Credit Risk Modeling and the Term Structure of Credit Spreads

Li Chen H. Vincent Poor [1]

March 15, 2003 (first draft); November 17, 2003 (this version)

[1]Princeton University, Princeton, NJ 08544, USA. Email Contacts: {lichen, poor}@princeton.edu.

Abstract

In this paper, by applying the potential approach to characterizing default risk, a class of simple affine and quadratic models is presented to provide a unifying framework of valuing both risk-free and defaultable bonds. It has been shown that the established models can accommodate the existing intensity based credit risk models, while incorporating a security-specific credit information factor to capture the idiosyncratic default risk as well as the one from market-wide influence. The models have been calibrated using the integrated data of both treasury rates and the average bond yields in different rating classes. Filtering technique and the quasi maximum likelihood estimator (QMLE) are applied jointly to the problem of estimating the structural parameters of the affine and quadratic models. The asymptotic properties of the QMLE are analyzed under two criteria: asymptotic optimality under the Kullback-Leibler criterion, and consistency. Relative empirical performance of the two models has been investigated. It turns out that the quadratic model outperforms the affine model in explaining the historical yield behavior of both Treasury and corporate bonds, while producing a larger error in fitting cross-sectional bond spread curves. Moreover, a modified fat-tail affine model is also proposed to improve the cross-sectional fitting abilities of the exist models. Meanwhile, our empirical study provides complete estimates of risk-premia for both market risk and credit default risk including jump event risk.

1 Introduction

Research in modeling and pricing of credit default risk has made a great progress during the past few decades. Since the ground breaking studies of Black and Scholes (1973) and Merton (1974), much of the literature has followed their steps to measure a firm's credit risk using its asset value and debt information. Although this "structural" framework has a sound economic interpretation of default, this type of models turns out to be less successful in practical applications due to the difficulties of obtaining the accurate asset value information and providing a realistic default boundary condition. Moreover, since in the traditional structural models, the default of a firm is anticipated by bond holders; or more strictly speaking, the default time is predictable, and thus the model implies the credit spread must approach to zero at the short end, which can hardly match the empirical observations. Although some alternative models have extended the existing framework to overcome this "short-term deviation,"[1] the lack of tractability prevents their extensive implementations in practice.

In response to the imperfections of the structural models, another different modeling strategy is pursued in the recent research works, in which a so-called "reduced-form" model is proposed to price credit risk. Compared with solid economic arguments in the structural models, a default event in the "reduced-form" framework is modeled as a Poisson-type jump which occurs completely unexpectedly. The stochastic structure of default is prescribed by an exogenously given intensity process. As pioneered by Jarrow and Turnbull (1995), Madan and Unal (1994), Lando (1998), and Duffie and Singleton (1999), the reduced-form framework is much more tractable than the structural model in that it provides a tractable credit risk valuation methodology, which has already been widely used in the risk-free interest rate models.

Besides an extensive discussion on how to refine the reduced-form models to better interpret the nature of credit risk, several empirical implementations of this type of models are carried out following the similar approaches in estimating the risk-free term structure models. Duffee (1999) build and test a three-factor affine model using the time-series of treasury rates and the corporate bond yields across investment-graded firms. The model produces a quite good fit for corporate bond yields. However, the model specifies the default risk premium only as a drift adjustment from the diffusion variable, which can not fully explain the excessive return on defaultable bonds. In the light of the work by El Karoui and Martellini (2002), and Jarrow, Lando and Yu (2003), Driessen extends the analysis of Duffee (1999) by introducing a constant jump event risk premium into the original model, the empirical results show that this default jump risk premium is statistically significant and serves as a crucial determinant of excess defaultable bond returns.

However, as the empirical evidence provided by Duffee (2002) and Cheridito, Filipović and Kimmel (2003), a reasonable risk premium structure should be flexible enough to produce a time-varying expected returns. Therefore the assumption of the constant jump risk premium may seem not appropriate. Moreover, Yu (2002) argues that in order to accurately estimate the credit risk premium, it is necessary to know the credit spread that is only caused by default risk, since as shown in Huang and Huang (2003), some non-default factors such as liquidity and tax effects are the main source of the actual spreads observed in the market. In this paper, we develop a potential approach to modeling the default risk. We will show this new framework can not only accommodate the existing reduced-form (intensity-based) credit risk models, but also provide us a more flexible risk premium structure given the recent results in Cheridito and Filipović (2003). As demonstrated by our empirical tests, the default jump risk premium turns out to be time varying and the physical default intensity is mainly determined by firm specific factors rather than the market-wide factors. Moreover, our tests also strongly reject the constant ratio assumption between the real world jump intensity and the risk-neutral one.

For simplicity of empirical implementations, a class of two-factor models is proposed. Both the short rate and a firm-specific credit related factor have been considered to capture the idiosyncratic default risk as well as the one from market-wide influence. The firm-specific variable characterizes the financial quality of a firm, which is called the firm's "credit index." In fact, this concept is originally proposed by Hull and White (2001), in which it is used to measure the distance to default of a firm[2]. By appropriately modeling the credit index of a firm, Chen and Filipović (2003a) proposes a simple model for credit migration and spread curves, in which the explicit formulas of corporate bond prices are derived with consideration of default risk. Furthermore, by adding another default indicator variable, the model has been demonstrated to be a hybrid of a structural and a reduced-form model in that the default can be triggered either by the successive credit downgradings or by an unpredictable jump of default indicator process. With the same setting as in Chen and Filipović (2003a), it is assumed that the higher the credit index value, the worse a firm's financial situation, and the zero-value of the corresponding credit index implies the perfect financial health of a firm. This model assumption allows us to consider the non-default factors, when analyzes the components of the credit spreads.

Besides using the popular affine framework to model these two factors, an alternative quadratic model is also established to compare their empirical performance. As documented in Ahn, Dittmar and Gallant (2002), Chen and Poor (2002), Leippold and Wu (2001, 2002), quadratic models not only empirically outperform the affine model in interpreting the historical Treasury yield movements, but also exhibit a nice analytical tractability comparable to affine models. The extension of existing risk-free quadratic term structure models to incorporate default risk has been discussed in Chen, Filipović and Poor (2003). We apply the filtering technique and quasi-maximum likelihood estimator (QMLE) to estimating both affine and quadratic models using the 25-year time series of Treasury rates and corporate bond yields. Our empirical results show that quadratic models provide a better description than the affine model not only for the dynamics of historical Treasury rates but also for the credit spreads.

Although the time-series properties of the models are usually the main concerns because of their implications in long-term investments and risk management, for the short term trading and pricing purposes, it is more interesting to see the models' cross-sectional term structure fitting ability. To serve this need, we also re-construct one-time Treasury yield curve and credit spread curves for four different rating classes using the current snapshot of more than 700 Treasury and corporate bond prices in the market. Moreover, a new affine fat-tail model is proposed here to capture the fat tail distribution of the short rate, and this fat-tail model together with the affine and quadratic diffusion models have been tested. It turns out that the affine fat-tail model shows the best cross-sectional fitting capacity.

Both the QMLE algorithm that is used to examine the time-series properties of the models and the nonlinear least squares method for fitting the current cross-sectional term structures are finally induced to an optimization problem. With regard to a fair large number of the parameters to estimate, here we apply a genetic algorithm (GA) instead of the gradient search method for the optimization purpose. The genetic algorithm performs a parallel and comprehensive search for the global optimum that uses a set or population, of points to conduct a search, not just a single point in the parameter space. The algorithm views the optimization process as a competition among the population of evolving candidate problem solutions. Four operations are applied to every generation of the search - evaluation, selection, crossover, and mutation. These operations are modeled after the evolutionary process of organisms in nature. This gives us the power to search noisy spaces littered with local optimum points and especially helpful in finding a global optimal solution.

The remainder of the paper is organized as follows. In Section 2, we propose the models used in the testing, and discuss the risk premium specifications and the related bond valuation problems. Section 3 introduces nonlinear filtering technique and the quasi maximum-likelihood estimator used to estimate the quadratic model. In Section 4, we briefly describe the estimation methodology and the dataset we use. The summary of the results on the models' time-series properties is presented in Section 5, while the cross-sectional properties of the models are investigated in Section 6. The appendix describes the asymptotic properties of the quasi maximum-likelihood estimator.

2 The Models

In this section, a class of simple two-factor affine and quadratic models are established for characterizing the joint dynamics of the risk-free spot interest rate and credit migration of a single firm, which provides a unifying framework of pricing both risk-free and defaultable bonds.

Consider a continuous trading economy with finite time horizon $[0, T^*]$ and a filtered probability space $(\Omega, (\mathcal{F}_t)_{0\le t\le T^*}, \mathcal{F}, \mathbb{P})$ representing the randomness in the economy during this time horizon. Let $B(t)$ denote the money market account and $B(t) = \exp\left(\int_0^t r_s ds\right)$, where r_s denotes the risk-free spot rate. Suppose there is no (approximate) arbitrage in the sense of Clark (1993), and thus we have an equivalent martingale measure $\mathbb{Q}$, under which the discounted gain processes on any traded asset is a $\mathbb{Q}$-martingale.

2.1 The Affine Model

Let us consider a two-factor affine model with a state process $Y := (Y^1, Y^2)$ taking values in $\bar{\mathbb{R}}^2_+ := \mathbb{R}^2_+ \cup \Delta$, which is the one-point compactification of $\mathbb{R}^2_+$. The default time is defined by

$$T_{\mathcal{D}} := \inf\{t : Y_t \in \{\Delta\}\}, \tag{1}$$

and $\{\Delta\}$ is an absorbing state. Here Y^1 represents the risk-free short rate up to the default time[3]; i.e.,

$$Y^1_t = r_t, \quad \forall\ t \in [0, T_{\mathcal{D}}], \tag{2}$$

and $Y^2 := Z_t$, where Z_t denotes the credit index of a firm (or index). Under the real world measure $\mathbb{P}$, the joint dynamics of the regular affine process Y is uniquely characterized by its generator

$$\begin{aligned}\mathcal{L}^a_{\mathbb{P}} f(y) &= \alpha_1 y_1 \partial^2_{y_1} f(y) + (\tilde{b}_1 + \tilde{\beta}_1 y_1)\partial_{y_1} f(y) + \alpha_2 y_2 \partial^2_{y_2} f(y) \\ &\quad + (b_2 + \tilde{\beta}_{21} y_1 + \tilde{\beta}_{22} y_2)\partial_{y_2} f(y) - (\tilde{c} + \tilde{\gamma}_1 y_1 + \tilde{\gamma}_2 y_2) f(y), \quad \forall\ f \in C^2_2(\mathbb{R}^2_+),\end{aligned} \tag{3}$$

where α_1, b_2, $\tilde{\beta}_{21}$, α_2, $\tilde{c}$, $\tilde{\gamma}_1$, and $\tilde{\gamma}_2$ are all positive, and

$$\tilde{b}_1 \ge \alpha_1.$$

The first two terms on the right hand side (RHS) of (3) characterize the diffusion and drift of the short rate, which are same as the Cox-Ingersoll-Ross model (1985). The third and fourth terms characterize the dynamics of the credit index process Y^2, which admits the correlation with the short rate Y^1 given by the mean reversion level $\beta_{21}Y^1$. The last term on the RHS represents the potential used for modeling the default risk. The default intensity depends on firm-specific factor Y^2 and market-wide factor Y^1, as given by killing rate $\tilde{c} + \tilde{\gamma}_1 Y^1 + \tilde{\gamma}_2 Y^2$.

Remark 2.1. *By convention, each measurable function f on $\mathbb{R}^2_+$ (or $\mathbb{R}^2$) is extended into $\bar{\mathbb{R}}^2_+$ (or $\bar{\mathbb{R}}^2$) by setting $f(\Delta) = 0$. Especially, we can write $1_{\{Y \neq \Delta\}} = e^{\langle 0, Y \rangle}$.*

In order to calculate the default probability and bond prices, we need to apply the basic affine property frequently, which is given by the following lemma. A general proof can be found in Duffie, Filipović and Schachermayer (2003, Theorem 2.7).

Lemma 2.2. *Given the affine state process defined in* (3) *under $\mathbb{P}$, for any $u, v \in R^2_-$, we have*

$$\mathbb{E}^{\mathbb{P}} \left[e^{\int_t^T \langle u, Y_s \rangle ds} e^{\langle v, Y_T \rangle} \mid \mathcal{F}_t \right] = e^{\varphi_{\mathrm{a}}(T-t) + \psi_{\mathrm{a}}(T-t) Y_t^1 + \phi_{\mathrm{a}}(T-t) Y_t^2}, \tag{4}$$

where $\varphi_{\mathrm{a}}(t)$, $\psi_{\mathrm{a}}(t)$, and $\phi_{\mathrm{a}}(t)$ solve the Riccati equations

$$\begin{aligned} \partial_t \varphi_{\mathrm{a}} &= \tilde{b}_1 \psi_{\mathrm{a}} + b_2 \phi_{\mathrm{a}} - \tilde{c}, \quad \varphi_{\mathrm{a}}(0) = 0, \\ \partial_t \psi_{\mathrm{a}} &= \alpha_1 (\psi_{\mathrm{a}})^2 + \tilde{\beta}_1 \psi_{\mathrm{a}} + \tilde{\beta}_{21} \phi_{\mathrm{a}} - \tilde{\gamma}_1 + u_1, \quad \psi_{\mathrm{a}}(0) = v_1, \\ \partial_t \phi_{\mathrm{a}} &= \alpha_2 (\phi_{\mathrm{a}})^2 + \tilde{\beta}_{22} \phi_{\mathrm{a}} - \tilde{\gamma}_2 + u_2, \quad \phi_{\mathrm{a}}(0) = v_2. \end{aligned} \tag{5}$$

Therefore given the affine model with the state process Y defined in (3), the default probability at time t with maturity $T (\geq t)$ is given by

$$\begin{aligned} \Pi_a(t,T) &= 1 - \mathbb{E}^{\mathbb{P}}[1_{\{Y_T \neq \Delta\}}] \\ &= 1 - e^{\varphi^{\mathrm{p}}_{\mathrm{a}}(T-t) + \psi^{\mathrm{p}}_{\mathrm{a}}(T-t) Y_t^1 + \phi^{\mathrm{p}}_{\mathrm{a}}(T-t) Y_t^2} 1_{\{T_{\mathcal{D}} > t\}}, \end{aligned} \tag{6}$$

where $\varphi^{\mathrm{p}}_{\mathrm{a}}(t)$, $\psi^{\mathrm{p}}_{\mathrm{a}}(t)$, and $\phi^{\mathrm{p}}_{\mathrm{a}}(t)$ solve (5) with $u = v = 0$.

2.2 Measure Change and Risk Premia Specifications

Using the potential approach to modeling default risk facilitates us to consider the different types of risk premia. Following the work done by El Karoui and Martellini (2002) and Jarrow, Lando and Yu (2003), the risk involved in credit products usually can be decomposed into three parts: market risk premium, diffusion default risk premium (through the drift adjustment from the diffusion) and jump default risk premium (through the jump intensity adjustment). In order to capture these three types of risk, we apply the recent results on the measure change for jump-diffusion (possibly non-conservative) processes in Cheridito and Filipović (2003), as given by the following Lemma.

