \left( \delta + \frac{1}{K_{ii}} \left( e^{-K_{ii}\delta} - 1 \right) \right),$$

$A$ is an $n \times n$ diagonal matrix where the $(i,i)$-th element is given by

$$A_{ii} \equiv \frac{1}{K_{ii}} \left( 1 - e^{-K_{ii}\delta} \right).$$
and \( \eta_{t+\delta} \equiv [\eta_{t+\delta,1}, \ldots, \eta_{t+\delta,n}] \).

It may be shown by standard methods that the joint distribution of \( \mu'\eta_{t+\delta} \) and \( \epsilon_{t+\delta} \) is joint Gaussian:

\[
\begin{pmatrix}
\epsilon_{t+\delta} \\
\mu'\eta_{t+\delta}
\end{pmatrix} \sim N(0_{n+1}, \Omega),
\]

(4.8)

where \( 0_{n+1} \) is an \((n+1)\)-vector of zeros. The parametric form of \( \Omega \) is algebraically untidy and so it is relegated to the appendix.

Estimation of this model is complicated by the fact that the first element of the state variable, \( X_{1,t} \), is unobserved. One may also wish to allow for measurement errors in the observed aggregate default rates. An advantage of the model is that it can be cast in standard state space form. This permits one to generalize the model to include observation error and to calculate the likelihood of a data set using the Kalman filter. The generalized model may then be estimated using ML techniques.

4.2.3 State Space Formulation

Using (4.7), (4.4) and (4.8), the observation and state equations required of the state space form may be written as:

**Observation Equation:** \( Y_t = Z\alpha_t + \xi_{t+\delta} \)

**State Equation:** \( \alpha_{t+\delta} = T\alpha_t + c^* + \nu_{t+\delta} \)

where

\[
\begin{align*}
\alpha_{t+\delta} &\equiv \begin{bmatrix} X_{t+\delta} \\ d_{t+\delta} \end{bmatrix}, & \nu_{t+\delta} &\equiv \begin{bmatrix} \epsilon_{t+\delta} \\ \mu'\eta_{t+\delta} \end{bmatrix}, & \nu_{t+\delta} &\sim N(0_n, \Omega), \\
T &\equiv \begin{bmatrix} A & 0_n \\ \mu' & 0 \end{bmatrix}, & c^* &\equiv \begin{bmatrix} c \\ \mu'g + \gamma^*t \end{bmatrix}, & Z &\equiv \begin{bmatrix} 0_n & I_n \end{bmatrix}. 
\end{align*}
\]

(4.11)

Here, \( c \equiv (I_n - e^{K\delta})\theta \) is an \( n \)-vector and \( \xi_{t+\delta} \sim N(0, \sigma^2 I_n) \) is a vector of random, serially uncorrelated, Gaussian measurement errors.

We transform the parameters so as to convert the problem into an unconstrained optimization (enforcing the positivity of variance) and then use standard ML methods for the Kalman filter (see, for example, Chapter 7 of Durbin and Koopman (2001)). These
methods allow us to estimate the parameters of the frailty variable, \( X_{1,t} \), at the same time as the parameters of the observable variables, \( X_{2,t}, X_{3,t}, \ldots, X_{n,t} \).

To calculate a starting parameter vector, we estimate the parameters of the processes followed by the observable variables using an OLS regression. We use these parameters to calculate the covariances between the observable variables, also. The remaining unknown parameters are started at randomly chosen values that are of orders of magnitude that correspond to the unconditional average observed default rates in the data set.

The program is then run until convergence is obtained at a maximum of the likelihood function. To avoid local maxima, this process was repeated using different starting values 500 times. The parameters corresponding to the maximum likelihood value in this set of 500 were taken as the estimates. Finally, following several diagnostic tests (reported below) the standard errors were calculated in the usual way based on the numerically evaluated information matrix.

4.3 Results

4.3.1 Data

The primary data source is quarterly charge-off data for US banks provided on the web site of the United States Federal Reserve Board. The data used runs from 1985 Q1 to 2008 Q2 inclusive and comprises 94 observations. We focus on charge-off rates for business loans, real estate loans, consumer loans, and credit cards. An important point to note is that charge-off rates differ from delinquency or default rates in two main ways:

1. They reflect losses rather than the fraction of borrowers that default. If mean recovery rates are greater than zero, loss rates will be lower than default rates. In the current context there are two ways to tackle this problem. The simplest is to assume that the loss given default is fixed at a value \( L \) that does not vary with time. This means that charge-off rates and default rates are related to each other by a fixed factor and so one can simply reinterpret the parameter vector \( \mu' \) as \( L\mu' \) and implement the model as before. \( \mu \) and \( L \) are then not separately identified.

Another, perhaps more satisfying approach, is to assume that \( \lambda_i^t \) is actually an adjusted default intensity that takes into account the recovery value obtained in the case of a default. That is, \( \lambda_i^t = \lambda_i^{\text{def},t} L_i^t \) where \( \lambda_i^{\text{def},t} \) is the default intensity and \( L_i^t \) is a time varying loss given default fraction for obligor \( i \). The main assumption
behind this approach is that, at the time of default, a fixed fraction of the value of
the loan just before default is lost. This is a standard approach in intensity based
default models (see Duffie and Singleton (1999) for more detail). The analysis then
proceeds just as with pure default rates. $\lambda_{it}^{\text{def},i}$ and $L_{it}^i$ are not separately identified.

