ECONOMETRICS AND RISK MANAGEMENT
ADVANCES IN ECONOMETRICS

Series Editors: Thomas B. Fomby and R. Carter Hill
CONTENTS

LIST OF CONTRIBUTORS vii

INTRODUCTION ix

FAST SOLUTION OF THE GAUSSIAN COPULA MODEL
Bjorn Flesaker 1

AN EMPIRICAL STUDY OF PRICING AND HEDGING COLLATERALIZED DEBT OBLIGATION (CDO)
Lijuan Cao, Zhang Jingqing, Lim Kian Guan and Zhonghui Zhao 15

THE SKEWED $t$ DISTRIBUTION FOR PORTFOLIO CREDIT RISK
Wenbo Hu and Alec N. Kercheval 55

CREDIT RISK DEPENDENCE MODELING WITH DYNAMIC COPULA: AN APPLICATION TO CDO TRANCHE
Daniel Totouom and Margaret Armstrong 85

PERTURBED GAUSSIAN COPULA
Jean-Pierre Fouque and Xianwen Zhou 103

THE DETERMINANTS OF DEFAULT CORRELATIONS
Kanak Patel and Ricardo Pereira 123
DATA MINING PROCEDURES IN GENERALIZED
COX REGRESSIONS
   Zhen Wei 159

JUMP DIFFUSION IN CREDIT BARRIER
MODELING: A PARTIAL INTEGRO-DIFFERENTIAL
EQUATION APPROACH
   Jingyi Zhu 195

BOND MARKETS WITH STOCHASTIC VOLATILITY
   Rafael DeSantiago, Jean-Pierre Fouque and Knut Solna 215

TWO-DIMENSIONAL MARKOVIAN
MODEL FOR DYNAMICS OF AGGREGATE
CREDIT LOSS
   Andrei V. Lopatin and Timur Misirpashaev 243

CREDIT DERIVATIVES AND RISK AVERSION
   Tim Leung, Ronnie Sircar and Thaleia Zariphopoulou 275
LIST OF CONTRIBUTORS

*Margaret Armstrong*  
Center for Industrial Economics (CERNA), École des Mines de Paris, Paris, France

*Lijuan Cao*  
Financial Studies, Fudan University, Shanghai, China

*Rafael DeSantiago*  
IESE Business School, University of Navarra, Barcelona, Spain

*Bjorn Flesaker*  
Quantitative Financial Research Group, Bloomberg LP, New York, NY

*Thomas B. Fomby*  
Department of Economics, Southern Methodist University, Dallas, TX

*Jean-Pierre Fouque*  
Department of Statistics and Applied Probability, University of California, Santa Barbara, Santa Barbara, CA

*Wenbo Hu*  
Quantitative Trading Group, Bell Trading, Chicago, IL

*Zhang Jingqing*  
Financial Studies, Fudan University, Shanghai, China

*Alec N. Kercheval*  
Department of Mathematics, Florida State University, Tallahassee, FL

*Lim Kian Guan*  
Lee Kong Chian School of Business, Singapore Management University, Singapore

*Tim Leung*  
Department of Applied Mathematics and Statistics, Johns Hopkins University, Baltimore, MD
LIST OF CONTRIBUTORS

Andrei V. Lopatin  Numerix LLC, Warrenville, IL
Timur Misirpashaev  Numerix LLC, Warrenville, IL
Kanak Patel  University of Cambridge, Department of Land Economy, Cambridge, UK
Ricardo Pereira  University of Cambridge, Department of Land Economy, Cambridge, UK
Ronnie Sircar  Operations Research and Financial Engineering Department, Princeton University, Princeton, NJ
Knut Solna  Department of Mathematics, University of California, Irvine, CA
Daniel Totouom  Fixed Income, BNP Paribas, New York, NY
Zhen Wei  Department of Statistics, Stanford University, Stanford, CA
Thaleia Zariphopoulou  Department of Mathematics, University of Texas, Austin, TX
Zhonghui Zhao  Department of Finance, Fudan University, Shanghai, China
Xianwen Zhou  Fixed Income Division, Lehman Brothers, New York, NY
Jingyi Zhu  Department of Mathematics, University of Utah, Salt Lake City, UT
INTRODUCTION

The main theme of this volume is credit risk and credit derivatives. Recent developments in financial markets show that appropriate modeling and quantification of credit risk is fundamental in the context of modern complex structured financial products. Moreover, there is a need for further developments in our understanding of this important area. In particular modeling defaults and their correlation has been a real challenge in recent years, and still is. This problem is even more relevant after the so-called subprime crisis that hit in the summer of 2007. This makes the volume very timely and hopefully useful for researchers in the area of credit risk and credit derivatives.

In this volume, we compile a set of points of view on credit risk when it is looked at from the perspective of Econometrics and Financial Mathematics. The volume comprises papers by both practitioners and theoreticians with expertise in financial markets in general as well as in econometrics and mathematical finance in particular. It contains nine contributions presented at the Advances in Econometrics Conference in Baton Rouge at Louisiana State University on November 3–5, 2006. It also features two additional invited contributions: Bjorn Flesaker (Chapter 1) offers an introduction to the popular Gaussian copula model and an original and efficient calibration method. Tim Leung, Ronnie Sircar, and Thaleia Zariphopoulou (Chapter 11) present another very interesting point of view based on risk aversion.