Lemma 2.3. *Given state process Y as defined in* (3) *and suppose $\tilde{b}_1 \geq \alpha_1$ and $\tilde{b}_2 \geq \alpha_2$, then for each $b_1, \beta_{21}, c, \gamma_1, \gamma_2 \in \mathbb{R}_+$ and $\beta_1, \beta_{22} \in \mathbb{R}$, satisfying*

$$b_1 \geq \alpha_1,$$

there exists an equivalent probability measure $\mathbb{Q}$, under which the dynamics of state process Y is given by

$$\begin{aligned}\mathcal{L}^a_{\mathbb{Q}} f(y) &= \alpha_1 y_1 \partial^2_{y_1} f(y) + (b_1 + \beta_1 y_1)\partial_{y_1} f(y) + \alpha_2 y_2 \partial^2_{y_2} f(y) \\ &\quad + (b_2 + \beta_{21} y_1 + \beta_{22} y_2)\partial_{y_2} f(y) - (c + \gamma_1 y_1 + \gamma_2 y_2) f(y), \quad \forall\ f \in C^2_2(\mathbb{R}^2_+).\end{aligned} \tag{7}$$

From Lemma 2.3, it is easy to see that the market risk premium is given by $\frac{1}{\sqrt{2\alpha_1 y_1}}[\tilde{b}_1 - b_1 + (\tilde{\beta}_1 - \beta_1)y_1]$; the diffusion default risk premium is represented by $\frac{\sqrt{y_2}}{\sqrt{2\alpha_2}}[(\tilde{\beta}_{22} - \beta_{22})] + \frac{\sqrt{y_1}}{\sqrt{2\alpha_1}}[\tilde{\beta}_{21} - \beta_{21}]$; and the jump risk premium is equal to $\frac{\tilde{c} + \tilde{\gamma}_1 Y^1 + \tilde{\gamma}_2}{c + \gamma_1 y_1 + \gamma_2 y_2}$. As argued by Jarrow et al. (2003), by adding a systematic jump risk premium, the model has a more flexible structure of default risk premia. For example, the model can imply a larger instantaneous intensity, and thus a larger spread as maturity approaches zero, which is demonstrated in Yu (2002). It will also generate a higher volatility in the intensity process suggesting larger fluctuations in yield spreads than what can be inferred from fluctuations of observed default intensities alone.

2.3 Valuing Treasury and Corporate Bonds

Given the dynamics of state process Y under the equivalent martingale measure $\mathbb{Q}$, we can derive the treasury bond prices and corporate bond prices of a firm using the affine property specified in Lemma 2.2.

Proposition 2.4. *Given the affine model* (7)*, the time t-price of a zero-coupon treasury bond with maturity $T \geq t$ is given by*

$$P^{\mathrm{tr}}_{\mathrm{a}}(t,T) = \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_t^T r_s ds} \mid \mathcal{F}_t\right] = e^{\varphi^{\mathrm{tr}}_{\mathrm{a}}(T-t) + \psi^{\mathrm{tr}}_{\mathrm{a}}(T-t) r_t} \tag{8}$$

where

$$\varphi_{\rm a}^{\rm tr}(t) = \frac{b_1}{\alpha_1}\log\left(\frac{2\rho_0 e^{\frac{1}{2}(\rho_0-\beta_1)t}}{(\rho_0-\beta_1)(e^{\rho_0 t}-1)+2\rho}\right),$$
$$\psi_{\rm a}^{\rm tr}(t) = -\frac{2(e^{\rho_0 t}-1)}{(\rho_0-\beta_1)(e^{\rho_0 t}-1)+2\rho_0},$$

with $\rho_0 = \sqrt{\beta_1^2+4\alpha_1}$, *and the corporate bond price with zero-recovery assumption can be obtained as*

$$\begin{aligned} P_{\rm a}^{\rm co}(t,T) &= \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_t^T r_s\,ds}1_{\{T\le T_{\mathcal{D}}\}} \mid \mathcal{F}_t\right] \\ &= \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_t^T Y_u^1\,du}1_{\{Y_T\neq\Delta\}} \mid \mathcal{F}_t\right] \\ &= e^{\varphi_{\rm a}^{\rm co}(T-t)+\psi_{\rm a}^{\rm co}(T-t)Y_t^1+\phi_{\rm a}^{\rm co}(T-t)Y_t^2}1_{\{T_{\mathcal{D}}>t\}}, \end{aligned} \tag{9}$$

where $\varphi_{\rm a}^{\rm co}(t)$, $\psi_{\rm a}^{\rm co}(t)$ *and* $\phi_{\rm a}^{\rm co}(t)$ *solve the Riccati equations:*

$$\begin{aligned} &\partial_t\varphi_{\rm a}^{\rm co} = b_1\psi_{\rm a}^{\rm co}+b_2\phi_{\rm a}^{\rm co}-c, \quad \varphi_{\rm a}^{\rm co}(0)=0, \\ &\partial_t\psi_{\rm a}^{\rm co} = \alpha_1(\psi_{\rm a}^{\rm co})^2+\beta_1\psi_{\rm a}^{\rm co}+\beta_{21}\phi_{\rm a}^{\rm co}-\gamma_1-1, \quad \psi_{\rm a}^{\rm co}(0)=0, \\ &\partial_t\phi_{\rm a}^{\rm co} = \alpha_2(\phi_{\rm a}^{\rm co})^2+\beta_{22}\phi_{\rm a}^{\rm co}-\gamma_2, \quad \phi_{\rm a}^{\rm co}(0)=0. \end{aligned} \tag{10}$$

Remark 2.5. *Given Proposition 2.4, the corresponding treasury yields and corporate bond yields are derived as*

$$D_{\rm a}^{\rm tr}(t,T,r_t) = -\frac{1}{T-t}[\varphi_{\rm a}^{\rm tr}(T-t)+\psi_{\rm a}^{\rm tr}(T-t)r_t], \tag{11}$$
$$D_{\rm a}^{\rm co}(t,T,Y_t) = -\frac{1}{T-t}[\varphi_{\rm a}^{\rm co}(T-t)+\psi_{\rm a}^{\rm co}(T-t)Y_t^1+\phi_{\rm a}^{\rm co}(T-t)Y_t^2]. \tag{12}$$

And then the spread is given by

$$S_{\rm a}(t,T,Y_t) = D_{\rm a}^{\rm co}(t,T,Y_t) - D_{\rm a}^{\rm tr}(t,T,r_t). \tag{13}$$

Moreover the coefficient function $\varphi_{\rm a}^{\rm co}$ *has an explicit formula as*

$$\phi_{\rm a}^{\rm co}(t) = -\frac{2\gamma_2(e^{\rho_1 t}-1)}{(\rho_1-\beta_{22})(e^{\rho_1 t}-1)+2\rho_1},$$

with $\rho_1 = \sqrt{\beta_{22}^2+4\alpha_2\gamma_2}$.

As defined in (7), the affine state process Y with potential can accommodate an intensity based default risk model in the sense of Lando (1998), and Duffie and

Singleton (1999). Indeed, let $\tilde{Y}$ denote the conservative regular affine process with generator

$$\tilde{\mathcal{L}}^a_{\mathbb{Q}} f(\tilde{y}) = \alpha_1 \tilde{y}_1 \partial^2_{\tilde{y}_1} f(\tilde{y}) + (b_1 + \beta_1 \tilde{y}_1)\partial_{\tilde{y}_1} f(\tilde{y}) + \alpha_2 \tilde{y}_2 \partial^2_{\tilde{y}_2} f(\tilde{y}) \\ + (b_2 + \beta_{21}\tilde{y}_1 + \beta_{22}\tilde{y}_2)\partial_{\tilde{y}_2} f(\tilde{y}), \quad \forall\, f \in C^2_2(\mathbb{R}^2_+),$$

with which we construct a two-factor risk-free affine model. Then we define the default time (on an enlarged probability space)

$$T_{\{\mathcal{D}\}} := \inf\left\{t \;\middle|\; \int_0^t \lambda(\tilde{Y}_s)ds \geq e\right\},$$

where e is a standard exponential random variable which is independent of $\mathcal{F}$, and

$$\lambda(\tilde{Y}) := \eta(c + \gamma_1 \tilde{Y}^1 + \gamma_2 \tilde{Y}^2)$$

models the intensity of default. By Proposition 2.4, the time t price (9) equals

$$\mathbb{E}^{\mathbb{Q}}\left[e^{-\int_t^T r_s ds} e^{-\int_t^T \lambda(\tilde{Y}_s)ds}\, \mathbf{1}(\tilde{Y}_T)\middle| \mathcal{F}_t\right]$$

which is the same as

$$\begin{aligned} \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_t^T r_s ds}\, \mathbb{E}^{\mathbb{Q}}\left[\int_t^T \lambda(\tilde{Y}_s)ds < e \mid \mathcal{F}\right] |\mathcal{F}_t\right] &= \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_0^T r_s ds} 1_{\{T \leq T_{\mathcal{D}}\}}\right] \\ &= P^{\text{co}}_{\text{a}}(t,T). \end{aligned}$$

Remark 2.6. *To accommodate the model proposed in Duffie and Singleton (1999), one can take $\lambda(\tilde{Y})$ as the product of the hazard rate and the loss function.*

Therefore we can see that the potential approach is essentially equivalent to the intensity based models for credit risk.

2.4 Alternative Affine Fat-Tail Model

As argued in the current literature (e.g., see Brigo and Mercurio (2001)), the traditional term structure models can not produce the market-implied short rate distribution as far as the tail distribution is concerned. Instead of the affine diffusion model in (7), here we replace the diffusion part by a specific jump in order to increase the weight in the tails of the state process Y. The basic jump measure introduced here is of the form

$$\mu_\zeta(d\xi) := \frac{\zeta(\zeta-1)}{\Gamma(2-\zeta)} \frac{d\xi}{\xi^{1+\zeta}}, \quad \forall\, \zeta \in (1,\, 2),$$

where $\Gamma(\cdot)$ denotes the Gamma function. This jump measure satisfies

$$\int_0^\infty \left(\xi \wedge \xi^2\right) \mu_\zeta(d\xi) < \infty.$$

Therefore by Duffie et al. (2003, Lemma 9.2), replacing the diffusion part $\alpha_i y_i \partial^2_{y_i} f(y)$ in (7) by the jump

$$\alpha_i y_i \int_{\mathbb{R}_{++}} \left(f(y+\xi e_i) - f(y) - \partial_{y_i} f(y)\xi\right) \mu_{\zeta_i}(d\xi),$$

for $i \in \{1,2\}$, leads to another positive affine process. Moreover as justified in Chen and Filipović (2003b), the alternative model produces a heavier tail distribution of Y_t^i in general and the smaller ζ_i, the more weight in the tail of Y_t^i.

Since

$$\int_0^\infty \left(e^{v\xi} - 1 - v\xi\right) \mu_\zeta(d\xi) = (-v)^\zeta, \quad v \in \mathbb{R}_-,$$

according to Duffie et al. (2003, Theorem 2.7), the treasury bond price at time t with maturity $T \geq t$ is given by [4]

$$P_{\mathrm{a}}^{\mathrm{tr}}(t,T) = e^{\varphi_{\mathrm{af}}^{\mathrm{tr}}(T-t) + \psi_{\mathrm{af}}^{\mathrm{tr}}(T-t) r_t} \tag{14}$$

where

$$\begin{aligned} \partial_t \varphi_{\mathrm{af}}^{\mathrm{tr}} &= b_1 \psi_{\mathrm{af}}^{\mathrm{tr}}, \quad \varphi_{\mathrm{af}}^{\mathrm{tr}}(0) = 0, \\ \partial_t \psi_{\mathrm{af}}^{\mathrm{tr}} &= \alpha_1 (-\psi_{\mathrm{af}}^{\mathrm{tr}})^{\zeta_1} + \beta_1 \psi_{\mathrm{af}}^{\mathrm{tr}} - 1, \quad \psi_{\mathrm{af}}^{\mathrm{tr}}(0) = 0. \end{aligned}$$

Moreover, the corporate bond price with zero-recovery assumption can be obtained as

$$P_{\mathrm{af}}^{\mathrm{co}}(t,T) = e^{\varphi_{\mathrm{af}}^{\mathrm{co}}(T-t) + \psi_{\mathrm{af}}^{\mathrm{co}}(T-t) Y_t^1 + \phi_{\mathrm{af}}^{\mathrm{co}}(T-t) Y_t^2} 1_{\{T_{\mathcal{D}} > t\}} \tag{15}$$

where $\varphi_{\mathrm{af}}^{\mathrm{co}}(t)$, $\psi_{\mathrm{af}}^{\mathrm{co}}(t)$ and $\phi_{\mathrm{af}}^{\mathrm{co}}(t)$ solve the Riccati equations:

$$\begin{aligned} \partial_t \varphi_{\mathrm{af}}^{\mathrm{co}} &= b_1 \psi_{\mathrm{af}}^{\mathrm{co}} + b_2 \phi_{\mathrm{af}}^{\mathrm{co}} - c, \quad \varphi_{\mathrm{af}}^{\mathrm{co}}(0) = 0, \\ \partial_t \psi_{\mathrm{af}}^{\mathrm{co}} &= \alpha_1 (-\psi_{\mathrm{af}}^{\mathrm{co}})^{\zeta_1} + \beta_1 \psi_{\mathrm{af}}^{\mathrm{co}} + \beta_{21} \phi_{\mathrm{af}}^{\mathrm{co}} - \gamma_1 - 1, \quad \psi_{\mathrm{af}}^{\mathrm{co}}(0) = 0, \\ \partial_t \phi_{\mathrm{af}}^{\mathrm{co}} &= \alpha_2 (-\phi_{\mathrm{af}}^{\mathrm{co}})^{\zeta_2} + \beta_{22} \phi_{\mathrm{af}}^{\mathrm{co}} - \gamma_2, \quad \phi_{\mathrm{af}}^{\mathrm{co}}(0) = 0. \end{aligned} \tag{16}$$

Finally, it should be noted that the limit case $\zeta_i \to 2$, for each $i \in \{1,2\}$, corresponds to the diffusion setup (7).

2.5 The Quadratic Model

As mentioned before, since our main focus is on the comparison of the empirical performance between the affine and quadratic models, we also propose a two-factor quadratic model with the same number of parameters as the previous affine models. Under the real world measure $\mathbb{P}$, the state process $X := (X^1, X^2)$ follows a regular quadratic process[5] in the state space $\bar{\mathbb{R}}^2 := \mathbb{R}^2 \cup \Delta$ with the infinitesimal generator

$$\begin{aligned}\mathcal{L}^q_{\mathbb{P}} f(x) = {} & \alpha_1 \partial^2_{x_1} f(x) + (\tilde{b}_1 + \tilde{\beta}_1 x_1)\partial_{x_1} f(y) \\ & + \alpha_2 \partial^2_{x_2} f(x) + (b_2 + \tilde{\beta}_{21} x_1 + \tilde{\beta}_{22} x_2)\partial_{x_2} f(x) \\ & - (\tilde{c} + \tilde{\gamma}_1 (x_1)^2 + \tilde{\gamma}_2 (x_2)^2) f(x), \quad \forall\ f \in C^2_2(\mathbb{R}^2),\end{aligned} \tag{17}$$

where $\tilde{\alpha}_1$, $\tilde{\alpha}_2$, $\tilde{c}$, $\tilde{\gamma}_1$, and $\tilde{\gamma}_2$ are all positive. For simplicity it is further assumed that the risk-free short rate $r_t := (X^1_t)^2$ up to the default time and the credit index is given by $Z_t := (X^2_t)^2$. It is easy to see that this quadratic framework embedded the short rate model as a generalized SAINTS model (e.g. see Ahn (1997)), and the default jump intensity given by $\tilde{c} + \tilde{\gamma}_1 (x_1)^2 + \tilde{\gamma}_2 (x_2)^2$. Since we can write the default indicator function as

$$1_{\{T_{\mathcal{D}} > t\}} = e^{\langle 0, X_t \rangle},$$

following the quadratic property (see e.g. Chen, Filipović and Poor (2003), Theorem 3.4), we can accordingly derive the default probability $\Pi_q(t, T)$ as given by

$$\Pi_q(t,T) = 1 - e^{\varphi^{\mathrm{p}}_{\mathrm{q}}(T-t) + \langle \Psi^{\mathrm{p}}_{\mathrm{q}}(T-t), X_t \rangle + \langle \Phi^{\mathrm{p}}_{\mathrm{q}}(T-t) X_t, X_t \rangle} 1_{\{T_{\mathcal{D}} > t\}}, \tag{18}$$

where $\varphi^{\mathrm{p}}_{\mathrm{q}}(t) \in \mathbb{R}$, $\Psi^{\mathrm{p}}_{\mathrm{q}}(t) \in \mathbb{R}^2$, and $\Phi^{\mathrm{p}}_{\mathrm{q}}(t) \in \mathbb{R}^{2\times 2}$ solve the Riccati equations

$$\begin{aligned}\partial_t \varphi^{\mathrm{p}}_{\mathrm{q}} &= \langle \alpha \Psi^{\mathrm{p}}_{\mathrm{q}}, \Psi^{\mathrm{p}}_{\mathrm{q}} \rangle + 2tr(\alpha, \Phi^{\mathrm{p}}_{\mathrm{q}}) + \langle \tilde{b}, \Psi^{\mathrm{p}}_{\mathrm{q}} \rangle - \tilde{c}, \quad \varphi^{\mathrm{p}}_{\mathrm{q}}(0) = 0, \\ \partial_t \Psi^{\mathrm{p}}_{\mathrm{q}} &= 4\alpha \Phi^{\mathrm{p}}_{\mathrm{q}} \Psi^{\mathrm{p}}_{\mathrm{q}} + (\beta)' \Psi^{\mathrm{p}}_{\mathrm{q}} + 2\Phi^{\mathrm{p}}_{\mathrm{q}} \tilde{b}, \quad \Psi^{\mathrm{p}}_{\mathrm{q}}(0) = 0, \\ \partial_t \Phi^{\mathrm{p}}_{\mathrm{q}} &= 4\Phi^{\mathrm{p}}_{\mathrm{q}} \alpha \Phi^{\mathrm{p}}_{\mathrm{q}} + \Phi^{\mathrm{p}}_{\mathrm{q}} \tilde{\beta} + (\tilde{\beta})' \Phi^{\mathrm{p}}_{\mathrm{q}} - \tilde{\gamma}_q, \quad \Phi^{\mathrm{p}}_{\mathrm{q}}(0) = 0,\end{aligned} \tag{19}$$

where

$$\alpha = \begin{pmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{pmatrix}, \quad \tilde{b} = \begin{pmatrix} \tilde{b}_1 \\ \tilde{b}_2 \end{pmatrix}, \quad \tilde{\beta} = \begin{pmatrix} \tilde{\beta}_1 & 0 \\ \tilde{\beta}_{21} & \tilde{\beta}_{22} \end{pmatrix}, \quad \tilde{\gamma}_q = \begin{pmatrix} \tilde{\gamma}_1 & 0 \\ 0 & \tilde{\gamma}_2 \end{pmatrix}.$$