2. Charge-off rates are determined not just by economic factors driving defaults by
the bank’s borrowers but also by bank managers’ decisions about when to recognize
losses for accounting and regulatory purposes. This makes charge-off rates more
complex to interpret but may make them a better reflection of the realizations of
credit risk as far as the bank is concerned.

The quarterly macroeconomic data is taken from Thompson Financial DataStream.

Figure 4.3.1 shows time-series plots of the charge-off series we study. The series are
interesting in that they suggest that the different sectors of the credit market exhibit
quite distinct cyclical behavior. All four series show some kind of peak (although of
quite different magnitudes) in the early 1990s and a rise (again of variable magnitude)
at the very end of the sample period. However, real estate charge-offs were very low in
the early 2000s in a period in which the business loan charge-offs were high. The credit
loan and other consumer loan charge-offs appear to have similar turning points but the
relative magnitudes of the peaks and troughs at times are very different indeed.
A feature of the plots in Figure 4.3.1 is that in several cases there are substantial reversals in that the charge-off rate jumps up and then jumps down even more sharply below its starting point. This is true in the late 1980s for real estate loans and in 2006 for credit card loans. These individual observations may well reflect timing effects, in that senior bank managers have decided to write off a large volume of distressed loans for accounting purposes. This appears more likely than the underlying default rates really fluctuating so substantially.

Such timing effects may be thought of as measurement errors. Although we can take into account some types of measurement errors in the state space formulation, such reversals are complex in nature in that they imply negatively autocorrelated errors, perhaps triggered by particular patterns in earlier defaults (such as a gradual run up in defaults). The spikes in the charge-off rate series they represent are also likely to make the observed series look non-Gaussian.

For these reasons, though it reduces the number of observations, we aggregate the quarterly observations and then estimate models using six-monthly and annual time periods. Employing longer time steps also has the advantage that the analysis of risk in banks is almost universally based on one-year horizons. It is easier, therefore, to related our results to other views on bank portfolio risk if we work with annual time steps.

The macroeconomic variables we include are as follows:

2. Industrial Production (IP) growth measured as the change from $t$ to $t + \delta$ in the natural log of the IP index.
4. Housing Market Index (HMI) levels.
5. 3 month Treasury bill rates in levels.

The estimation strategy was as follows. A model was estimated with a single observable variable present both in contemporaneous and one-period lagged form. Depending on which of the two was more significant, the degree of lag for the variable in question was chosen. The full model was estimated using all the variables (with the degree of lag determined as just described). Then, variables were successively eliminated based on which had the smallest t-statistic. At each stage, we checked to see if a Likelihood Ratio test
(with a 1% confidence level) could reject the restriction that the variables dropped had zero coefficients. When this test could be rejected, the process of eliminating variables was stopped.

4.3.2 Estimates

Table 4.1 shows ML estimates of the parameters in the model when applied to business loans, real estate Loans, credit card loans and other consumer loans with annual time steps.

The upper block of numbers shows the loadings on macroeconomic factors (elements of the vector $\mu$). Unemployment is consistently statistically significant and always has the intuitively reasonable sign in that higher unemployment is associated with a higher default rate. The largest unemployment loading occurs in the case of credit card loans. This is consistent with the notion that people use credit cards as a source of short-term financing and default in periods when it is difficult to find work because the aggregate unemployment rate is high.

IP is clearly significant in the case of business and real estate loans, and again has an intuitively reasonable sign. Real estate loans are a mixture of commercial and residential real estate. So IP, which is eliminated from the consumer loan models by the estimation procedure of successive elimination of insignificant variables, may be a significant factor for the corporate credit cycle.

Interest rates are only significant in the other consumer loans category. In this case, the coefficient is positive, implying that higher interest rates lead to greater defaults. Other consumer loans include car and consumer durable loans and so it is plausible that they are interest sensitive in this way.

Lastly the ratio of consumer debt to GDP turns out to be significant for credit card loans. The coefficient is negative, implying the perhaps unlikely result that higher debt is associated with lower defaults. The result is marginally significant and could possibly reflect a non-linear time trend not allowed for in this specification.

The second and fourth blocks of parameters in Table 4.1 exhibit the parameters of the different Ornstein-Ulenbeck risk factors, including both the unobservable frailty variable, $X_{1,t}$ and the observable macroeconomic variables, $(X_{2,t}, \ldots, X_{n,t})$. 

96
The convergence rates that appear in this second block suggest that the frailty variable reverts to its mean relatively slowly. With the exception of interest rates, the macroeconomic variables mean revert at faster rates.

The long run means of the latent variable reflect the average level over time of the default rate itself and hence are much larger for consumer loans, most notably in the case of credit card loans. The volatilities show a similar pattern, being higher for the consumer loans, which is to be expected given the higher fraction of such loans that default.

The lowest block of results in Table 4.1 consists of addendum items. These include time trends that are mostly significant but with signs that depend on the loan series in question. The consumer loan series show a positive trend in defaults while the more corporate loan categories, business and real estate loans, have negative trends.

The four models were estimated assuming that the Observation Equation (Equation (4.9)) contained additive measurement errors (as described in the specification above). However, the standard deviation of the measurement errors was not significant in any of the four cases, and so the model was re-estimated omitting measurement errors and it is the results from these estimations that are reported in Table 4.1.

Lastly the correlations between the macroeconomic factors are reported. These are mostly close to zero and insignificant suggesting that the factors are broadly orthogonal. The exception is the factors for other consumer loans; unemployment and interest rates have a strong correlation.