Chapters 2–6 discuss copula methods from various perspectives. In Chapter 2 Lijuan Cao, Zhang Jingqing, Lim Kian Guan, and Zhonghui Zhao compare Monte Carlo and analytic methods for pricing a collateralized debt obligation (CDO). Wenbo Hu and Alec N. Kercheval in Chapter 3 propose a method to calibrate a full multivariate skewed $t$-distribution. In Chapter 4 Daniel Totouom and Margaret Armstrong develop a new class of copula processes in order to introduce a dynamic dependence between default times. Jean-Pierre Fouque and Xianwen Zhou in Chapter 5 use an asymptotic analysis to construct corrections to the classical Gaussian copula. Kanak Patel and Ricardo Pereira in Chapter 6 present an empirical study of corporate bankruptcy within the framework of structural models.
Then, in Chapter 7 Zhen Wei examines hazard rate models using statistical data mining procedures. In Chapter 8 Jingyi Zhu introduces jumps in the first passage approach to default by a careful analysis of the associated partial integro-differential equation (PIDE). In Chapter 9 Rafael DeSantiago, Jean-Pierre Fouque, and Knut Solna show how to introduce stochastic volatility corrections in bond modeling. Andrei Lopatin and Timur Misirpashaev in Chapter 10 discuss a “top-down” approach using a Markovian projection technique to model and calibrate the portfolio loss distribution.

Eric Hillebrand (LSU) suggested the two co-editors Jean-Pierre Fouque and Knut Solna for this volume. We would like to thank him here and also acknowledge the presentation he gave on a joint work with Don Chance during the conference in Baton Rouge.

Jean-Pierre Fouque
Thomas B. Fomby
Knut Solna
FAST SOLUTION OF THE GAUSSIAN COPULA MODEL

Bjorn Flesaker

1. INTRODUCTION

This article describes a new approach to compute values and sensitivities of synthetic collateralized debt obligation (CDO) tranches in the market-standard, single-factor, Gaussian copula model with base correlation. We introduce a novel decomposition of the conditional expected capped portfolio loss process into “intrinsic value” and “time value” components, derive a closed form solution for the intrinsic value, and describe a very efficient computational scheme for the time value, taking advantage of a curious time stability of this quantity.

The underlying CDO structure, the Gaussian copula framework, and the base correlation concept will be described very briefly since they have been described in detail elsewhere, see for example, the papers by Li (2000); Andersen, Sidenius, and Basu (2003); Hull and White (2004); and McGinty and Ahluwalia (2004).

2. THE SYNTHETIC CDO STRUCTURE

A synthetic CDO tranche is a credit default swap characterized by a settlement currency; a portfolio, defined by a list of names and accompanying
notional amounts covered; a maturity date and premium payment schedule; a premium rate; and finally an attachment point and a detachment point, determining the beginning and end of the portfolio loss covered by the tranche. A tranche is known as an equity tranche if the attachment point equals zero, as a super-senior tranche if the detachment point equals the underlying portfolio notional, and as a mezzanine tranche if it is neither of the above. We can always analyze arbitrary tranches as the difference between two equity tranches: a long position in one detaching at the actual tranche’s detachment point along with a short position in one detaching at the actual tranche’s attachment point. This approach is necessary in the base correlation framework, where the correlation parameter is separately given for each detachment point. In the following sections, we will therefore focus on the modeling of equity tranches.

3. VALUATION ASSUMPTIONS

We follow the standard industry practice in credit default swap modeling and assume that we can value all cash flows by taking their expected value under a “risk neutral,” perhaps more accurately described as “risk adjusted,” probability measure and discounting the resulting quantities with the initial yield curve implied from the interest rate swap market. This is tantamount to assuming that interest rates are deterministic or that their dynamics are statistically independent of the default processes. We will further assume that each name upon default has a known fixed recovery rate as a fraction of par, which is an input to the model. We do not require the portfolio members to have homogeneous recoveries or, for that matter, notional amounts. We imply (“strip”) risk neutral default probabilities from single name CDS quotes for the underlying portfolio components, assuming that the survival function is piecewise exponential in time, that is in a manner consistent with hazard rates being constant in between available CDS maturity dates.

4. THE MODEL

The analysis takes as a starting point a function that provides a market implied unconditional risk neutral default probability $p_i(t)$ for each name $i$ to each date of interest $t$. Given a value $x$ of the common market factor, the content of the single-factor Gaussian copula model is that the event that
name $i$ defaults before time $t$ is independent of the default of all other names and occurs with probability:

$$p_i(t, x) = \Phi \left( \frac{\Phi^{-1}(p_i(t)) - a_i x}{\sqrt{1 - a_i^2}} \right)$$

where $a_i$ is the factor loading for name $i$ and $\Phi$ denotes the standard cumulative normal (Gaussian) distribution function. A historically important special case is the constant (compound) correlation model where $a_i \equiv \sqrt{\rho}$ and each (mezzanine) tranche is valued at its own “implied correlation.” This was the market standard in the early days of the synthetic CDO market. In 2004/2005, it was broadly replaced by the base correlation version of the model, which amounts to each attachment/detachment point having its own input level of $\rho$. In the following text, we will focus on the base correlation version, setting $a_i \equiv \sqrt{\rho}$, with the understanding that different tranches will typically be valued with different values of $\rho$.

The fractional loss and recovery processes for the underlying basket are given by $L(t)$ and $R(t)$, respectively, defined as follows:

$$L(t) = \frac{\sum_i \mathbb{1}_{\tau_i < t} N_i (1 - R_i)}{\sum_i N_i}$$

$$R(t) = \frac{\sum_i \mathbb{1}_{\tau_i < t} N_i R_i}{\sum_i N_i}$$

where $\tau_i$ denotes the default time of name $i$, $N_i$ is the notional amount of name $i$ in the basket, $R_i$ the fractional recovery upon default of name $i$, and $\mathbb{1}_A$ an indicator variable that takes on the value 1 if $A$ is true and 0 otherwise.