Under the risk-neutral measure $\mathbb{Q}$, it is assumed that state process X also follows a regular quadratic process with generator

$$\begin{aligned}\mathcal{L}^q_{\mathbb{Q}} f(x) &= \alpha_1 \partial^2_{x_1} f(x) + (b_1 + \beta_1 x_1)\partial_{x_1} f(y) \\ &\quad + \alpha_2 \partial^2_{x_2} f(x) + (b_2 + \beta_{21} x_1 + \beta_{22} x_2)\partial_{x_2} f(x) \\ &\quad - (c + \gamma_1 (x_1)^2 + \gamma_2 (x_2)^2) f(x), \quad \forall\ f \in C^2_2(\mathbb{R}^2_+).\end{aligned} \tag{20}$$

Then following the quadratic property, we can also derive the treasury bond prices and corporate bond prices.

$$P^{\text{tr}}_{\text{q}}(t,T) = \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_t^T r_s ds} \mid \mathcal{F}_t\right] = e^{\varphi^{\text{tr}}_{\text{q}}(T-t) + \psi^{\text{tr}}_{\text{q}}(T-t) X^1_t + \phi^{\text{tr}}_{\text{q}}(T-t)(X^1_t)^2} \tag{21}$$

where

$$\begin{aligned}\varphi^{\text{tr}}_{\text{q}}(t) &= \int_0^t \left(\alpha_1 \psi^{\text{tr}}_{\text{q}}(s)^2 + 2\alpha_1 \phi^{\text{tr}}_{\text{q}}(s) + b_1 \psi^{\text{tr}}_{\text{q}}(s)\right) ds, \\ \psi^{\text{tr}}_{\text{q}}(t) &= \frac{L_1(t)}{L_2(t)}, \quad \phi^{\text{tr}}_{\text{q}}(t) = \frac{L_3(t)}{L_2(t)},\end{aligned}$$

and

$$\begin{aligned}L_1(t) &:= \frac{\beta_1}{2\alpha_1}\left(2\rho_2 e^{\rho_2 t/2} - L_2(t)\right) \\ &\quad - \left(\frac{\beta_1 b_1}{2\alpha_1}\right)\frac{4}{\rho_2}\left(e^{\rho_2 t/2} - 1\right)\left(\beta_1\left(1 - e^{\rho_2 t/2}\right) + \frac{\rho_2}{2}\left(1 + e^{\rho_2 t/2}\right)\right), \\ L_2(t) &:= \rho_2\left(e^{\rho_2 t} + 1\right) - 2\beta_1\left(e^{\rho_2 t} - 1\right), \\ L_3(t) &:= 2\left(1 - e^{\rho_2 t}\right)\end{aligned}$$

with $\rho_2 := 2\sqrt{(\beta_1)^2 + 4\alpha_1}$. The corporate bond price with zero-recovery assumption can be obtained as

$$\begin{aligned}P^{\text{co}}_{\text{q}}(t,T) &= \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_t^T r_s\, ds} 1_{\{T < T_{\mathcal{D}}\}} \mid \mathcal{F}_t\right] \\ &= e^{\varphi^{\text{co}}_{\text{q}}(T-t) + \langle \Psi^{\text{co}}_{\text{q}}(T-t) X_t\rangle + \langle \Phi^{\text{co}}_{\text{q}}(T-t) X_t, X_t\rangle} 1_{\{T_{\mathcal{D}} > t\}}\end{aligned} \tag{22}$$

where $\varphi^{\text{co}}_{\text{q}}(t) \in \mathbb{R}$, $\Psi^{\text{co}}_{\text{q}}(t) \in \mathbb{R}^2$ and $\Phi^{\text{co}}_{\text{q}}(t) \in \mathbb{R}^{2\times 2}$ solve the Riccati equations:

$$\begin{aligned}\partial_t \varphi^{\text{co}}_{\text{q}} &= \langle \alpha \Psi^{\text{co}}_{\text{q}}, \Psi^{\text{co}}_{\text{q}}\rangle + 2tr(\alpha \Phi^{\text{co}}_{\text{q}}) + \langle b, \Psi^{\text{co}}_{\text{q}}\rangle - c, \quad \varphi^{\text{co}}_{\text{q}}(0) = 0, \\ \partial_t \Psi^{\text{co}}_{\text{q}} &= 4\alpha \Phi^{\text{co}}_{\text{q}} \Psi^{\text{co}}_{\text{q}} + (\beta)' \Psi^{\text{co}}_{\text{q}} + 2\Phi^{\text{co}}_{\text{q}} b, \quad \Psi^{\text{co}}_{\text{q}}(0) = 0, \\ \partial_t \Phi^{\text{co}}_{\text{q}} &= 4\Phi^{\text{co}}_{\text{q}} \alpha \Phi^{\text{co}}_{\text{q}} + \Phi^{\text{co}}_{\text{q}} \tilde{\beta} + (\tilde{\beta})' \Phi^{\text{co}}_{\text{q}} - \gamma_q, \quad \Phi^{\text{co}}_{\text{q}}(0) = 0,\end{aligned} \tag{23}$$

where

$$b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_1 & 0 \\ \beta_{21} & \beta_{22} \end{pmatrix}, \quad \gamma_q = \begin{pmatrix} 1+\gamma_1 & 0 \\ 0 & \gamma_2 \end{pmatrix}.$$

Given (58) and (22), the corresponding treasury yields and corporate bond yields are derived as

$$D_{\mathrm{q}}^{\mathrm{tr}}(t,T,X_t^1) = -\frac{1}{T-t}[\varphi_{\mathrm{q}}^{\mathrm{tr}}(T-t) + \psi_{\mathrm{q}}^{\mathrm{tr}}(T-t)X_t^1 + \phi_{\mathrm{q}}^{\mathrm{tr}}(T-t)(X_t^1)^2], \tag{24}$$

$$D_{\mathrm{q}}^{\mathrm{co}}(t,T,X_t) = -\frac{1}{T-t}[\varphi_{\mathrm{q}}^{\mathrm{co}}(T-t) + \langle \Psi_{\mathrm{q}}^{\mathrm{co}}(T-t)X_t\rangle + \langle \Phi_{\mathrm{q}}^{\mathrm{co}}(T-t)X_t, X_t\rangle], \tag{25}$$

$$S_q(t,T,X_t) = D_{\mathrm{q}}^{\mathrm{co}}(t,T,X_t) - D_{\mathrm{q}}^{\mathrm{tr}}(t,T,X_t^1). \tag{26}$$

Remark 2.7. *Since the potential approach for the quadratic model is equivalent to the intensity based models, the state process $X = (X^1, X^2)$ is equivalent to a simple two-factor Itô process before the default time, as given by*

$$dX_t = (\tilde{b} + \tilde{\beta}X_t)dt + \sqrt{2\alpha}dW_t, \tag{27}$$

where

$$\sqrt{2\alpha} = \begin{pmatrix} \sqrt{2\alpha_1} & 0 \\ 0 & \sqrt{2\alpha_2} \end{pmatrix},$$

and W_t is a standard 2-dimensional $\mathbb{P}$-Brownian motion.

It is worth noting that the parameters in the models are not fully identifiable, since the state variable X^2 (or Y^2 in the affine model) is subject to arbitrary scale. Without loss of generality, in what follows, we will fix γ_2 to be equal to 1. Then the remaining parameters are identifiable.

3 The Nonlinear Filtering Technique and Quasi Maximum Likelihood Estimator

3.1 Nonlinear Filtering

For affine models, the linear Kalman filter has been widely used for calibrating the joint specification of the model under both real-world measure $\mathbb{P}$ and risk-neutral measure $\mathbb{Q}$. (See e.g., Chen and Scott (1995), Duan and Simonato (1999) for its applications in risk-free affine term structure models, and Duffee (1999), Driessen (2002) for those in credit risk models). Therefore we will only illustrate the nonlinear filtering approach to estimating the quadratic model in this section. A general timehomogenous state-space model with the discrete observations $\{D_{t_i}\}$ at an increasing sequence of times $\{t_i\}_{0\leq i\leq n}$ is given by

$$\begin{aligned} dX_t &= F(X_t,\theta)dt + G(X_t,\theta)dW_t && (28)\\ D_{t_i} &= H(X_{t_i},\theta) + n_i(\theta), \quad \text{for every} \quad t_i \geq 0, && (29) \end{aligned}$$

where $H : \mathbb{R}^N \times \Theta \mapsto \mathbb{R}^m$, is an m-dimensional nonlinear function of the state vector and the parameter vector θ, which is assumed to lie in the compact parameter space Θ. $\{n_i(\theta)\}$ is an independently and identically distributed (i.i.d.) Gaussian noise sequence with means zero and covariance matrices $\{Q_i(\theta)\}$. Here we assume that $Q_i(\theta) = \varepsilon^2 I_m$ for all i, and that $t_0 = 0$.

According to Remark 2.7, for the specific case of the quadratic model proposed in the previous section, we have

$$\begin{aligned} F(X_t,\theta) &= \tilde{b} + \tilde{\beta}X_t, && (30)\\ \text{and} \quad G(X_t,\theta) &= \sqrt{2\alpha}, && (31) \end{aligned}$$

where $\theta := (\alpha, b, \tilde{b}, \beta, \tilde{\beta}, c, \tilde{c}, \gamma, \tilde{\gamma})$ and we set

$$\begin{aligned} H^k(X_t,\theta) &= D_{\mathrm{q}}^{\mathrm{tr}}(t, T_k, X_t^1), \quad \forall\ k = \{1, 2, ..., d\}, && (32)\\ H^k(X_t,\theta) &= D_{\mathrm{q}}^{\mathrm{co}}(t, T_k, X_t), \quad \forall\ k = \{d+1, d+2, ..., m\}, && (33) \end{aligned}$$

where $H^k(X_t,\theta)$ represents the k-th component of the function $H(X_t,\theta)$. In this case, $\{D_{t_i}\}$ is a set of both risk-free and corporate bond yields with different time to maturities $\{T_1, ..., T_m\}$, which can be observed from the market.

3.1.1 Time Propagation

Given the above model, for each i, the conditional density of X_t given $X_{t_{i-1}}$ and θ evolves according to the Kolmogorov forward equation:

$$\begin{aligned} \frac{\partial p(X_t|X_{t_{i-1}},\theta))}{\partial t} &= -\sum_{k=1}^{N} \frac{\partial}{\partial x_t^k}\{p(X_t|X_{t_{i-1},\theta})F_k(X_t,\theta)\}\\ +\frac{1}{2}\sum_{j=1}^{N}\sum_{k=1}^{N} &\frac{\partial^2}{\partial x_t^j \partial x_t^k}\{p(X_t|X_{t_{i-1},\theta})[G(X_t,\theta)G^{'}(X_t,\theta)]_{jk}\}, \end{aligned} \qquad (34)$$

for $t \in [t_{i-1}, t_i]$ and with the initial condition

$$p(X|X_{t_{i-1}},\theta) = \delta(X - X_{t_{i-1}}) \qquad (35)$$

where $\delta(\cdot)$ denotes the Dirac measure. This equation describes the propagation of the conditional density through the inter-sample interval $[t_{i-1}, t_i]$.

Since the computation of the entire density function is infeasible, we resort instead to the conditional mean as an estimator of the state vector. In particular, for $i = 1, 2, ..., n$, let $\hat{X}_{t|t_i}$ denote the estimator of X_t conditioned on $\{D_{t_j}\}_{0 \le j \le i}$ and $\hat{\Sigma}_{t|t_i}$ denote the covariance matrix of X_t conditioned on $\{D_{t_j}\}_{0 \le j \le i}$. Then, the propagation of the conditional mean and covariance over $[t_{i-1}, t_i]$ can also be derived from (34) (see, e.g., Maybeck (1982), Vol 2):

$$\frac{d\hat{X}_{t|t_{i-1}}}{dt} = \widehat{F(X_t, \theta)} \tag{36}$$

and

$$\begin{aligned} \frac{d\hat{\Sigma}_{t|t_{i-1}}}{dt} &= \{\widehat{F(X_t, \theta)X'}(t) - \widehat{F(X_t, \theta)}\hat{X}'_{t|t_{i-1}}\} \\ &\quad + \{X_t\widehat{F'(X_t}, \theta) - \hat{X}_{t|t_{i-1}}\widehat{F'(X_t, \theta)}\} \\ &\quad + G(X_t, \widehat{\theta)G(}X_t, \theta)' \end{aligned} \tag{37}$$

where $\vec{D}_{t_i} = \{D_{t_k}\}_{0 \le k \le i}$, and a carat over a quantity denotes conditional expectation of that quantity given $\vec{D}_{t_i}$; i.e.,

$$\widehat{\cdot} = \mathbb{E}^{\mathbb{P}}\left\{\cdot|\vec{D}_{t_i}\right\} \tag{38}$$

By substituting (30) and (31) into (36) and (37), we have the following proposition.