Using standard Kalman filter smoothing algorithms (see, for example, Chapter 4 of Durbin and Koopman (2001)) one may generate estimates of the unobserved frailty process. Such estimates are shown in Figure 4.3.2.

The frailties extracted from different loan series exhibit some similarities. For example, for all four series, there is a sharp fall in the frailty variable just before 1995. There is also a marked rise in the frailties at the very end of the sample period in all cases except for credit card loans (for which only a small increase occurs).

As remarked when discussing the parameter estimates in Table 4.1, the frailty series inherit some of the properties of the corresponding charge-off rates in that the levels and volatilities of the consumer loan charge-off rates series tend to be higher than those of the corporate charge-off rates.
Figure 4.2: Smoothed estimates of the frailty variable. The dotted lines show the standard errors.
Figure 4.3: Plots of loss distributions over 1, 2, 3, 4 and 5 years. The left most curve shows the 1 year and the right most curve shows the 5 year loan loss distributions conditional on the last data point in the sample, Q2 2008.
Table 4.2 shows diagnostic statistics for the forecast errors of the model using the parameter estimates that were given in Table 4.2. Tests are reported for normality, heteroskedasticity and serial autocorrelation. These tests are explained in Section 2.12 of Durbin and Koopman (2001). In brief, the normality test relies on the test statistic
\[ n \left( \frac{\text{Skewness}^2}{6} + \frac{(\text{Kurtosis} - 3)^2}{24} \right) \]
which, for a null or normality, is asymptotically \( \chi^2 \)-distributed with 2 degrees of freedom, where “skewness” and “kurtosis” are the usual coefficients and \( n \) is number of observations. The heteroskedasticity test compares the ratio of the sum of squared residuals for the first and second half of the sample period. This statistic is asymptotically \( F_{n/2,n/2} \) distributed. For serial correlation, the portmanteau Box-Ljung test is used. This is based on the statistic
\[ n(n + 2) \sum_{j=1}^{k} \frac{c_j^2}{n - j} \]
for an integer \( k \) where \( c_j \) is the sample correlation between residual and their \( j \)-th lag.

It is found that, when the models are implemented with annual data, none of these statistics exceeds the 5% confidence level.

Next, Table 4.3 shows estimates of the state space model for the four loan charge-off rate series implemented on six-monthly data. The advantage of employing a shorter period length is of course the fact that the sample is larger. On the other hand, data distortions arising from banks’ decisions to write off loans in a concentrated fashion are likely to be more of an issue.

Again, the approach of successively dropping the variable with the smallest t-statistic and at each stage checking that one cannot reject the restriction at a 1% confidence level in a likelihood ratio test that the parameters on the variables so far omitted equal zero. This leads to a different set of variables, as one may notice from a comparison of Tables 4.1 and 4.3.

For the models implemented using six-monthly data, it is found that the weights on the frailty variable are similar to those in the case of annual data for the corporate charge of rate series (business and real estate loans) but are somewhat lower in the
case of consumer charge-off rates (credit cards and other consumer). In the case of business loans, GDP growth replaced IP as the second significant factor. It again has the intuitively reasonable sign but the magnitude of the effect is smaller than in the case of IP with annual data.

The convergence parameters of the frailty Ornstein-Uhlenbeck processes (shown in the second block of Table 4.3) are consistently higher than is the case of models estimated using annual. This, together with the fact that the volatilities in the fourth block of Table 4.3 are not consistently higher in the case of six-monthly data) suggests that the frailty variable is more economically significant (it is more persistent), and its higher volatility means that it contributes a larger fraction of overall volatility in the case of annual data.

The addendum items in Table 4.3 show that the trend is only now significant for the consumer charge-off rate series, unlike in the case of annual data. On the other hand, measurement errors turn out to be significant for the consumer charge-off rate series in contrast to the case with annual data.

Finally, Table 4.4 reports diagnostic tests for the models estimated using six-monthly data. The statistics are distinctly higher in this case and indeed normality of the frailty innovations can be rejected at a 5% level and of GDP growth innovations at a 1% level.

### 4.4 A Note on Portfolio Heterogeneity

In the model derived above, the default probabilities of the obligors were assumed to be homogeneous. However, the model is easily reinterpreted to allow for heterogeneity if we choose instead to model the average default probability across obligors. This can be seen as follows.

Suppose that obligors have heterogeneous default probabilities so that \( P(\tau^i \leq t + \delta | \tau^i > t, X) \) now depends on \( i \). The random fraction of the undealted portfolio at time \( t \) that defaults by time \( t + \delta, D_{t,t+\delta} \), then has the property that

\[
\mathbb{E}[ND_{t,t+\delta}|X] = \bar{NP}(t, \delta|X) \quad \text{and} \\
\sigma^2(ND_{t,t+\delta}|X) = \sum_i P(\tau^i \leq t + \delta | \tau^i > t, X)(1 - P(\tau^i \leq t + \delta | \tau^i > t, X))
\]

\[
\leq \bar{N} \max_{\{i\}} P(\tau^i \leq t + \delta | \tau^i > t, X)(1 - P(\tau^i \leq t + \delta | \tau^i > t, X))
\]
where $N$ is the number of obligors and

$$\bar{P}(t, \delta|X) \equiv \frac{1}{N} \sum_{i} P(\tau^i \leq t + \delta | \tau^i > t, X)$$

is the average default probability. Therefore, we have that, conditional on the path of $X$ and as the obligor pool becomes large ($N \to \infty$),

$$D_{t,t+\delta} = \bar{P}(t, \delta|X) \text{ almost surely.}$$

The next step is to model $\bar{P}(t, \delta|X)$ directly. Letting

$$\bar{P}(t, \delta|X) = 1 - e^{-\int_{t}^{t+\delta} \Lambda(X_u)du},$$

then conditional on the path of $X$, $D_{t,t+\delta} = 1 - e^{-\int_{t}^{t+\delta} \Lambda(X_u)du}$ almost surely, as was the case in the homogeneous model. The remaining analysis therefore proceeds as before.