As described above, we will focus on the valuation of equity tranches, which for a given portfolio can be characterized by their detachment point, generally expressed as a percentage of the portfolio notional. Given such a fractional detachment point, $D$, we are concerned with the risk neutral expectations of the capped loss and recovery processes, where the latter is used in the calculation of the premium leg for super-senior tranches (and, somewhat hypothetically, any other tranche with a detachment point higher than the maximum portfolio loss):

$$\hat{L}(t) = \min[L(t), D]$$

$$\hat{R}(t) = \max[R(t) - (1 - D), 0]$$
With precise knowledge of these quantities for each date between the current time and maturity, the present value of the default leg and of the unit spread premium leg of a synthetic CDO equity tranche detaching at \( D \), as a fraction of the underlying portfolio notional, can be estimated by taking the expectation of the discounted loss and premium cash flows under the risk neutral pricing measure as follows:

\[
V_d = \mathbb{E} \int_0^T P(t) dL(t)
\]

\[
= P(T) \mathbb{E} \hat{L}(T) + \int_0^T f(t) P(t) \mathbb{E} \hat{L}(t) dt
\]  

(6)

\[
V_p = \sum_{j=1}^{J} P(t_j) \delta(t_{j-1}, t_j) \left[ D - \int_{t_{j-1}}^{t_j} (\mathbb{E} \hat{L}(t) + \mathbb{E} \hat{R}(t)) dt \right]
\]

(7)

where \( P(t) \) denotes the discount factor to time \( t \), \( f(t) \) is the instantaneous forward interest rate for time \( t \), \( t_j \) denotes premium payment dates, with \( t_J = T \), and \( \delta(t_{j-1}, t_j) \) is the daycount fraction between consecutive premium payment dates. The second expression in Eq. (6) follows from integration by parts of the integral in the first expression and by exchanging the order of integration and expectation (which is admissible by Fubini’s theorem, since all relevant quantities are explicitly bounded). The integral term subtracted from the initial tranche notional in Eq. (7) represents a slight approximation to the reduction in notional by tranche losses and write-down from above (the representation would be exact if discount rates were zero and/or all tranche losses occurred at the end of each premium period). The model value of the tranche, aside from premium accrual, is given by \( V_d/C_0 - cV_p \), where \( c \) is the contractual premium rate, and the breakeven premium (sometimes referred to as replacement deal spread) on a tranche with no upfront payment is found as the ratio \( V_d/V_p \).

The capped recovery process is only of interest for super-senior tranches, and given the assumption of known recovery per name, \( \mathbb{E} \hat{R}(t) \) can be found by a minor variation of the routine used to solve for the capped loss process, \( \mathbb{E} \hat{L}(t) \). We evaluate the time integrals above by first solving for the expected capped losses (and, if necessary, recoveries) for a discrete set of points in time that include the premium payment dates. We fit a cubic spline to the resulting function of time, and evaluate the integrals analytically under the mild assumption of piecewise constant forward interest rates between the knot points of the spline.

BJORN FLESAKER
5. PRICING

For each \( \{x, t\} \), \( L(t) \) is then the sum of a set of independent and generally heterogeneous binary random variables. Note that, in principle, we need to know the expected capped loss for each point in time until maturity for each possible value of the Gaussian state variable in order to calculate the unconditional expected capped loss required in Eqs. (6) and (7).

\[
\mathbb{E}[\hat{L}(t)] = \int_{-\infty}^{\infty} \mathbb{E}[\min(L(t), D)|x] \phi(x) \, dx
\]

where \( \phi(x) \) is the standard Gaussian probability density function.

Fig. 1 shows the conditional capped expected loss surface, \( \mathbb{E}[\min(L(t), D)|x] \), as a function of time to maturity, \( t \), and the value of the common factor value, \( x \). In this figure, we are looking at a 125 name CDX investment grade portfolio with a 7% detachment point and a 30%...
correlation. The qualitative behavior seen in the figures is robust across a wide range of spread levels, detachment points and correlations.

6. A DECOMPOSITION

Consider an equity tranche detaching at $D$, and assume that $\rho > 0$. Let $x^*_t$ be the unique value of $x$ where $\mathbb{E}[L(t)|x] = D$. Then we can rearrange the expression for the expected capped loss on an equity tranche as:

$$
\mathbb{E}[\hat{L}(t)] = \int_{-\infty}^{x^*_t} (D - \mathbb{E}[\max(D - L(t), 0)|x])\phi(x)dx \\
+ \int_{x^*_t}^{\infty} (\mathbb{E}[L(t)|x] - \mathbb{E}[\max(L(t) - D, 0)|x])\phi(x)dx
$$

We can further sort the integrals into “intrinsic value” minus “time value,” where we have liberally borrowed terminology from the option pricing literature:

$$
\mathbb{E}[\hat{L}(t)] = \left( \int_{-\infty}^{x^*_t} D\phi(x)dx + \int_{x^*_t}^{\infty} \mathbb{E}[L(t)|x]\phi(x)dx \right) \\
- \left( \int_{-\infty}^{x^*_t} \mathbb{E}[\max(D - L(t), 0)|x]\phi(x)dx \right) \\
+ \int_{x^*_t}^{\infty} \mathbb{E}[\max(L(t) - D, 0)|x]\phi(x)dx \right)
$$

The idea of calling the first pair of integrals the conditional intrinsic value comes from the fact that they result from pulling an expectation operator through an option-like pay-off, just as what is done to obtain the intrinsic value of a stock option. By analogy, we call the remainder the conditional time value, although “volatility value” or “convexity value” might be more descriptive, since it arises from the interaction of the conditional volatility of the losses with the convexity of the pay-off function, as in the usual illustration of Jensen’s inequality. As we will see, this decomposition is useful because of the analytical tractability of the intrinsic value and the numerical tractability of the time value. Figs. 2–5 illustrate the decomposition of the conditional expected loss surface from Fig. 1.