Proposition 3.1. *Given the quadratic model of* (17)*, we have the following time-updates:*

$$\hat{X}_{t_i|t_{i-1}} = U\Lambda(\Delta_i)U^{-1}\hat{X}_{t_{i-1}|t_{i-1}} + UV^{-1}(\Lambda(\Delta_i) - I_2)U^{-1}\tilde{b} \tag{39}$$

$$\begin{aligned} \textit{and} \quad \hat{\Sigma}_{t_i|t_{i-1}} &= U(\Lambda(\Delta_i)U^{-1}\hat{\Sigma}_{t_{i-1}|t_{i-1}}(U\Lambda(\Delta_i)U^{-1})' \\ &\quad + U\left[\frac{\nu_{ij}\exp((v_i + v_j)\Delta_i) - 1}{v_i + v_j}\right]_{NN} U' \\ \textit{for} \quad & 0 \le t_i \le t_n \end{aligned} \tag{40}$$

where $\Delta_i = |t_i - t_{i-1}|$, $U := (u_1, ..., u_N) \in \mathbb{R}^{N \times N}$, $V := diag[v_i]_N$, $\Lambda(t) := diag[\exp(v_i t)]_N$, *such that*

$$U^{-1}\tilde{\beta}U = V,$$

and

$$\mathcal{V} := [\nu_{ij}]_{NN} = 2U^{-1}\alpha(U^{-1})'.$$

3.1.2 Measurement Updating

As to the measurement update, the conditional mean and covariance are given by the following equations:

$$\hat{X}_{t_i|t_i} = \int X p(X|\vec{D}_{t_i}) dX \tag{41}$$

$$\text{and} \quad \hat{\Sigma}_{t_i|t_i} = \int X X' p(X|\vec{D}_{t_i}) dX - \hat{X}_{t_i|t_i} \hat{X}'_{t_i|t_i} \tag{42}$$

According to Bayes' formula, we have

$$p(X_{t_i}|\vec{D}_{t_i}) = \frac{p(D_{t_i}|X_{t_i}, \vec{D}_{t_{i-1}}) p(X_{t_i}|\vec{D}_{t_{i-1}})}{p(D_{t_i}|\vec{D}_{t_{i-1}})} \tag{43}$$

$$= \frac{p(D_{t_i}|X_{t_i}) p(X_{t_i}|\vec{D}_{t_{i-1}})}{p(D_{t_i}|\vec{D}_{t_{i-1}})} \tag{44}$$

and

$$p(D_{t_i}|\vec{D}_{t_{i-1}}) = \int p(D_{t_i}|X_{t_i}) p(X_{t_i}|\vec{D}_{t_{i-1}}) dX_{t_i}, \tag{45}$$

so that

$$\hat{X}_{t_i|t_i} = X_{t_i} \widehat{p(\vec{D}_{t_i}}|X_{t_i}) / p(\widehat{\vec{D}_{t_i}|X}_{t_i}) \tag{46}$$

$$\text{and} \quad \hat{\Sigma}_{t_i|t_i} = X X' \widehat{p(\vec{D}_{t_i}}|X_{t_i}) / p(\widehat{\vec{D}_{t_i}|X}_{t_i}) - \hat{X}_{t_i|t_i} \hat{X}'_{t_i|t_i} \tag{47}$$

Maybeck (1982) pointed out that approximating the updating expectation by a series expansion of $p(X|\vec{D}_{t_i})$ would incur a considerable measurement error. He assumes, instead, that the conditional mean and covariance can be expressed as power series of the innovations $\{D_{t_i} - E\{D_{t_i}|\vec{D}_{t_{i-1}}\}\}$, and uses a linear approximation, since the innovations are relatively small.

Here we restate the final results for the updating step according to Maybeck (1982):

$$\hat{X}_{t_i|t_i} = \hat{X}_{t_i|t_{i-1}} + K_{t_i}[D_{t_i} - H(\hat{X}_{t_i|t_{i-1}}, \theta) - \gamma_{t_i|t_{i-1}}(\theta)] \tag{48}$$

$$\text{and} \quad \hat{\Sigma}_{t_i|t_i} = \hat{\Sigma}_{t_i|t_{i-1}} - K_{t_i} h(\hat{X}_{t_i|t_{i-1}}, \theta) \hat{\Sigma}_{t_i|t_{i-1}}, \tag{49}$$

where

$$K_{t_i} = \hat{\Sigma}_{t_i|t_{i-1}} h(\hat{X}_{t_i|t_{i-1}}, \theta)' A_{t_i}^{-1} \tag{50}$$

with

$$\begin{aligned}
A_{t_i} &= h(\hat{X}_{t_i|t_{i-1}},\theta)\hat{\Sigma}_{t_i|t_{i-1}}h(\hat{X}_{t_i|t_{i-1}},\theta)' \\
&\quad -\gamma_{t_i|t_{i-1}}(\theta)\gamma_{t_i|t_{i-1}}(\theta)' + Q_i(\theta) \qquad (51)\\
h(\hat{X}_{t_i|t_{i-1}},\theta) &= \frac{\partial H(\hat{X}_{t_i|t_{i-1}},\theta)}{\partial X} \qquad (52)\\
\text{and} \quad \gamma^k_{t_i|t_{i-1}}(\theta) &= \frac{1}{2}tr\left[\frac{\partial^2 H^k(\hat{X}_{t_i|t_{i-1}},\theta)}{\partial X^2}\hat{\Sigma}_{t_i|t_{i-1}}\right] \qquad (53)\\
\text{for} \quad & 1 \le k \le m.
\end{aligned}$$

3.2 Quasi-Maximum Likelihood Estimator(QMLE)

Given θ, using the above filtering technique, we can calculate the conditional mean estimator of the state vector $\{\hat{X}_{t_i|t_i}\}_{0\le i\le n}$. In order to estimate the parameter vector θ, we will use a quasi-maximum likelihood estimator. The asymptotic properties of this estimator are given in the appendix. Let $\{\hat{n}_i(\theta)\}$ denote the one-step prediction error defined by

$$\hat{n}_i(\theta) = D_{t_i} - H(\hat{X}_{t_i|t_{i-1}},\theta), \qquad (54)$$

and let $l_{\hat{n}_i}(\hat{n}_i,\theta)$ denote the true log-likelihood function for $\hat{n}_i(\theta)$. Since it is quite difficult to obtain an analytical expression for $l_{\hat{n}_i}(\hat{n}_i,\theta)$, we instead give a quasi log-likelihood function $\hat{l}_{\hat{n}_i}(\hat{n}_i,\theta)$ as follows:

$$\begin{aligned}
\log p(\{\hat{n}_i(\theta)\}) &= \sum_{i=1}^{n}\hat{l}_{\hat{n}_i}(\hat{n}_i,\theta)\\
&= \sum_{i=1}^{n} -\frac{1}{2}(\log|M_i(\theta)| + (\hat{n}_i(\theta)-\gamma_{t_i|t_{i-1}}(\theta))'M_i(\theta)^{-1}\\
&\quad (\hat{n}_i(\theta)-\gamma_{t_i|t_{i-1}}(\theta)) + m\log(2\pi)) \qquad (55)
\end{aligned}$$

where $\gamma_{t_i|t_{i-1}}(\theta)$ is given by (53) and

$$M_i(\theta) = h(\hat{X}_{t_i|t_{i-1}},\theta)\hat{\Sigma}_{t_i|t_{i-1}}h(\hat{X}_{t_i|t_{i-1}},\theta)' + Q_i(\theta). \qquad (56)$$

Here we are essentially assuming that the $\{\hat{n}_i(\theta)\}$ are mutually independent Gaussian random variables with means $\{\gamma_{t_i|t_{i-1}}(\theta)\}$ and covariances $\{M_i(\theta)\}$. The rationale for this assumption will be discussed in the following section.

Since the Jacobian transfer matrix is given by

$$\frac{\partial(\hat{n}_1,\hat{n}_2,...,\hat{n}_n)}{\partial(D_{t_1},D_{t_2},...,D_{t_n})} = I_n \qquad (57)$$

we have

$$\begin{aligned}\log p(D_{t_1}, D_{t_2}, ..., D_{t_n}; \theta) &= \sum_{i=1}^{n} \log q_i(\vec{D}_{t_i}, \theta) \\ &= \sum_{i=1}^{n} l_{\hat{n}_i}(D_{t_i} - H(\hat{X}_{t_i|t_{i-1}}, \theta), \theta) \\ &\cong \sum_{i=1}^{n} \hat{l}_{\hat{n}_i}(D_{t_i} - H(\hat{X}_{t_i|t_{i-1}}, \theta), \theta) \end{aligned} \tag{58}$$

where

$$q_i(\vec{D}_{t_i}, \theta) = \begin{cases} p(D_{t_i}|\vec{D}_{t_{i-1}}, \theta), & \text{for} \quad i > 1 \\ p(D_{t_1}, \theta), & \text{for} \quad i = 1 \end{cases} \tag{59}$$

After filtering $\{\hat{X}_{t_i|t_{i-1}}\}_{0\leq i\leq n}$, by using this quasi log-likelihood function, we can obtain the QMLE from the nonlinear optimization:

$$\hat{\theta}_n = \arg\max_{\theta} \left\{ \frac{1}{n} \sum_{i=1}^{n} \hat{l}_{\hat{n}_i}(D_{t_i} - H(\hat{X}_{t_i|t_{i-1}}, \theta), \theta) \right\} \tag{60}$$

For implementing the QMLE, since $Q_i(\theta) = \varepsilon^2 I_m$, by applying Woodbury's formula, M_i^{-1} can be simplified as

$$M_i^{-1} = \varepsilon^{-2}[I_m - h_i(\varepsilon^2 \hat{\Sigma}_{t_i|t_{i-1}}^{-1} + h_i h_i^{'})^{-1} h_i^{'}]. \tag{61}$$

Similarly, we can simplify $|M_i|$ as

$$|M_i| = \varepsilon^{2(m-N)} |\hat{\Sigma}_{t_i|t_{i-1}}| |\varepsilon^2 \hat{\Sigma}_{t_i|t_{i-1}}^{-1} + h(t_i, \hat{X}_{t_i|t_{i-1}}, \theta)^{'} h(t_i, \hat{X}_{t_i|t_{i-1}}, \theta)|. \tag{62}$$

4 Estimation Methodology and Data Description

4.1 Data

As we need to test two different aspects of the empirical properties of the models: time-series property and cross-sectional fitting ability, two different datasets are constructed and used. In order to compare the performance of the affine diffusion (AD) model and quadratic (QD) model in explaining the historical movements of treasury yields and credit spreads, we apply the time series of monthly observations on treasury yields with maturity 3 month, 1 year, 10 year and 30 year, and the average 1-year credit spreads from four different Moody's rating classes: Aaa, Aa, A, Baa. The data are collected from the Global Insight (formerly DRI) and provided by Wharton Research Data Services. The sampling period ranges from January 1978 through December 2002 with totally 300 samples from each different type of the time series. Since the corporate bond yields are subject to a large liquidity effect, by choosing the 1-year average corporate bond spread data to estimate the two models' structural parameters instead of using the individual bond prices, we can avoid the possible measurement errors caused by the stale price problems in the individual bond price data. To estimate the default event risk premium γ, we use the dataset of historical default records for four different rating categories given by Moody's special comment (2002), which contains the cumulative default rates from 1970-2001.

When examining the term structure fitting abilities of the affine diffusion (AD) model, the affine fat-tail (AF) jump model and the quadratic model (QD), we collect both treasury and corporate bond data from www.Bondpage.com. This dataset includes a snapshot of 50 observations of treasury note and bond prices and more than 600 month-end quoted prices of corporate bonds issued by the investment-grade firms. All the bonds are non-callable, non-convertible, and without sinking fund provisions. They all have at least one year remaining to maturity and share the same settlement date.

The summary statistics of all sample data are shown in Table 1- Table 2.

[Table 1 about here.]

[Table 2 about here.]

4.2 Estimation Methodology and Two-Step Optimization Algorithm

As described above, filtering technique together with QMLE is applied to estimating the model using the 25-year time-series of monthly bond yields. Different from many existing literature in empirical estimations of credit risk, where the risk-free parameters and defaultable parameters are estimated separately (e.g., see Driessen (2002), and Duffee (1999)), we jointly estimate the risk-free parameters and defaultable parameters instead of separately, which would be more efficient since the corporate bond price depends on both the risk-free rate and the credit index.

Meanwhile, a nonlinear least squares algorithm is applied to estimating the parameters to fit the term structure of both treasury rates and the corporate bond spreads using a snapshot of the market data. the corresponding objective function can be written as

$$F(\theta) = \sum_{i=1}^{N} (p_i - P_i(\theta, \vec{T}_i, c_i))^2,$$

where p_i denotes the observed price of bond i, and $P(\theta, \vec{T}_i, c_i)$ denotes the model-implied bond price with the parameter set (θ), the semiannual coupon rate c_i and coupon payment dates $\vec{T}_i = (T_{i,1}, T_{i,2}, ..., T_{i,m_i})$. Here without loss of generality, zero-recovery at default is assumed when calculating corporate bond prices (see, Remark 2.6). Therefore the estimator is given by

$$\theta^* = \arg\min_{\theta} \{F(\theta)\}.$$

However, two problems come up. First, although the data includes 600 non-callable corporate bond prices, no individual firm has more than 10 observations. Hence the credit index estimation for each individual firm is subject to substantial uncertainty. Duffee (1999) encountered the similar problems when estimating the default intensity of each firm. A way to overcome this problem is to form four rating groups, Aaa, Aa, A, Baa, and estimate a typical credit index value, Y^2_{Aaa}, Y^2_{Aa}, Y^2_{A}, Y^2_{Baa}, for each of these groups, respectively.

The second problem is in estimating the AF model with the difficulty of estimating the jump parameter ζ. The parameter ζ turns out to be dominant over the other parameters. Changing the value ζ results in significant value changes of other parameters, but the differences between measurement errors are rather small, which implies that estimating the parameter ζ by minimizing the mean square error is infeasible. Therefore, instead we fix both ζ_i, $(i = 1, 2)$ at 1.1, when implementing the optimization algorithm.

Finally, both empirical tests finally lead to an optimization problem. With regard to a fair large parameter set θ, the traditional gradient search methods are susceptible to getting 'stuck' at local optima. Here we apply a genetic algorithm (GA) for optimization that uses a set, or population, of points to conduct a search, not just a single point in the parameter space. This gives us the power to search noisy spaces littered with local optimum points. Instead of relying on a single point to search through the space, the GA looks at many different areas of the parameter space at once, and uses all of this information to guide it. Four operations are applied to every generation of the search: evaluation, selection, crossover, and mutation. This makes the GA especially helpful in finding a global optimal solution. Although the search is not precise meaning that there is no guarantee that the global maximum will be found, the result should still be a good approximation of the optimum. In order to refine the optimal solution, a two-step optimization method is employed. First, we apply the genetic algorithm to search for a global optimal solution[6]. Then we use the gradient search method to improve the accuracy of the optimization by starting from this solution. Given the explicit formulas for bond prices, this optimization method becomes quite fast and efficient.

5 Estimation Results

5.1 Time Series Properties of the Affine and Quadratic Models

The estimation results and the summary statistics of the fitting errors are displayed in Table 3 and Table 4.

The risk-free parameters of both affine and quadratic models are estimated using the 25-year time series of monthly treasury yields with maturity 3-month, 1-year, 10-year and 30-year. Under both the real world and risk-neutral measures, the short rate follows a mean-reverting process, while the real world dynamics shows a much larger mean-reverting rate (i.e., $|\tilde{\beta}_{11}| > |\beta_{11}|$), which implies the existence of a positive market risk premium for Treasury bonds. It is worth noting that we also provide the estimates of the risk compensation on the mean level of the interest rate. This risk premium is independent of the interest rate volatility, which substantially improves the model's performance in forecasting future yields as demonstrated in Duffee (2002). As shown in Table 3 and Table 4, the difference between $\tilde{b}_1$ and b_1 are significantly positive as 0.03249 on average for the affine model and at average 0.00977 for the quadratic model, which constitutes an extra excess return for Treasury bonds. The results also indicate that one-factor affine and quadratic short rate model are incapable to price the short-term bonds accurately with regard to the relative large root mean square errors (RMSE): the RMSE for 3-month bills is 48.342 b.p. for the affine model and 30.353 b.p. for the quadratic model. However, this performance of both models becomes reasonably well at pricing the long-term instruments in view of the RMSE for fitting 30-year Treasury yields is only around 15 b.p. for the affine model and 12 b.p. for the quadratic model. It is also noticed that the fitted yield curves are consistently higher than actual yields (with the mean fitting error for 3-month bill at -28.482 b.p. (-12.132%) for the affine model and -22.372 b.p. (-5.132%) for the quadratic model at the short end and are underpriced for long-maturity bonds. The similar error patterns are also documented in Duffee (1999), which was taken as the evidence of the models' misspecification. Finally, in terms of the pricing performance for treasury bonds, the quadratic model generally outperforms the affine model in that it might be able to capture the nonlinearity of the relevant time series.