This research does not include any obligor specific/idiosyncratic factors that drive default intensities. However, in this framework one can, in principle, include such factors. Each one becomes a state variable in the model since $\bar{P}(t, \delta|X)$ depends on all the obligor specific default probabilities. The inclusion of such factors comes at the expense of introducing additional unknown parameters into the estimation problem.

### 4.5 Concluding Comments

This chapter has developed and implemented on aggregate US bank charge-off data simple techniques for modeling the dependence of default rates on macroeconomic variables. Intuitively convincing effects are found: unemployment plays a key role in affecting the loan default cycle and other macroeconomic variables affect default rates differently depending on the category of loans.

It is clear, however, that the charge-off rates examined are also significantly affected by unobserved frailties orthogonal to the macroeconomic variables. These frailties contribute significantly to the overall volatility in charge-off rates especially when we employ annual data.

While smoothed estimates of the frailties extracted from charge-off rates for different loan categories are clearly connected (for example, they all fall sharply in 1994, they are
not very closely correlated, suggesting that different parts of the bank credit market are subject to rather different influences. This observation has important implications for the choices banks face in diversifying their portfolios and for regulation.
4.6 Appendix

The \((n+1) \times (n+1)\)-dimensional matrix \(\Omega\) in (4.8) is calculated as follows. The upper left \(n \times n\) block results from a standard integration of the multivariate Ornstein-Uhlenbeck process in that was given in (4.2). For \(i = 1, \ldots, n\) and \(j = 1, \ldots, n\), the \((i,j)\)-th element is given by

\[
\Omega_{ij} = \frac{1 - e^{-((K_{ii} + K_{jj})\Delta)}}{K_{ii} + K_{jj}}(\Sigma \Sigma^T)_{ij},
\]

The remaining elements of \(\Omega\) are calculated by taking the covariances of \(\epsilon_{t+\delta}\) and \(\mu' \eta_{t+\delta}\) and the variance of \(\mu' \eta_{t+\delta}\) using (4.6) and (4.5). For \(i = 1, \ldots, n\),

\[
\Omega_{n+1,i} = \Omega_{i,n+1} = \frac{\mu_i}{\sigma_i} \sum_{k=1}^{n} \Sigma_{ik} \left( \frac{1 - e^{-K_{kk}\delta}}{K_{kk}} - \frac{1 - e^{-(K_{ii} + K_{kk})\delta}}{K_{ii} + K_{kk}} \right)
\]

and

\[
\Omega_{n+1,n+1} = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\mu_i \mu_j \Sigma_{ij}}{K_{ii} K_{jj}} \left( \frac{1 - e^{-(K_{ii} + K_{kk})\delta}}{K_{ii} + K_{kk}} - \frac{1 - e^{-K_{ii}\delta}}{K_{ii}} - \frac{1 - e^{-K_{jj}\delta}}{K_{jj}} + \delta \right).
\]
Table 4.1: Parameter Estimates with Annual Data