If we have the corner case of $\rho = 0$, $x^*_t$ will fail to exist and the approach described above will break down. In this case, however, the pricing problem
is greatly simplified by the fact that the common factor is irrelevant, and we are left with a set of unconditionally independent defaults, which can be modeled as described in Section 9. For values of $\rho$ very close to zero (say inside 1%) we may need to interpolate between the results obtained at $\rho = 0$ and values obtained at a “safe” correlation level (e.g., 1%), to avoid numerical instability. Note that $\rho < 0$ is incompatible with the single-factor framework, even if a correlation matrix with all off-diagonal elements greater than $-(N-1)^{-1}$ is positive definite (Fig. 6).

7. INTRINSIC SIMPLICITY OF THE INTRINSIC VALUE

Using the Gaussian copula expression for the conditional default probabilities, we can simplify the intrinsic value calculations further, to
get the following closed form solution:

\[
\int_{-\infty}^{x_T^*} D\phi(x)\,dx + \int_{x_T^*}^{\infty} \mathbb{E}[L(t)|\lambda]\phi(x)\,dx \\
= D\Phi(x_T^*) + \int_{x_T^*}^{\infty} \sum_i (1 - R_i) N_i \Phi\left( \frac{\Phi^{-1}(p_i(t))}{\sqrt{1 - \rho}} - \sqrt{\frac{\rho}{1 - \rho}}x \right) \phi(x)\,dx \\
= D\Phi(x_T^*) + \sum_i (1 - R_i) N_i \left[ p_i(t) - \Phi_2(\Phi^{-1}(p_i(t)), x_T^*; \sqrt{\rho}) \right] \tag{11}
\]

where \( \Phi_2(x, y; \rho) \) denotes the bivariate normal distribution function with correlation \( \rho \). The last line follows from the properties of Gaussian integrals (see e.g., Andersen & Sidenius, 2005). To optimize the computational speed of the model it may be worth writing a tailored bivariate normal routine that takes a vector of the first argument along with constant second argument.
and (positive) correlation, since this function will typically be called a large number of times in a given pricing call.

8. TIME STABILITY OF THE TIME VALUE

As illustrated in Fig. 6, the time value function associated with a detachment point is nearly invariant in time, once centered around the critical value of the common factor, $x^*_n$. As such, it might have been more appropriately called “volatility value” or “convexity value,” since it arises from the interaction of the negative convexity of the loss cap with the conditional volatility of the portfolio loss, given the common factor. We are relying on the time stability as a computational heuristic, which has proven remarkably resilient in practical computations, even if the exact conditions for its approximate validity are hard to establish. It is perhaps worth noting in this context that the popular “Large Pool Model” approximation described in

Fig. 4. The Time Value and Intrinsic Value Components of the 0–7% Tranche Loss of a 125 Name Investment Grade Portfolio for Different Quarterly Time Slices as a Function of the Common Factor Value with the Maturity Axis Suppressed.
the base correlation paper by McGinty and Ahluwalia (2004) (and
foreshadowed by Vasicek (1991)) effectively assumes that the time value is
identically equal to zero (and thus trivially satisfying time stability). For
typical transaction and market parameters, the contributions to the final
tranche value from the time values tend to be relatively small compared to
the corresponding intrinsic value contributions, thus making the overall
calculation fairly robust to the performance of this approximation.

We estimate the time value and its contribution to the single name spread
and default sensitivities by a slightly modified version of Hull and White’s
(2004) bucketing algorithm, as explained further below, for a set of Gauss–
Laguerre quadrature points in each direction from $x^*_t$ for the common factor
at maturity. We numerically integrate these time values and their derivatives
against the Gaussian density for each calculation date $t$, taking care to re-
center the quadrature points around the corresponding values of $x^*_t$. We use
Gauss–Laguerre quadrature in each direction to integrate the smooth
function against a Gaussian density on each half-line around the kink in the

Fig. 5. The Time Value Component of the 0–7% Tranche Loss of a 125 Name
Investment Grade Portfolio for Different Quarterly Time Slices as a Function of the
Common Factor Value with the Maturity Axis Suppressed.
time value function at $X^*_t$. We have found 11 points in each direction to be satisfactory, generally producing breakeven premiums to well within 1 basis point of accuracy, and providing valuation and single name spread sensitivities in less than a second of CPU time for a typical mezzanine tranche of a 125 name index portfolio, not counting the time taken to load and strip the single name curves.