Table 3 and Table 4 also summarize the defaultable parameter estimates for the affine and quadratic model, respectively. For both the affine model and the quadratic model, the second state variable exhibits a strong mean reversion trend under the real world measure, where the mean reversion rates $|\beta_{22}|$ are around 5 for both the affine model and the quadratic model. Our empirical results also imply a mean reversion dynamics for the second state variable under the risk-neutral measure, although rather weak compared with the one under the real world measure with the half life around 1.08 year for the affine model and and almost 4-year for the quadratic model. However, in contrast, Driessen (2002) and Duffee (1999) documented a weak mean-averting risk-neutral dynamics of the credit-related factor for most individual corporate bonds. The difference may come from the different data source used in the test. Since instead of using the yield data of individual corporate bonds in Driessen (2002), we use the average bond yields for four rating classes, which generate a more stable dynamics of the credit-related variable. Some of the parameter estimates show a systematic tendency across the four rating classes. For both models, the mean level of the parameter b_1 increases as the credit rating decreases, which is consistent with our assumption that the higher the credit index value, the worse the credit quality. For the quadratic model, the diffusion parameter α_2 also increases as the credit rating goes down, which implies a higher volatility for the lower rated bonds. For the affine model, since the volatility also depends on the state variable Y^2, the estimates of α_2 are similar for all rating classes. Finally, the results also show that both models can explain quite well the historical credit spread dynamics with regard to less than 15 b.p. RSME for fitting errors of all four rating classes. However, the quadratic model once again shows a relatively better empirical pricing performance than the affine model.

[Table 3 about here.]

[Table 4 about here.]

Figure 1 through Figure 4 provide the comparison between the fitted yields $\{H(\hat{X}_{t_i|t_{i-1}}, \theta)\}$ and actual yields $\{D_{t_i}\}$ observed from the market. The plots suggest that the quadratic model fares better in capturing the first conditional moment of yield changes for both Treasury bonds and corporate bonds even for the high-rate regime 1979-1982. Moreover the plots also show that both the affine and quadratic models fit the long-term bond dynamics better than short term.

[Figure 1 about here.]

[Figure 2 about here.]

[Figure 3 about here.]

[Figure 4 about here.]

5.2 Jump Risk Premium and Non-Default Factors

In the previous part, the default intensity under risk-neutral measure has been estimated for both affine and quadratic models. Given Moody's historic default rates, we can perform the estimation for the physical default intensity for the two models. Actually in Driessen (2002), the author estimated the default risk premium when considering a simple relationship of the intensities under the two measures; that is

$$\frac{h_t^{\mathbb{P}}}{h_t^{\mathbb{Q}}} = \mu,$$

which means that the ratio between the intensities are constant over time. However, as argued in Jarrow, Lando and Turnbull (2001) and Yu (2002), simply taking this constant μ as the default jump risk premium is doubtful. First, the constant ratio would imply a simple dependent structure between the term structure of credit spreads and default rates which is inconsistent with the actual market data. Moreover, as demonstrated in Delianedis and Geske (2001), and Huang and Huang (2003), credit risk accounts for only a small fraction of the observed corporate-Treasury yield spreads for investment grade bonds of all maturities, while a major portion of the spread is attributable to some non-Default factors such as taxes, liquidity and market risk factors. In order to refine the default risk factor, Jarrow et al. (2001) propose to subtract the average value of the Aaa-implied intensity from the intensities implied from other ratings. Yu (2002) further suggests to subtract the non-default factor from the default intensity, although a fully specified model of non-default sources of the spread remains elusive.

Since our models already assume that the zero-value of the credit index implies the perfect financial health of a firm, we naturally conjecture that the default risk part of the corporate bond spread is given by

$$S_d(t, T, X_t) = S(t, T, X_t^1, X_t^2) - S(t, T, X_t^1, 0),$$

where S is defined in (13) for the affine model and (26) for the quadratic model, which means the spread related to non-default factors is given by these formulas when setting the credit index Z to be zero. Figure 5 compares the corporate bond spreads and their non-default part for the affine model and similar result holds for the quadratic model. It is interesting to see that for the Aaa and Aa rated bonds, the model implied spreads are almost solely contributed by the non-default part, which is consistent with the findings in the empirical literature (e.g., see Huang and Huang (2003)).

[Figure 5 about here.]

As to determining the intensity structure under the real world, we use annually default rate data given by Moody's. The approach is same as the one applied in Driessen (2002), except that our real world default intensity structure is more flexible with three parameters $\tilde{c}$, $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$ to estimate. For the detail of the methodology, we refer to Driessen (2002). The results are summarized in Table 5 and Figure 6-8 display the default probability structures for Aa, A and Baa rated bonds, where 'Q-Prob' denotes the risk-neutral default probability, 'Unadjusted P-Prob' denotes the physical default probability without adjusting the default intensities, and 'Adjusted P-Prob' represents the one using our estimates of the real world default intensities.

[Table 5 about here.]

[Figure 6 about here.]

We find our estimate results strongly reject the constant ratio assumption between the two intensities in Driessen (2002). As shown in Table 5, constant default intensity $\tilde{c}$ and the intensity parameter $\tilde{\gamma}_1$ influenced by the market decreases significantly compared the ones under the risk-neutral measure. However, the firm-specific default factor $\tilde{\gamma}_2$ only has a small change. This result implies the physical default risk is relatively more influenced by the firm-specific factors than other factors with respect to the risk-neutral situation.

6 Fitting the Term Structure of Credit Spreads

In the previous section, we examine the time-series properties of the affine and quadratic models, which is relevant to their implications in long-term investments and risk management. However, for the short term trading and pricing purposes, it is more interesting to see the term structure fitting ability of the models. Here we investigate the cross-sectional properties of the three models: the affine diffusion (AD) model, the affine fat-tail (AF) model and the quadratic (QD) model using a snapshot of the market data including 50 Treasury bonds and more than 650 corporate bonds from the four Moody's rating classes. As described in Section 4.2, a simple nonlinear least square optimization is applied. In order to test the robustness of this nonlinear optimization method, thirty independent experiments are performed and the estimate results are summarized in Table 6.

[Table 6 about here.]

We can see that for all the three models, the credit index processes are mean reverting under the risk-neutral measure , which is consistent with the previous estimate results using historic yield data. The significant non-zero estimate values for β_{21} confirms the fact that the short rate does impact on the dynamics of credit indices. Now we examine the performance of our model fitting the yield curves. The averaged fitting error for short term corporate (treasury) bonds (with time to maturity less than four years) is around 15 b.p. (3 b.p.) for the AD and AF models and around 17 b.p. (3 b.p.) for the QD model as shown in Table **??**. These results indicate that there exists a severe short-term distortion between model implied yields and the actual yields, which might be caused by the innate defects of our one-fact short rate model and the illiquidity of short term corporate bonds. However, all the three models exhibit a good fitting ability for long term yields with less than 2 b.p. RMSE for Treasury bonds and less than 8 b.p. for corporate bonds. It is also noticed that among all the three models, the affine fat-tail model has the best goodness-of-fit performance with regard to its smallest mean fitting error for all different bonds.

Another way to speak for the quality of a model is to see whether the values for credit indices vary too much for firms within one rating class. Therefore by fixing all the remaining parameters given by the preceding estimation, it is an interesting test to inversely solve for the credit index Y^2 (or X^2) of every individual bond. The summary statistics of the implied credit indices are shown in Table 7 and their term structures are displayed in Figure 9 – Figure 11.

As we expected, the values of credit indices increase as the corresponding investment grades go down. Aaa rated bonds have the lowest (best) mean value (around 0.03 for the AD and AF models, and around 0.6 for the QD model), while Baa rated bonds have the highest (worst) mean values (around 0.55 for the AD model, 0.3 for the AF model, and 0.95 for the QD model).

[Table 7 about here.]

As illustrated in Figure 9 – Figure 11, the AD model has quite a few Aaa-rated bonds imply negative credit index values, which is not allowed in our affine setup. This means that the fixed yield spread part explained as tax and liquidity effects is too large and has to be compensated by subtracting the credit-sensitive part. However, for the QD model, there exists a significant upward drift of the credit index value from short-term bonds to long-term bonds. This means that $T \mapsto -\frac{1}{T}\tilde{\Phi}_q^{\mathrm{co}}(T)$ is too flat, resulting in an underestimate of long term credit spreads which has to be compensated by larger values of credit indices. Finally, we concluded the AF model clearly outperforms AD and QD in terms of their flat term structure and relatively low standard deviations across all different time to maturities.

[Figure 7 about here.]

A Asymptotic Properties of the QMLE

Since we apply the quasi log-likelihood function instead of the true one to implement the QMLE, we need to verify its validity. We now consider this issue.

First we give the following lemma.

Lemma A.1. *Given a sequence of random variables* $\{D_{t_i}\}_{i\geq 1}$*, suppose* $\{g_n(\vec{D}_{t_n},\theta)\}$ *and* $\{k_n(\vec{D}_{t_n},\theta)\}$ *are two sequences of measurable functions (*$\vec{D}_{t_n} = \{D_{t_i}\}_{1\leq i\leq n}$*) that satisfy the following conditions*

i)

$$\begin{aligned} g_n(\vec{D}_{t_n},\theta) &\rightarrow s(\theta) \quad i.p. & (63)\\ and \quad k_n(\vec{D}_{t_n},\theta) &\rightarrow s(\theta) \quad i.p., & (64)\\ as \qquad & n\rightarrow\infty. \end{aligned}$$

ii) $\{g_n(\vec{D}_{t_n},\theta)\}$*,* $\{k_n(\vec{D}_{t_n},\theta)\}$ *and* $s(\theta)$ *are identifiable and smooth on* Θ*. (Note: the appropriate definitions of "identifiable" and "smooth" can be found in Peracchi (2000))*

iii) The sequences $\{\hat{\theta}^g_n\}$ *and* $\{\hat{\theta}^k_n\}$ *defined as follows,*

$$\begin{aligned} \hat{\theta}^g_n &= \arg\max_{\theta\in\Theta}\{g_n(\vec{D}_{t_n},\theta)\}, n=1,2,... & (65)\\ and \quad \hat{\theta}^k_n &= \arg\max_{\theta\in\Theta}\{k_n(\vec{D}_{t_n},\theta)\}, n=1,2,..., & (66) \end{aligned}$$

are convergent.

Then

$$\lim_{n\rightarrow\infty}(\hat{\theta}^g_n - \hat{\theta}^k_n) = 0 \quad i.p. \qquad (67)$$

Proof: Since we have

$$p\lim_{n\rightarrow\infty} g_n(\vec{Y}_{t_n},\theta) = p\lim_{n\rightarrow\infty} k_n(\vec{Y}_{t_n},\theta) = s(\theta), \qquad (68)$$

where $p\lim$ means the limit in probability, then it is easy to prove that

$$|\max_\theta g_n(\vec{Y}_{t_n},\theta) - \max_\theta k_n(\vec{Y}_{t_n},\theta)| \rightarrow 0 \quad \text{i.p..} \qquad (69)$$

Using the definitions (65) and (66), we can rewrite (69) as

$$|g_n(\vec{Y}_{t_n},\hat{\theta}^g_n) - k_n(\vec{Y}_{t_n},\hat{\theta}^k_n)| \rightarrow 0 \quad \text{i.p.} \qquad (70)$$

Because

$$|s(\hat{\theta}_n^g)-s(\hat{\theta}_n^k)| \leq |s(\hat{\theta}_n^g)-g_n(\vec{Y}_{t_n},\hat{\theta}_n^g)|+|g_n(\vec{Y}_{t_n},\hat{\theta}_n^g)-k_n(\vec{Y}_{t_n},\hat{\theta}_n^k)|+|s(\hat{\theta}_n^k)-k_n(\vec{Y}_{t_n},\hat{\theta}_n^k)| \tag{71}$$

and all of the three terms on the right-side of the above equation converge to zero in probability, as n goes to infinity, we have

$$|s(\hat{\theta}_n^g) - s(\hat{\theta}_n^k)| \to 0 \quad \text{i.p.} \quad \text{as} \quad n \to \infty. \tag{72}$$

This is equivalent to

$$\hat{\theta}_n^g - \hat{\theta}_n^k \to 0 \quad \text{i.p.} \quad \text{as} \quad n \to \infty \tag{73}$$

if $s(\theta)$ is identifiable on Θ. □

A.1 Optimality under the Kullback-Leibler Criterion

The quantity $p(\vec{D_{t_n}}, \theta_0)$ (see (58)) is the true likelihood function of the observation vector $\vec{D_{t_n}} = (D_{t_1}, D_{t_2}, ..., D_{t_n})$. From (59), we can write

$$p(\vec{D_{t_n}}, \theta_0) = \prod_{i=1}^{n} q_i(\vec{D}_{t_i}, \theta_0). \tag{74}$$

A natural way to analyze the consistency of the QMLE is to evaluate the divergence between the true likelihood functions and the likelihood function used in computing the QMLE, as $n \to \infty$. Here we will apply the Kullback-Leibler (K-L) divergence defined for a probability density p_1 with respect to a probability density p_2 as

$$D_{KL}(p_1, p_2) = \int \log \frac{p_1(x)}{p_2(x)} p_1(x) dx. \tag{75}$$

In our setting, we define an estimator $\bar{\theta}_n$ via the Kullback-Leibler criterion (Kullback-Leibler, 1951) as

$$\bar{\theta}_n = \arg\min_{\theta} \left\{ D_{KL}^n[p(\vec{D_{t_n}}, \theta_0), \hat{p}(\vec{D_{t_n}}; \theta)] \right\}, \tag{76}$$

where

$$\log \hat{p}(\vec{D_{t_n}}; \theta) = \sum_{i=1}^{n} \hat{l}_{\hat{n}_i}(D_{t_i} - H(\hat{X}_{t_i|t_{i-1}}, \theta), \theta). \tag{77}$$

Following (76), we have

$$
\begin{aligned}
\bar{\theta}_n &= \arg\min_{\theta}\left\{\frac{1}{n}D_{KL}\left\{p(\vec{D}_{t_n},\theta_0),\hat{p}(\vec{D}_{t_n};\theta)\right\}\right\} \\
&= \arg\min_{\theta}\left\{\frac{1}{n}E\left\{\log\frac{\prod_{i=1}^n p_i(\vec{D}_{t_i},\theta_0)}{\prod_{i=1}^n \hat{p}_i(\vec{D}_{t_i},\theta)}\right\}\right\} \\
&= \arg\min_{\theta}\left\{\frac{1}{n}E\left\{\sum_{i=1}^n \log p_i(\vec{D}_{t_i},\theta_0) - \frac{1}{n}\sum_{i=1}^n \hat{l}_i(\vec{D}_{t_i},\theta)\right\}\right\} \\
&= \arg\max_{\theta}\left\{\frac{1}{n}E\left\{\sum_{i=1}^n \hat{l}_i(\vec{D}_{t_i},\theta)\right\}\right\} \\
&= \arg\max_{\theta}\left\{\frac{1}{n}\sum_{i=1}^n \bar{l}_i(\theta)\right\} \qquad (78) \\
\text{where} \quad \bar{l}_i(\theta) &= E\left\{\hat{l}_i(\vec{D}_{t_i},\theta)\right\} = \int \hat{l}_i(\vec{D}_{t_i},\theta)p(\vec{D}_{t_i})d\vec{D}_{t_i} \qquad (79) \\
\text{with} \quad \log\hat{q}_i(\vec{D}_{t_i},\theta) &= \hat{l}_i(\vec{D}_{t_i},\theta) = \hat{l}_{\hat{n}_i}(D_{t_i} - H(\hat{X}_{t_i|t_{i-1}},\theta),\theta). \qquad (80)
\end{aligned}
$$

The third step of the above deduction is true because $E\{\sum_{i=1}^n \log p_i(\vec{D}_{t_i},\theta_0)\}$ is unconditional on θ.