<table>
<thead>
<tr>
<th></th>
<th>Business Loans</th>
<th>Real Estate Loans</th>
<th>Credit Cards</th>
<th>Other Consumer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estim SE t</td>
<td>Estim SE t</td>
<td>Estim SE t</td>
<td>Estim SE t</td>
</tr>
<tr>
<td><strong>Loadings of Default Intensity on Macroeconomic Variables ($\mu$)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unemployment</td>
<td>0.47 0.10</td>
<td>0.46 0.10</td>
<td>2.85 0.86</td>
<td>1.22 0.48</td>
</tr>
<tr>
<td>IP</td>
<td>-0.58 0.25</td>
<td>-0.60 0.25</td>
<td>-2.42</td>
<td></td>
</tr>
<tr>
<td>Interest Rate</td>
<td></td>
<td></td>
<td>-2.15</td>
<td>0.31 0.14</td>
</tr>
<tr>
<td>CCO/GDP</td>
<td></td>
<td></td>
<td>-1.04</td>
<td>2.21</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-2.07</td>
<td></td>
</tr>
<tr>
<td><strong>Mean Reversion Rates of Frailty and Macroeconomic Variables ($e^{-K}$)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frailty</td>
<td>0.73 0.15</td>
<td>0.74 0.16</td>
<td>0.84 0.14</td>
<td>0.79 0.22</td>
</tr>
<tr>
<td>Unemployment</td>
<td>0.46 0.17</td>
<td>0.46 0.18</td>
<td>0.43 0.19</td>
<td>0.14 0.16</td>
</tr>
<tr>
<td>IP</td>
<td>0.28 0.17</td>
<td>0.34 0.20</td>
<td>0.70 0.12</td>
<td></td>
</tr>
<tr>
<td>Interest Rate</td>
<td></td>
<td></td>
<td>0.17</td>
<td>5.98</td>
</tr>
<tr>
<td>CCO/GDP</td>
<td></td>
<td></td>
<td>3.37</td>
<td></td>
</tr>
<tr>
<td><strong>Means of Frailty and Macroeconomic Variables ($\theta$)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frailty</td>
<td>1.48 0.18</td>
<td>1.50 0.17</td>
<td>12.06</td>
<td>1.29 0.88</td>
</tr>
<tr>
<td>Unemployment</td>
<td>-0.06 0.21</td>
<td>-0.11</td>
<td>-0.11</td>
<td>-0.21 0.15</td>
</tr>
<tr>
<td>IP</td>
<td>0.28 0.08</td>
<td>0.29</td>
<td>5.35</td>
<td>0.29 0.22</td>
</tr>
<tr>
<td>Interest Rate</td>
<td></td>
<td>3.47</td>
<td></td>
<td>1.30</td>
</tr>
<tr>
<td>CCO/GDP</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Volatilities of Frailty and Macroeconomic Variables (diag($\Sigma$))</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frailty</td>
<td>0.18 0.06</td>
<td>0.18</td>
<td>2.17</td>
<td>0.40 0.21</td>
</tr>
<tr>
<td>Unemployment</td>
<td>0.79 0.13</td>
<td>0.79</td>
<td>0.85</td>
<td>0.17 0.28</td>
</tr>
<tr>
<td>IP</td>
<td>0.44 0.10</td>
<td>0.42</td>
<td>1.63</td>
<td>0.28 5.91</td>
</tr>
<tr>
<td>Interest Rate</td>
<td></td>
<td>4.63</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCO/GDP</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Addendum Items ($\gamma, \Sigma_{ij}$)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trend</td>
<td>-0.04 0.01</td>
<td>-0.04</td>
<td>0.31</td>
<td>0.14 0.04</td>
</tr>
<tr>
<td>Measurement Error</td>
<td>Insignificant</td>
<td>Insignificant</td>
<td>Insignificant</td>
<td>Insignificant</td>
</tr>
<tr>
<td>F1 vs F2 Correl</td>
<td>0.02 0.23</td>
<td>0.10</td>
<td>0.02</td>
<td>-0.64 0.13</td>
</tr>
<tr>
<td></td>
<td>Normality</td>
<td>Heterosk</td>
<td>Serial Corr</td>
<td></td>
</tr>
<tr>
<td>--------------------------------</td>
<td>-----------</td>
<td>----------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td><strong>Business Loans</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frailty</td>
<td>2.75</td>
<td>1.39</td>
<td>7.40</td>
<td></td>
</tr>
<tr>
<td>Unemployment</td>
<td>1.60</td>
<td>1.91</td>
<td>5.91</td>
<td></td>
</tr>
<tr>
<td>IP</td>
<td>3.37</td>
<td>1.74</td>
<td>10.91</td>
<td></td>
</tr>
<tr>
<td><strong>Real Estate Loans</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frailty</td>
<td>2.70</td>
<td>1.42</td>
<td>7.21</td>
<td></td>
</tr>
<tr>
<td>Unemployment</td>
<td>1.02</td>
<td>1.87</td>
<td>7.05</td>
<td></td>
</tr>
<tr>
<td>IP</td>
<td>3.64</td>
<td>1.85</td>
<td>10.93</td>
<td></td>
</tr>
<tr>
<td><strong>Credit Card Loans</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frailty</td>
<td>1.78</td>
<td>1.33</td>
<td>6.02</td>
<td></td>
</tr>
<tr>
<td>Unemployment</td>
<td>1.33</td>
<td>0.64</td>
<td>6.85</td>
<td></td>
</tr>
<tr>
<td>CCO/GDP</td>
<td>0.09</td>
<td>2.23</td>
<td>10.78</td>
<td></td>
</tr>
<tr>
<td><strong>Other Consumer Loans</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frailty</td>
<td>3.05</td>
<td>1.93</td>
<td>5.81</td>
<td></td>
</tr>
<tr>
<td>Unemployment</td>
<td>0.96</td>
<td>2.25</td>
<td>4.03</td>
<td></td>
</tr>
<tr>
<td>Interest Rate</td>
<td>4.28</td>
<td>0.75</td>
<td>6.52</td>
<td></td>
</tr>
</tbody>
</table>
Table 4.3: Parameter Estimates with Six-Monthly Data