9. THE TIME VALUE COMPUTATION

We compute the conditional time value for each value of the state variable $x$ under consideration by dividing the positive part of the conditional loss distribution into a set of ranges that we will refer to as “buckets.” Specifically, we have a zero loss point, followed by equally spaced boundaries for portfolio losses up to a point slightly above the detachment

Fig. 6. The Time Value Component of the 0–7% Tranche Loss of a 125 Name Investment Grade Portfolio for Different Quarterly Time Slices as a Function of the Difference of the Common Factor Value and the Critical Level of the Common Factor for Each Maturity, with the Maturity Axis Suppressed.
point. The last loss bucket is a “trapping state” that will catch the probability of all losses exceeding the detachment point. We keep track of two quantities per bucket: the probability of the portfolio loss being in the bucket, and the expected portfolio loss given that the loss is in the bucket. The initial conditions are set as the entire probability mass being located at the zero loss point and the expected loss in each bucket being equal to the bucket’s mid-point. We then proceed to iterate over the names in the portfolio. For each name, we iterate over the loss buckets, and from those “source” buckets that already have attained positive probability, we re-allocate the probability of a default to the bucket that contains the sum of the name’s loss given default (LGD) and the expected loss in the source bucket. Once we have gone through all the names, we compute the conditional time value of the tranche as the capped expected portfolio loss minus the expected capped portfolio loss, based on the resulting conditional loss distribution.

An important feature of the algorithm is that we can compute a full set of single name spread sensitivities at relatively little extra cost. This is done by passing in a set of perturbed conditional default probabilities along with the original ones, corresponding to the effect of the desired spread perturbations, which are not restricted to be small. From the fully built-up conditional loss distribution, that is after having iterated over all names in the portfolio, we calculate perturbed values of the expected capped loss by treating the losses as being constrained to lie exactly on the points defined by the expected value of each loss bucket. Each name’s loss will in turn be de-convolved and re-convolved with the remaining loss distribution, where we split the loss probability between neighboring points when the distance between loss points is not commensurate with the name’s LGD. This requires the solution of a triangular and highly banded system of simultaneous linear equations for each name, which is usually extremely fast and accurate. A certain amount of care is required to avoid problems with numerical instability arising from the transition matrix being near-singular when the conditional single name default probability is close to 1.

REFERENCES


AN EMPIRICAL STUDY OF PRICING AND HEDGING COLLATERALIZED DEBT OBLIGATION (CDO)

Lijuan Cao, Zhang Jingqing, Lim Kian Guan and Zhonghui Zhao

ABSTRACT

This paper studies the pricing of collateralized debt obligation (CDO) using Monte Carlo and analytic methods. Both methods are developed within the framework of the reduced form model. One-factor Gaussian Copula is used for treating default correlations amongst the collateral portfolio. Based on the two methods, the portfolio loss, the expected loss in each CDO tranche, tranche spread, and the default delta sensitivity are analyzed with respect to different parameters such as maturity, default correlation, default intensity or hazard rate, and recovery rate. We provide a careful study of the effects of different parametric impact. Our results show that Monte Carlo method is slow and not robust in the calculation of default delta sensitivity. The analytic approach has comparative advantages for pricing CDO. We also employ empirical data to investigate the implied default correlation and base correlation of
the CDO. The implication of extending the analytical approach to incorporating Levy processes is also discussed.

1. INTRODUCTION TO COLLATERALIZED DEBT OBLIGATION

In recent years, due to the burgeoning credit derivatives market, there has been much research work on the collaterized debt obligation (CDO). A CDO is an asset-backed security whose payment depends on the collateral portfolio. There are different types of CDOs. A CDO whose collateral is made up of cash assets such as corporate bonds or loans is called cash CDO, while a CDO whose collateral is made up of credit default swaps is called a synthetic CDO. The structure of a CDO consists of partitions of the collateral portfolio into different tranches of increasing seniority. The CDO in effect transfers credit risk from the portfolio holder to investors. Investors of CDO are called protection sellers, while the issuer of CDO is called protection buyer. A particular tranche of a CDO is defined by its lower and an upper attachment point. The tranche with a lower attachment point $L$ and a higher attachment point $H$ will bear all the losses in the collateral portfolio in excess of $L$ and up to $H$ percent of the initial value of the portfolio. The portfolio loss is absorbed in ascending order of tranches, starting with the equity tranche, then the mezzanine tranche, and eventually the senior tranche.

As compensation for taking potential loss, the protection seller receives a periodic premium payment from the issuer of CDO until the maturity of the CDO or at the time when the tranche is expended through loss. The premium is paid from the interest income of the collateral portfolio. Interest is distributed to the tranches starting with the senior tranche, then the mezzanine tranche, and eventually the equity tranche. As the equity tranche absorbs the first layer of loss, the premium of this tranche is the largest among all the tranches. An example of CDO is illustrated in Fig. 1, where the portfolio is composed of 100 loans.

Each loan has $10 million notional amount. The equity tranche absorbs the first losses within $[0\%, \ 3\%]$ of the initial portfolio notional amount. The mezzanine tranche absorbs losses within $[3\%, \ 14\%]$. The senior tranche absorbs the remaining loss within $[14\%, \ 100\%]$. The premium of the tranches is paid as a percentage of the outstanding notional amounts of the corresponding tranches. For example, if there is a 1% loss in the
portfolio, the 3% portfolio value of the equity tranche is then reduced to 2% due to the loss. This amounts to 1/3 or a tranche loss of 33.3% in value. Consequently, the equity tranche pays only the pre-determined interest rate on a remaining 66.7% of tranche capital. For practical details of a CDO, see for example Elizalde (2004). Finger (2004), as well as Bluhm and Overbeck (2004), discusses the standard pricing model framework for synthetic CDO and some of the outstanding implementation and application issues.