As shown in (Gallant and White, 1988), under regularity conditions, we have the generalized version of the uniform law of large numbers:

$$
\frac{1}{n}\sum_{i=1}^n \hat{l}_i(\vec{D}_{t_i},\theta) - \frac{1}{n}\sum_{i=1}^n \bar{l}_i(\theta) \to 0 \quad \text{a.s.} \qquad (81)
$$

Thus according to Lemma A.1, we have

$$
(\hat{\theta}_n \quad - \quad \bar{\theta}_n) \to 0 \quad \text{i.p.} \qquad (82)
$$

This means that the QMLE $\hat{\theta}_n$ will asymptotically minimize the K-L divergence between $p(\vec{D_{t_n}})$ and $\hat{p}(\vec{D_{t_n}})$; i.e., it is asymptotically optimal under the K-L criterion.

A.2 General Consistency and Asymptotical Normality

As mentioned in Bollerslev and Wooldridge (1992), since the score of the normal log-likelihood has the martingale difference property when the first two conditional moments are correctly specified, the Gaussian distributed QMLE is generally consistent and asymptotically normally distributed. In particular we have the following result from Bollerslev and Wooldridge (1992):

Lemma A.2. *Given a quasi log-likelihood function, if the first two conditional moments are correctly specified, under regularity conditions (for the sake of readability, these are listed in the appendix) we have*

$$A_n^{-1} B_n A_n^{-1} \sqrt{n}(\theta_n^* - \theta_0) \to \mathcal{N}(0, I) \tag{83}$$

where

$$\begin{aligned} A_n &= \frac{1}{n}\sum_{i=1}^{n} E\left\{ \left(\frac{\partial \mu_i(\theta_0)}{\partial \theta}\right)' \Omega_i^{-1}(\theta_0) \frac{\partial \mu_i(\theta_0)}{\partial \theta} \right. \\ &+ \left. \frac{1}{2}\frac{\partial \Omega_i(\theta_0)'}{\partial \theta}[\Omega_i^{-1}(\theta_0) \otimes \Omega_i^{-1}(\theta_0)] \frac{\partial \Omega_i(\theta_0)}{\partial \theta} \right\} \end{aligned} \tag{84}$$

and

$$B_n = \frac{1}{n}\sum_{i=1}^{n} E\left\{ \frac{\partial \hat{l}_i(\theta_0)'}{\partial \theta} \frac{\partial \hat{l}_i(\theta_0)}{\partial \theta} \right\} \tag{85}$$

where $\mu_i(\theta)$ and $\Omega_i(\theta_0)$ represent the conditional mean and covariance respectively, given the observations $\{D_k\}_{1 \le k \le i}$, $\hat{l}_i(\theta)$ is the quasi-likelihood function given the observations $\{D_k\}_{1 \le k \le i}$, and $\otimes$ represents the Kronecker product.

The validity of the assumed quasi-likelihood function (defined in (55)) is shown by the following derivation. According to (28) and (54), we have

$$\hat{n}_i(\theta) = n_i(\theta) + [H(X_{t_i}, \theta) - H(\hat{X}_{t_i|t_{i-1}}, \theta)]. \tag{86}$$

Because, in our case, H is a quadratic function of the state vector, we can rewrite the above equation as

$$\begin{aligned} \hat{n}_i^k(\theta) &= n_i^k(\theta) + h^k(\hat{X}_{t_i|t_{i-1}}, \theta)'(X_{t_i} - \hat{X}_{t_i|t_{i-1}}) \\ &+ \frac{1}{2}(X_{t_i} - \hat{X}_{t_i|t_{i-1}})' h_2^k(\hat{X}_{t_i|t_{i-1}}, \theta)(X_{t_i} - \hat{X}_{t_i|t_{i-1}}) \\ &\qquad \text{for} \quad 1 \le k \le m. \end{aligned} \tag{87}$$

where

$$h_2^k(X, \theta) = \frac{\partial^2 H^k(X, \theta)}{\partial X^2}, \quad \text{for} \quad 1 \le k \le m.$$

Now we can prove that $\{\hat{n}_i(\theta)\}$ are mutually independent random variables that satisfy:

$$
\begin{aligned}
E\{\hat{n}_i(\theta)\} &= \gamma_{t_i|t_{i-1}}(\theta) && (88)\\
\text{and} \quad Var(\hat{n}_i(\theta)) &= M_i(\theta) + N_i(\theta) && (89)
\end{aligned}
$$

where

$$
N_i(\theta) = E\{(\hat{n}_i(\theta) - n_i(\theta))(\hat{n}_i(\theta) - n_i(\theta))'\} - \{\gamma_{t_i|t_{i-1}}(\theta)\gamma_{t_i|t_{i-1}}(\theta)'\}. \qquad (90)
$$

Define $\{l'_{\hat{n}_i}(\vec{D}_{t_i}, \theta)\}$ to be a sequence of Gaussian log-likelihood functions with means and variances defined in (88) and (89). Thus, according to Lemma 2, if we apply $\sum_{i=1}^{n} l'_{\hat{n}_i}(\vec{D}_{t_i}, \theta)$ to the QMLE, the obtained estimator $\hat{\theta}_n$ is generally consistent and asymptotical normally distributed. Here we assume that $\left\{l'_{\hat{n}_i}(\vec{D}_{t_i}, \theta)\right\}$ satisfies the regularity conditions in Lemma 2.

However, because we are unable to calculate $\{N_n(\theta)\}$, instead as shown in (55), we take $Var(\hat{n}_i(\theta))$ as $M_i(\theta)$ which means that the second moment is misspecified. But we can still achieve the consistency under certain conditions according to Lemma A.1.

Proposition A.3. *Suppose $l'_{\hat{n}_i}(\vec{D}_{t_i}, \theta)$ and $\hat{l}_{\hat{n}_i}(\vec{D}_{t_i}, \theta)$ satisfy the conditions in Lemma 1, and for any $\theta \in \Theta$,*

$$
\begin{aligned}
\lim_{n\to\infty} E[||X_{t_n} - \hat{X}_{t_n|t_{n-1}}||^3] &= 0 && (91)\\
\textit{and} \quad \lim_{n\to\infty} E[||X_{t_n} - \hat{X}_{t_n|t_{n-1}}||^4] &= 0, && (92)
\end{aligned}
$$

where $||\cdot||$ denotes the Euclidean norm. Then

$$
\hat{\theta}_n \to \theta_0 \quad \textit{i.p.}. \qquad (93)
$$

Proof: Since $l'_{\hat{n}_i}(\vec{Y}_{t_i}, \theta)$ is a Gaussian log-likelihood function whose first two conditional moments are correctly specified, according to Lemma A.1 and Lemma A.2, we need only to prove that

$$
\frac{1}{n}\sum_{i=1}^{n} l'_{\hat{n}_i}(\vec{Y}_{t_i}, \theta) - \frac{1}{n}\sum_{i=1}^{n} \hat{l}_{\hat{n}_i}(\vec{Y}_{t_i}, \theta) \to 0 \quad \text{i.p.}, \qquad (94)
$$

uniformly in Θ. This is equivalent to showing that $l'_{\hat{n}_i}(\theta) - \hat{l}_{\hat{n}_i}(\theta)$ converges uniformly to zero in probability as i goes to infinity. Since the only difference between $l'_{\hat{n}_i}(\theta)$ and $\hat{l}_{\hat{n}_i}(\theta)$ is their covariance, we need to show that the differences of those covariances will vanish uniformly as n goes to infinity. So what we need to prove is that

$$\varphi_i(\theta) \to 0 \quad \text{as} \quad i \to \infty, \quad \text{uniformly on} \quad \Theta \tag{95}$$

where $\varphi_i(\theta)$ is defined in (90).

According to (87) and (90), we have that

$$||\varphi_n(\theta)|| \to 0, \tag{96}$$

if

$$\lim_{i\to\infty} E\{||X_{t_i} - \hat{X}_{t_i|t_{i-1}}||^3\} = 0 \tag{97}$$

$$\text{and} \quad \lim_{i\to\infty} E\{||X_{t_i} - \hat{X}_{t_i|t_{i-1}}||^4\} = 0, \tag{98}$$

where $||\cdot||$ denotes the Euclidean norm. This completes the proof. □

Since generally we cannot guarantee the conditions (91) and (92), in the next section, we will use Monte Carlo analysis to confirm the performance of the proposed estimator.

Notes

[1]Recent work can be found in Zhou (1997), and Duffie and Lando (2001).

[2]Similar concept has also been used in Albanese et al. (2003).

[3]The existence of the extended risk-free short rate process r_t that satisfies (2) and (3) has been proved in Chen and Filipović (2003a), Lemma 4.1.

[4]The deductions can be found in Chen and Filipović (2003b).

[5]For the notion and general results of quadratic processes, we refer to Chen, Filipović and Poor (2003).

[6]For the detail of genetic algorithm for optimization, we refer to Goldberg (1989).

References

Ahn, Dong-Hyun. "Generalized Squared-Autoregressive-Independent-Variable Nominal Term Structure Model." Working paper, Kenan-Flagler Business School, University of North Carolina, Chapel Hill (1997).

Ahn, D. H., R. F. Dittmar, and A. R. Gallant. "Quadratic Term Structure Models: Theory and Evidence." *The Review of Financial Studies* 15 (2002), 243–288.

Albanese, C., J. Campolieti, O. Chen, and A. Zavidonov. "Credit Barrier Models." to appear in Risk Magazine, 2003.

Black, F., and M. Scholes. "The Valuation of Options and Corporate Liabilities." *Journal of Political Economy* 81 (1973), 637–654.

Bollerslev, T., and J. M. Wooldridge. "Quasi-Maximum Likelihood Estimation of Dynamic Models with Time-Varying Covariances." *Econometric Reviews*, 11 (1992), 143–172.

Brigo, D., and F. Mercurio. *Interest Rate Models: Theory and Practice.* Springer-Verlag, Berlin 2001.

Chen, L., and D. Filipović. "A Simple Model for Credit Migration and Spread Curves." Working paper, Princeton University (2003a).

Chen, L., and D. Filipović. "Credit Derivatives in An Affine Framework." Working paper, Princeton University (2003b).

Chen, L., D. Filipović, and H. V. Poor. "Quadratic Term Structure Models for Risk-free and Defaultable Rates." forthcoming in *Mathematical Finance* (2003).

Chen, L., and H. V. Poor. "Parametric Estimation of Quadratic Models for the Term Structure of Interest Rate." Working paper, Princeton University, 2003.

Chen, R., and L. Scott. "Multi-Factor Cox-Ingersoll-Ross Models of the Term Structure: Estimates and Tests from a Kalman Filter Model." Working paper, University of Georgia (1995).

Cheridito, P., and D. Filipović. "Equivalent and Absolutely Continuous Measure Changes for Jump-Diffusion Processes." Working paper, Princeton University (2003).

Cheridito, Patrick, D. Filipović, and R. Kimmel. "Market Price of Risk Specifications for Affine Models: Theory and Evidence." Working paper, Princeton University, 2003.

Clark, S. A. "The Valuation Problem in Arbitrage Price Theory." *Journal of Mathematical Economics*, 22 (1993), 463–478.

Cox, J. C., J. E. Ingersoll, and S. A. Ross. "A Theory of the Term Structure of Interest Rates." *Econometrica*, 53 (1985), 385–407.

Delianedis, G., and R. Geske. "The Components of Corporate Credit Spreads: Default, Recovery, Tax, Jumps, Liquidity, and Market Factors." Working paper, UCLA, 2001.

Driessen, J. "Is Default Event Risk Priced in Corporate Bonds." Working paper, University of Amsterdam (2002). Duan, J., and J. G. Simonato. "Estimating

Exponential-Affine Term Structure Models by Kalman Filter." *Review of Quantitative Finance and Accounting*, 13 (1999), 111–135.

Duffee, G. "Estimating the Price of Default Risk." *The Review of Financial Studies*, 12 (1999), 197–226.

Duffee, G. "Term Premia and Interest Rate Forecasts in Affine Models" *The Journal of Finance*, 15 (1999), 405–443.

Duffie, D., D. Filipović, and W. Schachermayer. "Affine Processes and Applications in Finance." *The Annals of Applied Probability* 13 (2003), 984–1053.

Duffie, D., and D. Lando. "Term Structures of Credit Spreads with Incomplete Accounting Information." *Econometrica*, 69 (2001), 633–664. Duffie, D., and K.

J. Singleton."Modeling Term Structures of the Defaultable Bonds." *The Review of Financial Studies* 12 (1999), 687–720.

El Karoui, N., and L. Martellini. "A Theoretical Inspection of the Market Price for Default Risk." Working paper, Ecole Polytechnique (2002).

Gallant, A. R., and H. White. *A Unifed Theory of Estimation and Inference for Nonlinear Dynamic Models*, Basil Blackwell, Oxford, 1988.

Goldberg, D. E. *Genetic Algorithms in Search, Optimization, and Machine Learning.* Addison-Wesley Pub Co., 1989.

Huang, J., and M. Huang. "How Much of the Corporate-Treasury Yield Spread is Due to Credit Risk?" Working paper, Stanford University, 2003.

Hull, J., and A. White. "Valuing Credit Default Swaps II: Modeling Default Correlations." *The Journal of Derivatives*, 8 (2001), 12–22.

Jarrow, R. A., D. Lando, and S. A. Turnbull. "A Markov Model for the Term Structure of Credit Risk Spreads." *The Review of Financial Studies*, 10 (1997), 481–523.

Jarrow, R. A., D. Lando, and F. Yu. "Default Risk and Diversification: Theory and Applications." Working paper, Cornell University (2003).

Jarrow, R. A., and S. Turnbull. "Pricing Options on Financial Securities Subject to Default Risk." *Journal of Finance* 50 (1995), 53–86.

Kullback, S., and R. A. Leibler. "On Information and Sufficiency," *Annals of Mathematical Statistics* 22 (1951), 79–86.

Lando, D. "On Cox Processes and Credit-Risky Securities." *Review of Derivatives Research*, 1 (1998), 99–120.

Leippold, M., and L. Wu. "Design and Estimation of Quadratic Term Structure Models." Working paper, 2001.

Leippold, M., and L. Wu "Asset Pricing Under The Quadratic Class." Forthcoming, *Journal of Financial and Quantitative Analysis*, 2002.

Madan, D. B., and H. Unal. "Pricing the Risks of Default." Working paper, The Wharton Financial Institution Center, 1996.

Maybeck, P. S. *Stochastic Models, Estimation and Control*, Academic Press, London, 1982.

Merton, R. "On the Pricing of Corporate Debt: the Risk Structure of Interest Rates." *Journal of Finance* 29 (1974), 449–470. Moody's Special Comment. "Default and

Recovery Rates of Corporate Bond Issuers: A Statistical Review of Moody's Ratings Performance 1970-2001." Moody's Investors Service, Global Credit Research, 2002.

Peracchi, F. *Econometrics*, John Wiley and Sos Ltd., New York, 2000.

White, H. "Maximum Likelihood Estimation of Misspecified Models." *Econometrica*, 50 (1982), 1–25. Yu, F. "Modeling Expected Return on Defaultable Bonds." *Journal of Fixed Income*, 12 (2002), 69–81.

Zhou, C., "A Jump-Diffusion Approach to Modeling Credit Risk and Valuing Defaultable Securities." Finance and Economics Discussion Series, Board of Governors of the Federal Reserve, 1997.