<table>
<thead>
<tr>
<th></th>
<th>Business Loans</th>
<th>Real Estate Loans</th>
<th>Credit Cards</th>
<th>Other Consumer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estim</td>
<td>SE</td>
<td>t</td>
<td>Estim</td>
</tr>
<tr>
<td>Loadings of Default Intensity on Macroeconomic Variables ($\mu$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unemployment</td>
<td>0.25</td>
<td>0.06</td>
<td>3.99</td>
<td>0.10</td>
</tr>
<tr>
<td>Interest Rate</td>
<td>-0.03</td>
<td>0.01</td>
<td>-2.69</td>
<td>-0.10</td>
</tr>
<tr>
<td>GDP</td>
<td>-0.06</td>
<td>0.03</td>
<td>-2.20</td>
<td></td>
</tr>
<tr>
<td>HMI</td>
<td>-0.04</td>
<td>0.01</td>
<td>-2.20</td>
<td></td>
</tr>
<tr>
<td>Mean Reversion Rates of Frailty and Macroeconomic Variables ($e^{-0.5K}$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frailty</td>
<td>0.88</td>
<td>0.07</td>
<td>12.77</td>
<td>0.92</td>
</tr>
<tr>
<td>Unemployment</td>
<td>0.46</td>
<td>0.14</td>
<td>3.43</td>
<td>0.22</td>
</tr>
<tr>
<td>Interest Rate</td>
<td>0.91</td>
<td>0.05</td>
<td>17.17</td>
<td>0.91</td>
</tr>
<tr>
<td>GDP</td>
<td>0.25</td>
<td>0.13</td>
<td>1.85</td>
<td></td>
</tr>
<tr>
<td>HMI</td>
<td>0.92</td>
<td>0.08</td>
<td>11.81</td>
<td></td>
</tr>
<tr>
<td>Means of Frailty and Macroeconomic Variables ($\theta$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frailty</td>
<td>0.75</td>
<td>0.11</td>
<td>6.72</td>
<td>0.90</td>
</tr>
<tr>
<td>Unemployment</td>
<td>-0.09</td>
<td>0.10</td>
<td>-0.91</td>
<td>-0.05</td>
</tr>
<tr>
<td>Interest Rate</td>
<td>4.55</td>
<td>1.14</td>
<td>3.98</td>
<td>5.00</td>
</tr>
<tr>
<td>GDP</td>
<td>2.69</td>
<td>0.14</td>
<td>18.64</td>
<td></td>
</tr>
<tr>
<td>HMI</td>
<td>5.24</td>
<td>1.35</td>
<td>3.88</td>
<td></td>
</tr>
<tr>
<td>Volatilities of Frailty and Macroeconomic Variables (diag($\Sigma$))</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frailty</td>
<td>0.08</td>
<td>0.02</td>
<td>4.63</td>
<td>0.06</td>
</tr>
<tr>
<td>Unemployment</td>
<td>0.50</td>
<td>0.07</td>
<td>7.58</td>
<td>0.67</td>
</tr>
<tr>
<td>Interest Rate</td>
<td>0.89</td>
<td>0.10</td>
<td>8.70</td>
<td>0.86</td>
</tr>
<tr>
<td>GDP</td>
<td>1.29</td>
<td>0.25</td>
<td>5.16</td>
<td></td>
</tr>
<tr>
<td>HMI</td>
<td>0.89</td>
<td>0.10</td>
<td>8.95</td>
<td></td>
</tr>
<tr>
<td>Addendum Items ($\gamma, \Sigma_{ij}, \sigma(\xi)$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trend</td>
<td>Insignificant</td>
<td>Insignificant</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>Measurement Error</td>
<td>Insignificant</td>
<td>Insignificant</td>
<td>0.06</td>
<td>0.07</td>
</tr>
<tr>
<td>F1 vs F2 Correlation</td>
<td>-0.13</td>
<td>0.18</td>
<td>-0.74</td>
<td>-0.53</td>
</tr>
<tr>
<td>F1 vs F3 Correlation</td>
<td>-0.26</td>
<td>0.17</td>
<td>-1.56</td>
<td></td>
</tr>
<tr>
<td>F2 vs F3 Correlation</td>
<td>0.37</td>
<td>0.16</td>
<td>2.29</td>
<td></td>
</tr>
</tbody>
</table>
Table 4.4: Diagnostic Tests with Six-Monthly Data

<table>
<thead>
<tr>
<th></th>
<th>Normality</th>
<th>Heterosk</th>
<th>Serial Corr</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Business Loans</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frailty</td>
<td>8.38*</td>
<td>0.88</td>
<td>15.70</td>
</tr>
<tr>
<td>Unemployment</td>
<td>2.85</td>
<td>0.76</td>
<td>2.29</td>
</tr>
<tr>
<td>GDP</td>
<td>24.11**</td>
<td>1.09</td>
<td>19.13</td>
</tr>
<tr>
<td><strong>Real Estate Loans</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frailty</td>
<td>2.37</td>
<td>0.51</td>
<td>2.42</td>
</tr>
<tr>
<td>Interest Rates</td>
<td>19.89</td>
<td>0.60</td>
<td>6.56</td>
</tr>
<tr>
<td>Unemployment</td>
<td>0.79</td>
<td>1.65</td>
<td>7.97</td>
</tr>
<tr>
<td>HMI</td>
<td>9.72</td>
<td>2.01</td>
<td>4.40</td>
</tr>
<tr>
<td><strong>Credit Card Loans</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frailty</td>
<td>1.05</td>
<td>0.69</td>
<td>1.91</td>
</tr>
<tr>
<td>Interest Rates</td>
<td>0.71</td>
<td>1.32</td>
<td>8.24</td>
</tr>
<tr>
<td>Unemployment</td>
<td>3.37</td>
<td>5.69</td>
<td>9.08</td>
</tr>
<tr>
<td><strong>Other Consumer Loans</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frailty</td>
<td>13.92</td>
<td>0.67</td>
<td>7.01</td>
</tr>
<tr>
<td>Unemployment</td>
<td>14.14</td>
<td>19.46</td>
<td>7.82</td>
</tr>
</tbody>
</table>
4.7 Further Research - Extension to a Common Frailty Model

The purpose in this section is to extend the model to the case where there is just one frailty that is common to all four charge-off series. There are at least two reasons for this approach.