The major risk in a CDO is default risk of the entities of the portfolio collateral. Such default risk can be modeled in two primary types of models that describe default processes in the credit risk literature: structural models and reduced form models. Structural models determine the time of default using the evolution of firms’ structural variables such as asset and debt values. Reduced form models determine the default process as a stochastic Poisson process with random default intensity. Empirically, the results in the literature show that the structural models under-predict the default probability while the reduced form models could predict the default process well.

More specifically, the problem of pricing CDO is equivalent to determining the premium of each tranche. There are three important components in the pricing of CDO: modeling credit risk, handling default correlations among collateral portfolio, and calculating the portfolio loss. The last two components are not common to simpler credit derivatives such as single-name credit default swaps. For an understanding of the

---

Fig. 1. Illustration of a CDO Structure.
background of default correlation and portfolio loss in the context of a CDO, we provide a brief summary of the literature in Table 1 where existing literature is surveyed in terms of the methodology for calculating default correlation and portfolio loss, and whether default delta sensitivity is studied or not.

Most of the existing studies used Copula to treat default correlation, except for Duffie and Garleanu (2001) where default correlation is treated by using dependent default intensity. Based on Copula, different analytic methods are proposed for calculating the portfolio loss. Most of the works demonstrated that the analytic methodologies constitute a powerful tool for evaluating CDO.

In the framework of reduced form models, there are basically three methodologies for treating default correlation among multiple assets in the collateral portfolio: conditionally independent default model of Duffie and Garleanu (2001), contagion model of Jarrow and Yu (2001), and Copula method by Li (2000). Conditionally independent default model handles default correlation by simulating correlated default intensities based on a common set of state variables. The major disadvantage of the conditionally independent default model is that the correlation generated by the model is often too small in empirical data with high default correlation. In the contagion model, the default of one firm triggers the default of the other related firms, and the default times tend to be concentrated in certain time periods. The disadvantage of the contagion method is that it is difficult to calibrate the parameters of the model. The resulting model is thus hard to implement. The other method of treating default correlation is the Copula method. Using Copula in default correlation modeling is originally proposed by Li (2000). The Copula function is actually a correlated multivariate function defined by the marginal default probability distribution. A variety of functions can be used as Copula, such as $t$-Student and Gaussian. The Copula method is simple and easy to implement.

Apart from its simplicity, another advantage of using Copula is that the portfolio loss in a CDO can be analytically computed without relying on Monte Carlo simulations that can be computationally intensive and time-consuming. Andersen, Sidenius, and Basu (2003) describe an analytic method of calculating the portfolio loss based on a one-factor Gaussian Copula. Gibson (2004) describes the analytic method more explicitly. Besides the portfolio loss, Andersen, Sidenius, and Basu (2003) also propose an analytic method for calculating the default delta sensitivity.

This paper studies the pricing and hedging of CDO by comparing the analytic method of Gibson (2004) with the Monte Carlo method. There has
Table 1. A Summary of Literature Review in CDO.