List of Figures

Figure 1: Comparison of Actual Yields and Fitted Yields for the Affine Model

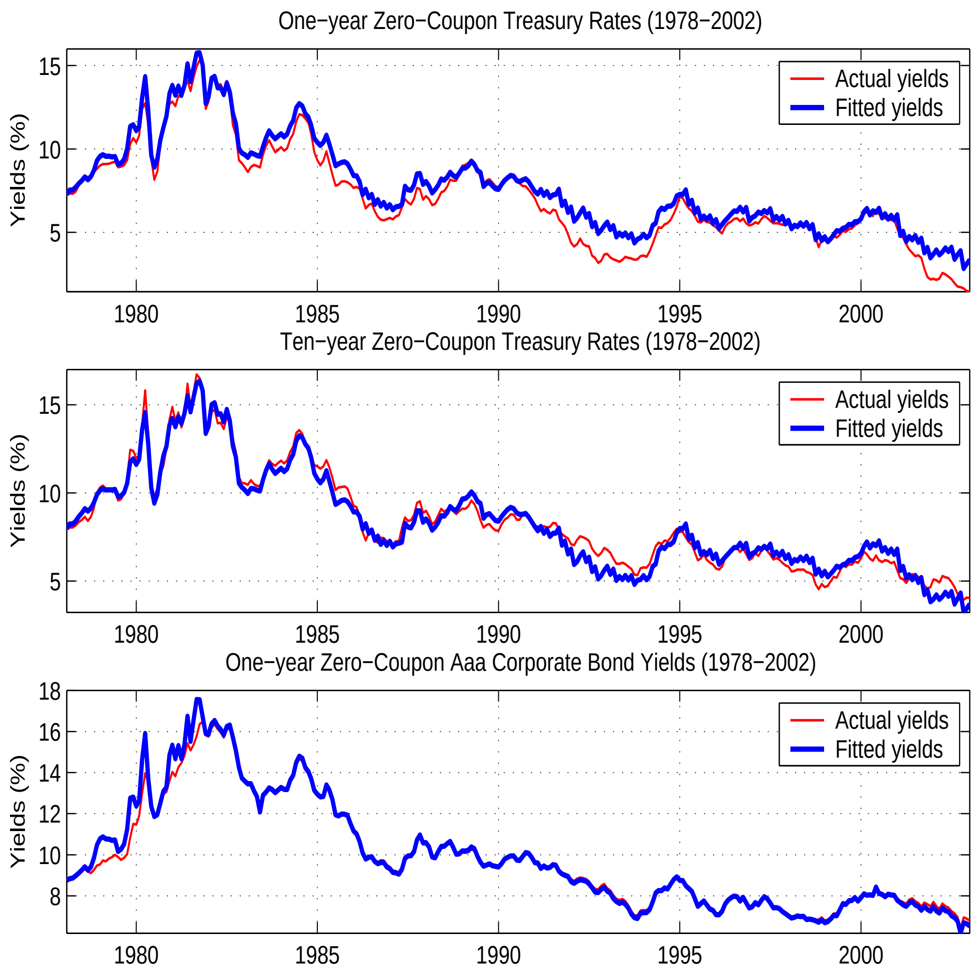


Figure 2: Comparison of Actual Yields and Fitted Yields for the Affine Model (Cont'd)

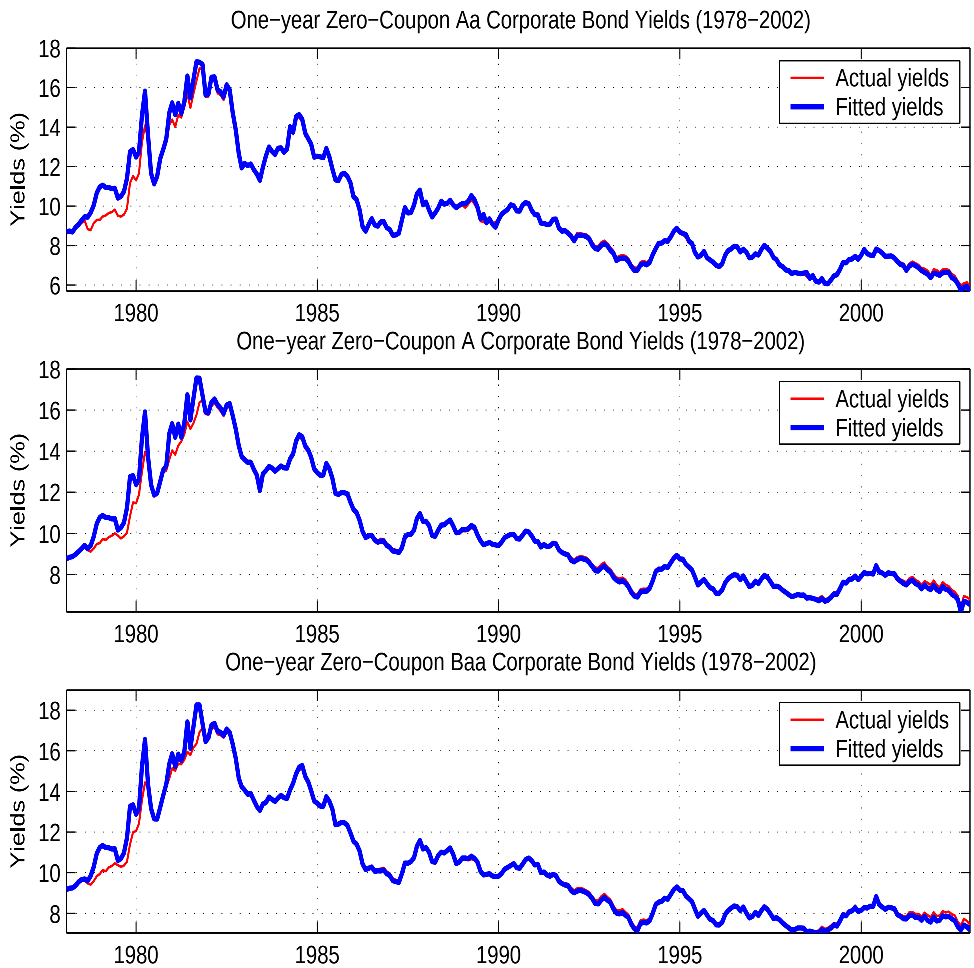


Figure 3: Comparison of Actual Yields and Fitted Yields for the Quadratic Model

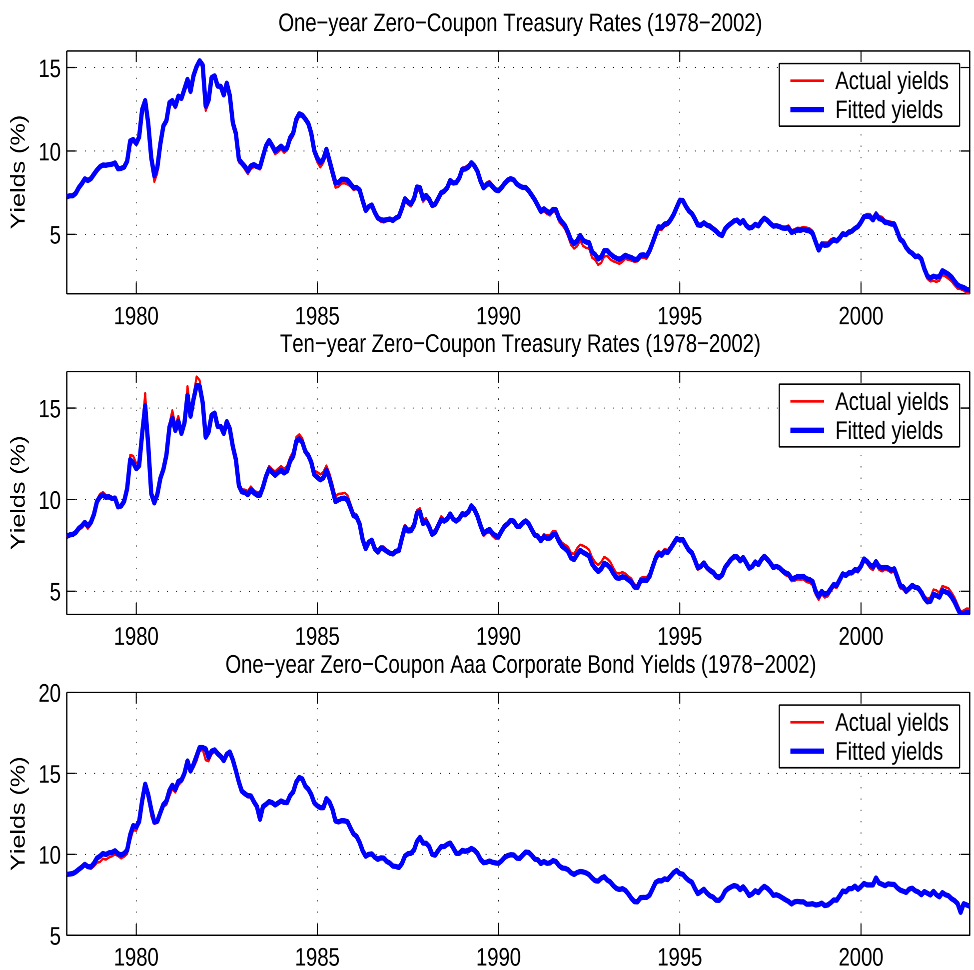


Figure 4: Comparison of Actual Yields and Fitted Yields for the Quadratic Model (Cont'd