First, if one interprets the frailty variable as an unobserved or unmeasured macroeconomic variable that drives charge-off rates, then it may be that this variable is common to all the charge-off series. In the models estimated above, amongst other features of the results, the common trough in all the filtered frailty series in 1994 (see Figure 4.3.2) lends some support to this interpretation. Further, if indeed a single frailty is able to capture the variation in aggregate default rates across the four series here, then we may be able to filter this variable more precisely by using the entire data set to simultaneously estimate this single series.

Second, given the similarity between the respective series in Figures 4.3.2 and 4.3.1, one may suspect that the frailty, rather than being an explanatory variable in its own right, is simply being forced to capture variation in default rates that the observable macroeconomic variables cannot explain. By allowing only one frailty variable to explain all four default rate series, the frailty cannot simply follow one.

The state space formulation is modified as described in the next subsection.

4.7.1 State Space Formulation

The observation and state equations are straightforward extensions to those in (4.9), (4.10), (4.11) and (4.12). I now specify,

\begin{align*}
\text{Observation Equation:} \quad Y_t &= Z \alpha_t + \xi_{t+\delta} \\
\text{State Equation:} \quad \alpha_{t+\delta} &= T \alpha_t + c^* + \nu_{t+\delta}
\end{align*}

\begin{align*}
&\quad \text{(4.13)} \\
&\quad \text{(4.14)}
\end{align*}
where
\[\alpha_{t+\delta} \equiv \begin{bmatrix} X_{t+\delta} \\ d_{t,t+\delta} \end{bmatrix}, \quad \nu_{t+\delta} \equiv \begin{bmatrix} \epsilon_{t+\delta} \\ \mu' \eta_{t+\delta} \end{bmatrix}, \quad \nu_{t+\delta} \sim \mathcal{N}(0_{n+s}, \Omega), \quad (4.15)\]
\[T \equiv \begin{bmatrix} A & 0_{n \times s} \\ \mu' A & 0_s' \end{bmatrix}, \quad c^* \equiv \begin{bmatrix} c \\ \mu' g + \gamma^* t \end{bmatrix}, \quad Z \equiv \begin{bmatrix} 0_{n+s} & I_{n+s} \end{bmatrix}. \quad (4.16)\]

Here, \(d_{t,t+\delta}\) and \(\gamma^*\) are now \(s\)-dimensional vectors and \(\mu\) is an \(n \times s\) dimensional matrix. \(s\) is the number of default rate series used (\(s = 4\) in this chapter). The vector \(d_{t,t+\delta}\) now contains the default rates from each of the charge-off series. The model estimation is carried out exactly as before.

### 4.7.2 Results

The parameter estimates and diagnostic tests are shown in Tables 4.5 and 4.6, and Figure 4.7.2 shows a plot of the smoothed estimate of the common frailty variable.

There are three main points to note here. First, Table 4.5 shows that the same macroeconomic variables appear to be statistically significant in driving default rates as found in the case where each charge-off series was examined separately. One may expect that macroeconomic variables that were previously found to be insignificant in driving default rates now appear to have explanatory power. The reason is that the frailty is now more
smooth (see Figure 4.7.2), does not follow the default rate time series as closely, and so is less able to drive other variables out of the analysis. However, this is not the case.

Second, the estimates of the dependence of default intensities on the macroeconomic variables appears to be relatively robust. Although the coefficients in the matrix $\mu$ have changed size, these changes are not large, and none have changed sign.

Third, the frailty remains important for driving default intensities for credit card and other consumer loans. Table 4.5 shows that the dependence of the estimated default intensity for these two series remains statistically significant. However, in the case of business and real estate loans, this dependence is not statistically significant. Also, although the rate of mean reversion of the frailty variable shown in Table 4.5 is greater than in the previous section, its volatility is still statistically important.

4.7.3 Concluding Comments

The further analysis here has extracted a smoothed frailty variable time series that is common to the four charge-off series used. The variable no longer follows any individual series. However, only two of the four series have statistically significant dependence on the frailty variable here. Interestingly, these are economically related series: credit card and consumer loans. One possibility to further extend this research is to include two frailties. One may expect that the commercial business loan and real estate loan series may depend on a common frailty that is not necessarily closely related to that that the credit card and consumer loan series depend upon.
Table 4.5: Parameter Estimates (Single Model with Common Frailty), Annual Data

<table>
<thead>
<tr>
<th>Business Loans</th>
<th>Real Estate Loans</th>
<th>Credit Cards</th>
<th>Other Consumer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estim</td>
<td>SE</td>
<td>t</td>
</tr>
<tr>
<td>Frailty</td>
<td>0.72</td>
<td>0.50</td>
<td>1.44</td>
</tr>
<tr>
<td>Unemployment</td>
<td>0.32</td>
<td>0.08</td>
<td>4.00</td>
</tr>
<tr>
<td>IP</td>
<td>-0.44</td>
<td>0.25</td>
<td>-1.76</td>
</tr>
<tr>
<td>Interest Rate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCO/GDP</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Loadings of Default Intensity on Frailty and Macroeconomic Variables ($\mu$)

| Frailty        | 0.40  | 0.13 | 3.07 |
| Unemployment   | 0.46  | 0.15 | 3.06 |
| IP             | 0.26  | 0.17 | 1.53 |
| Interest Rate  | 0.76  | 0.14 | 5.43 |
| CCO/GDP        | 0.60  | 0.17 | 3.53 |