<table>
<thead>
<tr>
<th>Publications</th>
<th>Default Correlation Methods</th>
<th>Portfolio Loss Methods</th>
<th>Default Delta Sensitivity</th>
<th>Conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Andersen, Sidenius, and Basu (2003)</td>
<td>One-factor Copula</td>
<td>Analytical approach by a recursion-based probability calculation</td>
<td>Uses Brute-force method and an analytical method</td>
<td>Proposed a number of techniques to improve the efficiency with which prices and hedge parameters can be calculated for credit basket derivatives.</td>
</tr>
<tr>
<td>Gibson (2004)</td>
<td>Same as above</td>
<td>Same as above</td>
<td>Brute-force method</td>
<td>The value of the senior tranche decreases as correlation increases. In contrast, the equity tranche value increases as default correlation increases. CDO tranches are sensitive to the business cycle.</td>
</tr>
<tr>
<td>Laurent and Gregory (2003)</td>
<td>One-factor Copula</td>
<td>Analytic approach based on Fourier method</td>
<td>N.A.</td>
<td>Proposed an analytic approach based on Fourier method to calculate the conditional loss distribution on a portfolio as a convolution of the conditional loss distributions of each entity in the portfolio.</td>
</tr>
<tr>
<td>Burtschell, Gregory, and Laurent (2005)</td>
<td>Copulas</td>
<td>Laurent and Gregory’s analytical approach</td>
<td>N.A.</td>
<td>Compare some popular Copula functions such as Gaussian Copula model, stochastic correlation extension to Gaussian Copula, Student’s t Copula model, double t factor model, clayton and Marshall–Olkin Copula.</td>
</tr>
<tr>
<td>Peixoto (2004)</td>
<td>Copula</td>
<td>Monte Carlo and analytical methods</td>
<td>N.A.</td>
<td>Compare Monte Carlo and analytical method in the pricing of CDO. Both prices are within one standard deviation.</td>
</tr>
<tr>
<td>Mina and Stern (2003)</td>
<td>One-factor Copula</td>
<td>Analytical approach based on Fourier transform method</td>
<td>Analytic method</td>
<td>Senior Tranche price depends on the best while Equity Tranche on the worst names in a portfolio. Mezzanine behavior varies over time. Loan-equivalent hedges depend on the entire portfolio. Equity Tranche value rises</td>
</tr>
<tr>
<td>Publicationsa</td>
<td>Default Correlation Methods</td>
<td>Portfolio Loss Methods</td>
<td>Default Delta Sensitivity</td>
<td>Conclusions</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------------------------</td>
<td>------------------------</td>
<td>----------------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Chen and Zhang (2003)</td>
<td>One-factor Copula</td>
<td>Analytical approach based on Fourier transform method</td>
<td>N.A.</td>
<td>while Senior Tranche value drops when correlation increases. FFT/FI generates loss distributions more accurate than those by the Monte Carlo simulations.</td>
</tr>
<tr>
<td>Duffie and Garleanu (2001)</td>
<td>Dependent default intensity</td>
<td>Monte Carlo</td>
<td>N.A.</td>
<td>Illustrated the effects of correlation and prioritization for the market valuation, diversity score and risk of CDO in a simple jump-diffusion setting for correlated default intensities.</td>
</tr>
<tr>
<td>Hull and White (2004)</td>
<td>One-factor Copula</td>
<td>Two analytic methods: recursive approach and iterative numerical procedure</td>
<td>N.A.</td>
<td>The procedures are attractive alternatives to Monte Carlo simulation and have advantages over the fast Fourier transform approach. Implied correlations are typically not the same for all tranches.</td>
</tr>
<tr>
<td>Andersen and Sidenius (2004)</td>
<td>One-factor Copula</td>
<td>Analytical recursive method</td>
<td>Analytic method</td>
<td>This paper extends the standard Gaussian Copula model by using random recovery rates and random systematic factor loadings. It is capable of producing correlation skews similar to those observed in the market.</td>
</tr>
<tr>
<td>Author(s)</td>
<td>Method Description</td>
<td>Approach</td>
<td>Notes</td>
<td></td>
</tr>
<tr>
<td>---------------------------------</td>
<td>-------------------------------------------------------------------------------------</td>
<td>----------------------</td>
<td>----------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Kalemanova et al. (2005)</td>
<td>Copula with normal inverse Gaussian (NIG) distribution</td>
<td>Analytic approach</td>
<td>Proposed a modification of the LHP model replacing the Student's $t$ distribution with the NIG. The employment of the NIG distribution does not only speed up the computation time significantly but also brings more flexibility into the dependence structure.</td>
<td></td>
</tr>
<tr>
<td>Bluhm and Overbeck (2004)</td>
<td>One-factor Copula</td>
<td>Analytic approach</td>
<td>Analytic techniques constitute a powerful tool for the evaluation of CDO.</td>
<td></td>
</tr>
<tr>
<td>Morokoff (2003)</td>
<td>Copula</td>
<td>Monte Carlo</td>
<td>This paper describes a multiple-time step simulation approach that tracks cash flows over the life of a CDO deal to determine the risk characteristics of CDO tranches.</td>
<td></td>
</tr>
<tr>
<td>Hurd and Kuznetsov (2005)</td>
<td>N.A.</td>
<td>Affine Markov Chain</td>
<td>Combined a continuous time Markov Chain with an independent set of affine processes that yield a flexible framework for which computations are very efficient.</td>
<td></td>
</tr>
</tbody>
</table>

*aThe publication name can be found in the Reference section of this paper.*
been few studies performing such comparisons, and it is important to be able to decide which models to use in practice. The CDO data employed in Peixoto (2004) are used in the empirical investigation in this paper. In addition, this paper provides a careful study of the effects of different parametric impact. The portfolio loss, the expected loss allocated to each tranche, the tranche spread, and the default delta sensitivity are analyzed with respect to different parameters such as maturity, default correlation, default intensity or hazard rate, and recovery rate. In the current literature, the default delta sensitivity is discussed only in Andersen, Sidenius, and Basu (2003), Gibson (2004), Mina and Stern (2003), and Andersen and Sidenius (2004). By providing a more thorough study of the delta sensitivity with respect to some key parameters, this study will help in the hedging performance of CDO. Furthermore, the implied default correlation and base correlation are also empirically investigated.

The remainder of the paper is organized as follows. Section 2 describes the methodology of pricing CDO using the analytic method and the Monte Carlo method. The methodology of calculating default hedge ratio is described in Section 3. Section 4 presents the empirical results. The implication of extending the analytical approach to incorporating Levy processes is also discussed. Section 5 contains the conclusions.

2. METHODOLOGY OF PRICING CDO

The model is set up in a filtered probability space \((\Omega, F, (F_t)_{t \geq 0}, P)\), where \(P\) is a pre-specified martingale measure. The filtration \((F_t)_{t \geq 0}\) satisfies the usual conditions and the initial filtration \(F_0\) is trivial. There is also a finite time horizon \(T\) with \(F = F_T\). The remaining notations used in this paper are described as follows.

- \(I_k\): Notional amount for asset \(k, k = 1, 2, \ldots, K\)
- \(R_k\): Recovery rate for asset \(k\)
- \(\lambda_k\): Default intensity for asset \(k\)
- \(\tau_k\): Default time for asset \(k\)
- \(T_i\): The payment date in CDO, \(i = 1, 2, \ldots, n\). We assume that for a standard CDO, all tranches are paid interest at the same time points
- \(l_i\): The total amount of loss in the portfolio at time \(T_i\)
- \(e_i\): The total amount of loss allocated to the tranche at \(T_i\)
- \(B_i\): The price of a default-free zero coupon bond with maturity \(T_i\) and face value of $1 at present time
The pricing of a CDO consists of pricing the single tranches that make up the CDO structure. For a single tranche with attachment points \([L, H]\), the cash flows can be described as follows. The seller of a CDO pays a periodic coupon to the investor of the CDO at each payment date \(T_i, i = 1, 2, \ldots, n\). The coupon paid at \(T_i\) for a tranche is calculated based on the outstanding notional amount in that tranche. Obviously, the initial dollar value of notional amount for the tranche is equal to \(H/C_0 \times L\). When default occurs in the portfolio and the portfolio loss exceeds \(L\), the investor of this CDO tranche has to pay to the seller of CDO the amount of loss in excess of \(L\). The loss between \([T_{i-1}, T_i]\) is assumed to be paid at \(T_i\). The maximal value of the total amount payable by the investor is equal to \(H/C_0 \times L\). Thus, the pricing of a tranche consists of calculating the premium leg corresponding to the payment by the seller and the default leg corresponding to the payment by the investor when there is default in excess of \(L\).