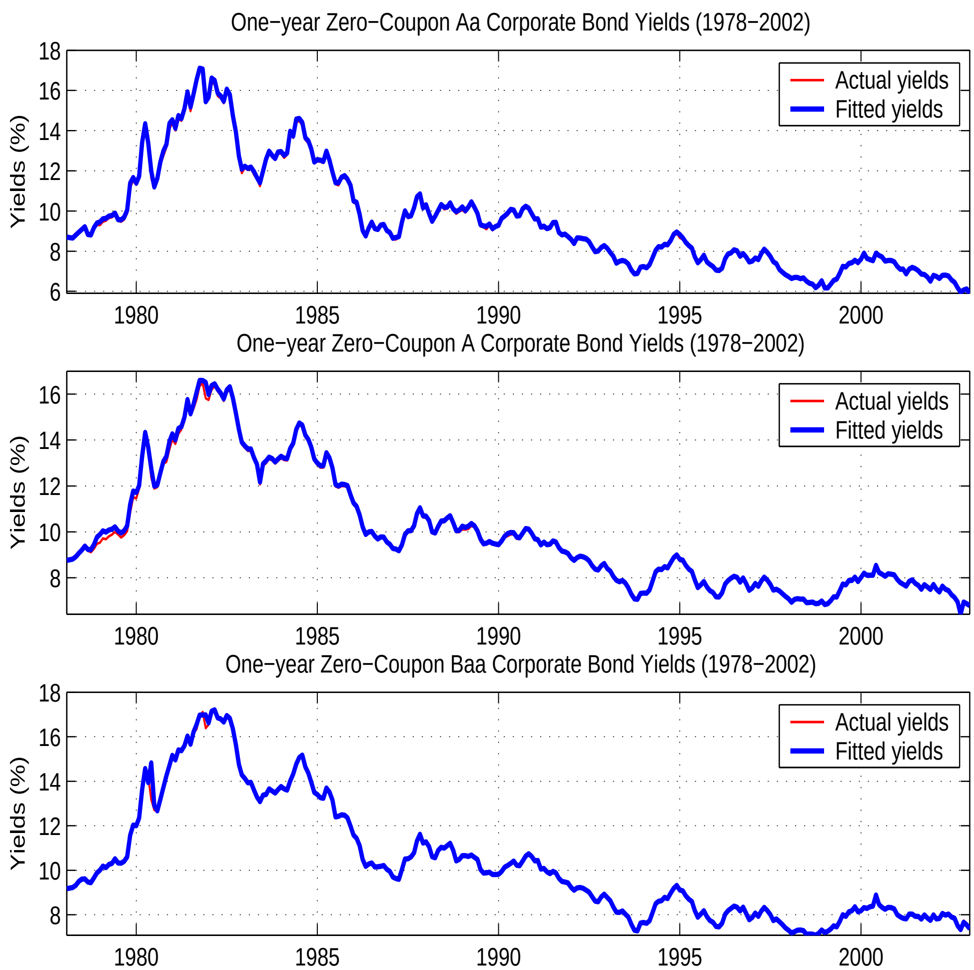


Figure 5: Corporate Bond Spreads with the Non-Default Part

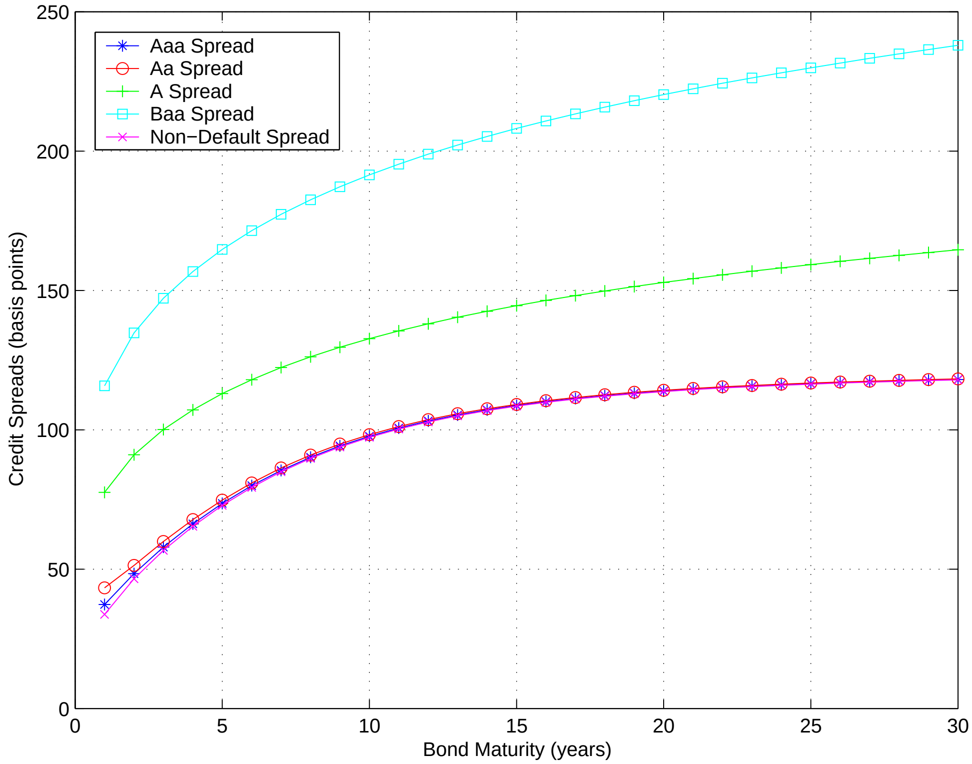


Figure 6: Default Probabilities for Aa Rated Bonds

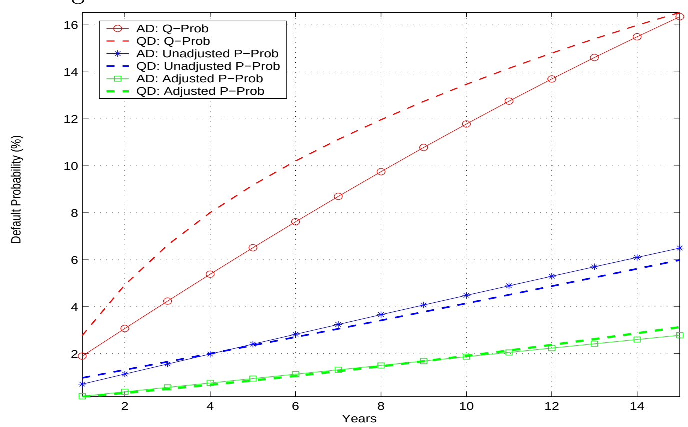


Figure 7: Default Probabilities for A Rated Bonds

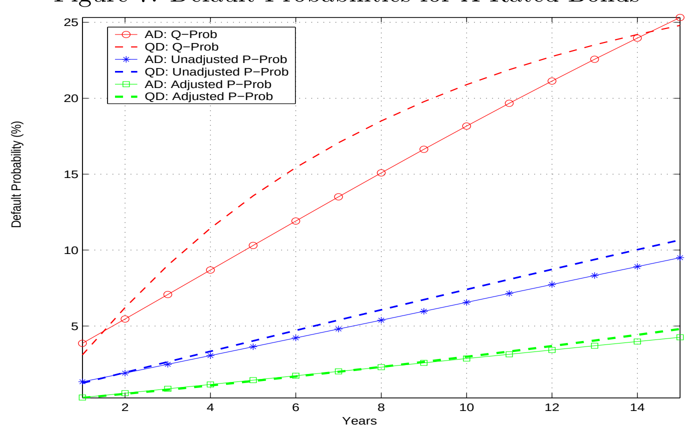


Figure 8: Default Probabilities for Baa Rated Bonds

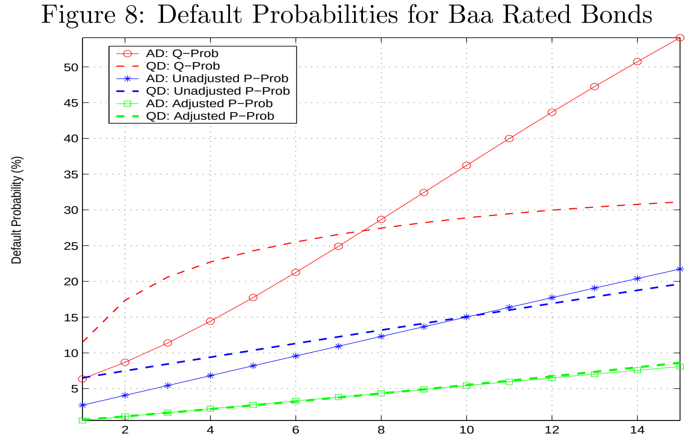


Figure 9: Credit Indices for the Affine Diffusion Model

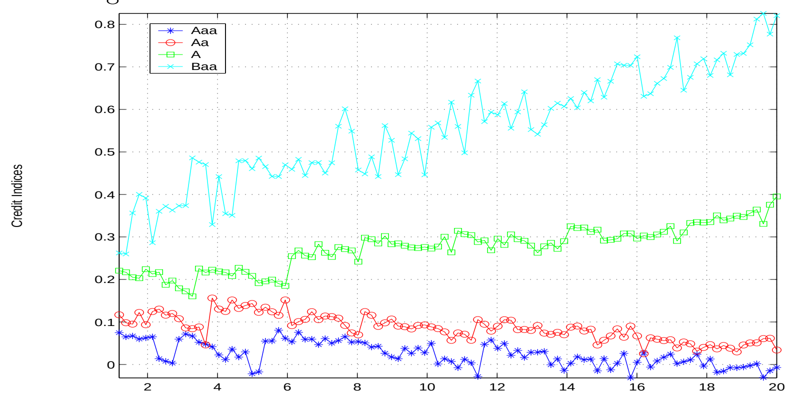


Figure 10: Credit Indices for the Affine Fat-Tail Model

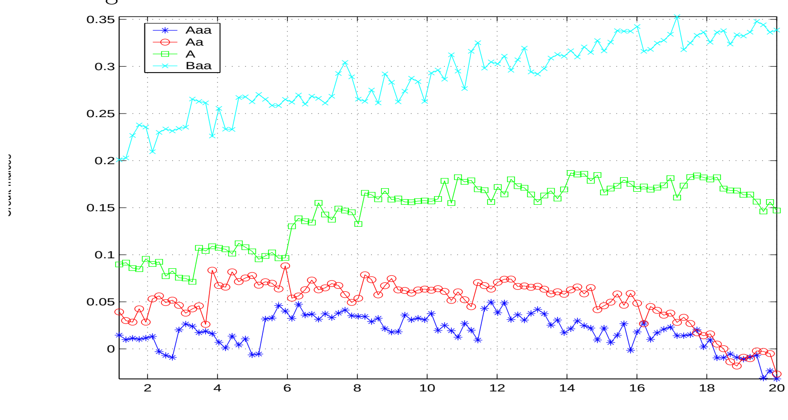
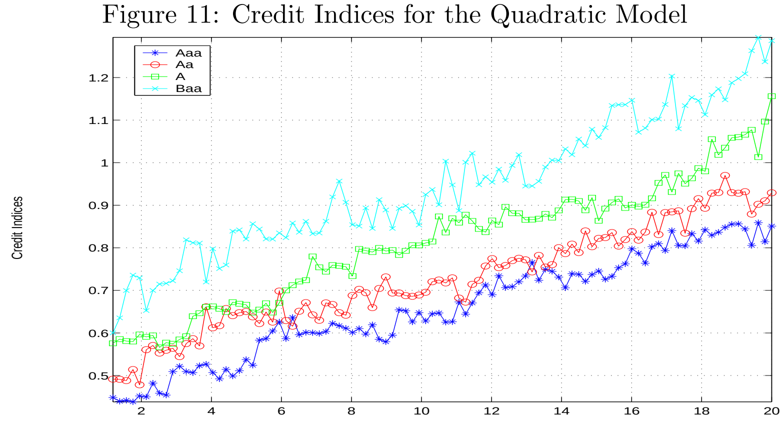


Figure 11: Credit Indices for the Quadratic Model



List of Tables

Table 1: The Summary of Statistics of Interest Rates

Treasury Rates (%)	3 month	1 Year	10 Year	30 Year
Period of Time Series	01/78-12/02	01/78-12/02	01/78-12/02	01/78-12/02
Number of Obs.	300	300	300	289
Mean	6.592	7.264	8.244	8.526
Std. Dev.	3.008	3.188	2.618	2.357
Median	5.77	6.52	7.86	8.14
Minimum	1.19	1.45	3.87	5.01
Maximum	16.3	16.32	15.32	14.68
1-Yr Corporate Bond Yields (%)	**Aaa**	**Aa**	**A**	**Baa**
Period of Time Series	01/78-12/02	01/78-12/02	01/78-12/02	01/78-12/02
Number of Obs.	300	300	300	300
Mean	9.206	9.504	9.837	10.298
Std. Dev.	2.266	2.578	2.534	2.637
Median	8.755	8.995	9.345	9.765
Minimum	6.15	5.93	6.42	7.09
Maximum	16.39	16.47	16.97	17.18

Means, standard deviations, and other statistics of the Treasury rates and the averaged 1-Yr. Corporate Bond Yields are summarized. The Treasury rates are collected with four different maturities 3-month, 1-year, 10-year, and 30-year.The corporate bond years are collected from four rating classes: Aaa, Aa, A, and Baa. The sample period ranges from January 1978 to December 2002 with totally 300 observations for all different rates.

Table 2: Summary of Statistics for a Snapshot of the Market Rates

	Maturity (years)						
Quality	0-2	2-4	4-6	6-8	8-10	10-20	Total
Treasury	3	12	8	6	4	17	50
Aaa	20	24	34	31	32	26	166
Aa	19	30	24	37	22	30	162
A	17	38	28	32	24	20	159
Baa	17	32	46	30	18	16	159

The snapshot of treasury and corporate bond price data are downloaded from www. Bondpage. com. This dataset includes over 700 month-end quoted prices of treasury bonds and corporate bonds issued by the investment-grade firms. Each bond contains the information about the rating given by Moody's, the maturity date, the coupon rate and payment frequency, and the bond's clean and dirty price. All the bonds are non-callable and have at least one year remaining to maturity. All the bonds share the same settlement date.

Table 3: Estimates of the Affine Model Using Filtering Technique and QMLE

Rating	Aaa	Aa	A	Baa
b_1	0.00156	0.00155	0.00156	0.00157
	(0.00021)	(0.00016)	(0.00018)	(0.00023)
$\tilde{b}_1$	0.0352	0.0325	0.0335	0.0350
	(0.0012)	(0.0015)	(0.0018)	(0.0013)
$\tilde{\beta}_{11}$	-0.0744	-0.0831	-0.0729	-0.0717
	(0.0064)	(0.0068)	(0.0067)	(0.0051)
β_{11}	-0.103	-0.117	-0.104	-0.103
	(0.051)	(0.052)	(0.068)	(0.045)
α_1	1.168e-4	7.046e-4	1.021e-4	8.544e-4
	(0.12e-5)	(0.16e-5)	(0.22e-5)	(0.32e-5)
b_2	0.0254	0.0328	0.0380	0.0422
	(0.014)	(0.016)	(0.015)	(0.019)
β_{22}	-0.725	-0.620	-0.614	-0.602
	(0.15)	(0.18)	(0.19)	(0.25)
$\tilde{\beta}_{22}$	-5.401	-5.741	-5.263	-5.812
	(1.32)	(2.09)	(2.01)	(1.83)
β_{21}	0.0595	0.190	0.481	0.638
	(0.063)	(0.073)	(0.064)	(0.098)
$\tilde{\beta}_{21}$	0.00352	0.00366	0.00381	0.00396
	(0.00021)	(0.00019)	(0.00012)	(0.00024)
α_2	4.658	4.766	4.714	4.643
	(0.43)	(0.49)	(0.46)	(0.45)
c	0.00115	0.00379	0.00221	0.000560
	(0.0011)	(0.0009)	(0.0016)	(0.0005)
γ_1	0.0248	0.0343	0.0481	0.0743
	(0.0034)	(0.0028)	(0.0044)	(0.0064)

Yields	Mean Error (basis points)	Mean Percentage Error (%)	$\sqrt{\text{Mean Square Error}}$ (basis points)
3-mo. treasury	-28.482	-12.453	48.342
1-yr. treasury	-22.806	-11.638	43.109
10-yr. treasury	6.1456	0.5404	27.639
30-yr. treasury	5.1374	0.2321	15.684
1-yr. Aaa rating bonds	-8.5162	-0.47533	15.064
1-yr. Aa rating bonds	7.7607	0.55684	14.651
1-yr. A rating bonds	-5.7947	-0.31322	14.444
1-yr. Baa rating bonds	-4.8276	-0.22293	14.073

The parameters are estimated from 25-year time-series of monthly treasury yields with maturity 3-month, 1-year, 10- year and 30-year and the 1-year corporate bond yields from four different Moody's rating classes: *Aaa*, *Aa*, *A*, *Baa*. The estimated values together with the standard errors (in parentheses) are presented. The standard errors are computed as with QMLE and corrected for heteroskedasticity as described in White (1982). Moreover the fitting error are presented with three different measures: mean error (actual-fitted), mean percentage error ((actual-fitted)/actual) and the square root of the mean square errors.

Table 4: Estimates of the Quadratic Model Using Nonlinear Filtering Technique and QMLE

Rating	Aaa	Aa	A	Baa
b_1	0.00404	0.00405	0.00354	0.00489
	(0.00062)	(0.00054)	(0.00034)	(0.00042)
$\tilde{b}_1$	0.0133	0.0141	0.0134	0.0148
	(0.0042)	(0.0048)	(0.0068)	(0.0057)
β_{11}	-0.000104	-0.000166	-0.000103	-0.000132
	(0.000075)	(0.000072)	(0.000078)	(0.000053)
$\tilde{\beta}_{11}$	-0.00200	-0.00231	-0.00437	-0.00282
	(0.0045)	(0.0058)	(0.0050)	(0.0055)
α_1	0.00141	0.00117	0.000745	0.000391
	(0.00022)	(0.00023)	(0.00019)	(0.00018)
b_2	0.119	0.194	0.230	0.295
	(0.084)	(0.073)	(0.068)	(0.063)
β_{22}	-0.188	-0.175	-0.168	-0.163
	(0.35)	(0.29)	(0.29)	(0.28)
$\tilde{\beta}_{22}$	-5.221	-5.307	-5.340	-5.548
	(2.21)	(2.10)	(1.91)	(1.95)
β_{21}	0.000932	0.000934	0.000955	0.000964
	(0.0034)	(0.0023)	(0.0026)	(0.0028)
$\tilde{\beta}_{21}$	0.00352	0.00366	0.00381	0.00396
	(0.0022)	(0.0018)	(0.0023)	(0.0025)
α_2	0.0124	0.0138	0.0342	0.0413
	(0.0076)	(0.0077)	(0.0066)	(0.0075)
c	9.40e-7	7.62e-7	6.34e-6	4.22e-6
	(1.43e-6)	(0.94e-6)	(1.6e-6)	(1.5e-6)
γ_1	0.00479	0.00679	0.0112	0.0148
	(0.00043)	(0.00038)	(0.00042)	(0.00046)

Yields	Mean Error (basis points)	Mean Percentage Error (%)	$\sqrt{\text{Mean Square Error}}$ (basis points)
3-mo. treasury	-22.372	-5.132	30.353
1-yr. treasury	-21.156	-4.589	28.137
10-yr. treasury	5.993	1.283	16.473
30-yr. treasury	4.343	0.332	12.420
1-yr. Aaa rating bonds	-7.443	-1.709	13.034
1-yr. Aa rating bonds	3.321	1.853	12.651
1-yr. A rating bonds	-2.158	-0.218	10.444
1-yr. Baa rating bonds	-0.048273	0.17175	9.899

The parameters are estimated from 25-year time-series of monthly treasury yields with maturity 3-month, 1-year, 10- year and 30-year and the 1-year corporate bond yields from four different Moody's rating classes: *Aaa*, *Aa*, *A*, *Baa*. The estimated values together with the standard errors (in parentheses) are presented. The standard errors are computed as with QMLE and corrected for heteroskedasticity as described in White (1982). Moreover the fitting error are presented with three different measures: mean error (actual-fitted), mean percentage error ((actual-fitted)/actual) and the square root of the mean square errors.

Table 5: Estimates of the Real World Default Intensity

Parameter	Affine	Quadratic
$\tilde{c}$	1.267e-7 (2.5e-6)	1.343e-6 (9.2e-6)
$\tilde{\gamma}_1$	0.00344 (0.0032)	0.000898 (0.0048)
$\tilde{\gamma}_2$	0.772 (0.12)	0.815 (0.31)
RMSE (b.p.)	3.3 b.p.	2.8 b.p.

Table 6: Parameter Estimates for Fitting the Current Term Structure of Treasury Rates and Credit Spreads

Parameter	AD	AF	QD
b_1	0.00952	0.0116	-0.0578
	(0.00024)	(0.000013)	(0.0087)
β_{11}	-0.0690	-0.151	-0.247
	(0.0050)	(0.0027)	(0.81)
α_1	0.00783	0.000673	0.00985
	(0.00034)	(0.00021)	(0.0014)
r	0.0117	0.0109	0.0195
	(0.00041)	(0.00072)	(0.00039)
b_2	0.0118	0.0197	0.37465
	(0.048)	(0.027)	(0.057)
β_{22}	-1.5697	-1.3606	-2.4328
	(1.03)	(0.73)	(0.92)
β_{21}	0.124	0.194	0.124
	(0.032)	(0.053)	(0.042)
α_2	5.323	2.501	0.00771
	(0.63)	(0.71)	(0.0047)
c	9.132e-7	1.570e-6	3.075e-5
	(0.13e-6)	(0.27e-6)	(9.5e-6)
γ_1	0.00153	0.00101	0.0296
	(0.00064)	(0.00042)	(0.00065)

Yield	RMSE (basis points)		
Maturity	AD Model	AF Model	QD Model
0-4 yr. Treasury	3.281	3.089	4.109
4-20 yr. Treasury	0.992	0.852	1.292
0-4 yr. Aaa rated bonds	14.689	9.477	17.870
4-20 yr. Aaa rated bonds	6.235	4.918	6.912
0-4 yr. Aa rated bonds	5.517	5.013	6.939
4-20 yr. Aa rated bonds	2.477	1.311	2.767
0-4 yr. A rated bonds	19.176	17.234	19.798
4-20 yr. A rated bonds	5.435	5.088	6.888
0-4 yr. Baa rated bonds	24.764	23.88	24.627
4-20 yr. Baa rated bonds	6.394	6.165	7.092

The parameters are estimated using the price data of more than 650 treasury and corporate bonds. Thirty independent experiments are performed. The mean estimates together with the standard errors (in parentheses) are presented. The fitting errors (RMSE) are also presented with respect to different maturities and four rating classes.

Table 7: Summary Statistics of Credit Indices

	AD	AF	QD
Rating	Mean (Stdev)	Mean (Stdev)	Mean (Stdev)
Z_{Aaa}	0.0250 (0.0282)	0.0290 (0.0174)	0.660 (0.122)
Z_{Aa}	0.0864 (0.0306)	0.0596 (0.0247)	0.727 (0.120)
Z_{A}	0.275 (0.0504)	0.156 (0.0336)	0.814 (0.144)
Z_{Baa}	0.554 (0.131)	0.300 (0.0372)	0.948 (0.159)