Mean Reversion Rates of Frailty and Macroeconomic Variables ($e^{-K}$)

| Frailty        | 2.03  | 0.51 | 3.98 |
| Unemployment   | -0.11 | 0.22 | 0.50 |
| IP             | 0.29  | 0.08 | 3.62 |
| Interest Rate  | 5.15  | 0.70 | 7.35 |
| CCO/GDP        | 0.24  | 0.22 | 1.09 |

Means of Frailty and Macroeconomic Variables ($\theta$)

| Frailty        | 0.63  | 0.15 | 4.20 |
| Unemployment   | 0.77  | 0.14 | 5.50 |
| IP             | 0.48  | 0.08 | 6.00 |
| Interest Rate  | 1.11  | 0.40 | 2.77 |
| CCO/GDP        | 0.66  | 0.12 | 5.50 |

Volatilities of Frailty and Macroeconomic Variables (diag($\Sigma$))
Table 4.6: Diagnostic Tests (Single Model with Common Frailty), Annual Data

<table>
<thead>
<tr>
<th>Business Loans</th>
<th>Normality</th>
<th>Heterosk</th>
<th>Serial Corr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frailty</td>
<td>2.00</td>
<td>1.46</td>
<td>7.01</td>
</tr>
<tr>
<td>Unemployment</td>
<td>1.07</td>
<td>1.86</td>
<td>5.92</td>
</tr>
<tr>
<td>IP</td>
<td>3.40</td>
<td>1.67</td>
<td>9.77</td>
</tr>
<tr>
<td>CCO/GDP</td>
<td>0.09</td>
<td>2.09</td>
<td>9.77</td>
</tr>
<tr>
<td>Interest Rate</td>
<td>3.97</td>
<td>0.77</td>
<td>6.45</td>
</tr>
</tbody>
</table>
Chapter 5

Summary of Contributions

This thesis contributes to the academic literature on dynamic asset pricing applied to fixed income securities. The contributions are to the pricing of Treasury bonds, the pricing of corporate bonds, and the modelling of default rates in portfolios of defaultable debt. This chapter summarises these contributions.

5.1 Pricing Treasury Bonds

Chapter 2 studies risk premia in Treasury bond returns. Specifically, it proposes a GDTSM in which market prices of risk are consistent with the important empirical relation found in the well known Fama-Bliss CRSP Treasury bond data set, that a linear combination of forward rates forecast excess returns to long maturity bonds over short maturity bonds. This relationship was first studied in Cochrane and Piazzesi (2005). Here, I make three main contributions.

First, I show that such market prices of risk are a special case of the essentially affine DTSMS introduced by Duffee (2002). Given this fact, the models described in this thesis are analytically and econometrically tractable.

Second, I introduce a novel method to initialise the ML estimation. Such methods are useful given the large number of local maxima in the likelihood function (see Duffee (2009)). The method builds on the insight of JSZ that, if no restrictions are imposed on market prices of risk, then a GDTSM can be efficiently estimated by OLS. I use the results on invariant transformations from Dai and Singleton (2000) to extend this method to the case where market prices of risk are linear in forward rates.
Third, I compare the forecasting ability of the GDTSM proposed here to that of an unrestricted VAR(1) estimated by OLS. The motivation for this step follows from the result of JSZ that no-arbitrage is irrelevant for forecasting future yields and so forecasts generated by a VAR(1)/OLS are identical to those of a GDTSM without restriction on market prices of risk. Improvements in out-of-sample forecasts here lend support to the form of market prices of risk that are used in the GDTSM here. I find that the model is able to improve in out-of-sample forecasts of the most explanatory principle component of bond yields, the level component.

5.2 Pricing Corporate Bonds

Chapter 3 studies the form of conditional risk premia in a commonly applied default intensity based model for pricing corporate bonds, the completely affine DDTSM. There are two main contributions.

First, I show that completely affine DDTSMs imply that the compensation for risk associated with shocks to default intensities (the credit spread risk premium) is related to the volatility of default intensities. This is an important point. Beginning with the seminal work of Duffee (1999), several authors have estimated completely affine DDTSMs. One benefit of such models is their tractability. Here, I make explicit the restrictions on the form of risk premia that such tractability forces.

It is not clear that the above mentioned restrictions on risk premia should be consistent with empirical fact. Therefore, second, using data on corporate bond indices, I run regressions that suggest that, though restrictive, completely affine DDTSMs may indeed be able to capture some part of the the time variation in risk premia in corporate bond spreads.

5.3 Portfolios of Defaultable Debt

Chapter 4 studies portfolios of defaultable debt. There are two main contributions.

First, I develop a default intensity model that can be used to understand credit cycles. The model is broadly similar to that used in Duffie, Saita, and Wang (2007) in that the default intensity of individual loans is driven by a set of macroeconomic factors. Further, building on the analysis of Duffie, Eckner, Horel, and Saita (2009), I also allow the default intensity to depend on a so-called frailty random process. The aim of the frailty is to capture common latent drivers of loan default intensities that cannot be
directly associated with observable macro-economic variables. The frailty allows us to capture additional co-movement in loan default intensity and explain additional default clustering. Assuming a large homogeneous portfolio, I derive the conditional distribution of the default rate.

Second, I estimate the model on US bank charge-off data in the corporate, real-estate and non-mortgage retail sectors. I find that unemployment, industrial production, and interest rates have significant and quantitatively plausible effects on aggregate default rates.
Bibliography