In the reduced form model of Li (2000), the risk-neural default probability of an asset at \(T_i\) is calculated by

\[
p(\tau_k \leq T_i) = 1 - e^{-\int_0^{T_i} \dot{\lambda}_k(u)du} \tag{1}
\]

If the default time \(\tau_k\) in each \(k\)th asset is known, the portfolio loss \(l_i\) at \(T_i\) can be calculated by

\[
l_i = \sum_{k=1}^{K} I_k(1 - R_k)1_{\{\tau_k \leq T_i\}}, \quad i = 1, 2, \ldots, n \quad \text{and} \quad 1_{\{\tau \leq t\}} = \begin{cases} 1 & \tau \leq t \\ 0 & \tau > t \end{cases}
\]

\[
(2)
\]

Given \(l_i\), the total amount of dollar loss allocated to the single tranche of \([L, H]\) at \(T_i\) is equal to

\[
e_i = \text{Max}(\text{Min}(l_i, H) - L, 0), \quad i = 1, 2, \ldots, n \tag{3}
\]

Thus the present value of the default leg (denoted as DL) is equal to the sum of the present values of the expected values of the loss paid by the investor of tranche to the seller of tranche at the various \(T_i\)’s, that is calculated by

\[
E(DL) = \sum_{i=0}^{n} B_i(E(e_i) - E(e_{i-1})) \tag{4}
\]

where by definition \(E(e_{-1}) = 0\). The expected loss between \([T_{i-1}, T_i]\) is equal to \((E(e_i) - E(e_{i-1}))\), which is assumed to be paid at \(T_i\). Obviously \(E(e_{0}) = 0\) as
well if the earliest first loss payment is at $T_1$. In this case, the summation indices in $E(DL)$ could also be written without loss of generality to start at $i = 1$ instead of 0.

Let $s$ denote the tranche spread, which is the annualized interest charge or coupon rate on the tranche. The expected value of the premium leg (denoted as $PL$) can be expressed as the sum of the present values of the expected values of the amount paid by the seller of the tranche to the investor, which is calculated by

$$E(PL) = \sum_{i=1}^{n} s \Delta t_i B_i (H - L - E(e_i))$$

where $\Delta t_i = T_i - T_{i-1}$ is denoted as a fraction of a year. $H - L - E(e_i)$ denotes the expected value of outstanding notional amount at $T_i$.

The equilibrium pricing of the tranche under risk-neutrality implies that $s$ is found by setting $E(DL) = E(PL)$. Thus,

$$s = \frac{\sum_{i=0}^{n} B_i (E(e_i) - E(e_{i-1}))}{\sum_{i=1}^{n} \Delta t_i B_i (H - L - E(e_i))}$$

From Eqs. (4)–(6), it can be observed that the crucial task of pricing is to calculate the expected value of the tranche loss $E(e_i)$ for each $T_i$. The methodologies of calculating $E(e_i)$ thus obtaining $s$ in the analytic method of Gibson (2004) and Monte Carlo method are described as below.

2.1. Analytic Method

The analytic method uses a continuous state variable $X_k$ taking values in $(-\infty, \infty)$ to represent the default status of an asset $k$. When $X_k$ approaches $-\infty$ from the right, the probability of default approaches 0. When $X_k$ approaches $+\infty$ from the left, the probability of default approaches 1. The (cumulative) probability distribution function of $X_k$ is the (unconditional) probability of default. A single (common) factor model of $X_k$ consists of a common factor $M$ and an individual factor $Z_k$.

$$X_k = a_k M + \sqrt{1 - a_k^2} Z_k, \quad k = 1, 2, \ldots, K$$

where $a_k$ represents the fraction of the common factor relative to $Z_k$. The value of $a_k$ is between $[0, 1]$ and the variances of the $X_k$’s are ones. $a_k a_j$ thus denotes the correlation of $X_j$ and $X_k$. It represents the default correlation
between asset \( k \) and \( j \). For simplicity, \( X_k, M, \) and \( Z_k \) are assumed to follow standard normal distributions. \( M \) and \( Z_k \) are independent variables. Eq. (7) is often termed the one-factor Gaussian Copula.

The probability distribution of \( X_k \) is equal to the risk-neutral probability of default in Eq. (1). Thus, the conditional default probability \( p(\tau_k \leq T_i | M) \) at \( T_i \) for asset \( k \) can be calculated by

\[
p(\tau_k \leq T_i | M) = N \left( \frac{N^{-1}(p(\tau_k \leq T_i)) - a_k M}{\sqrt{1 - a_k^2}} \right)
\]

Let \( p^K_i(k|M) \) denote the conditional probability of \( k \) defaults up to \( T_i \) in a reference portfolio of size \( K \). The analytic method calculates \( p^K_i(k|M) \) using the following recursive algorithm:

\[
p^K_{i+1}(0|M) = p^K_i(0|M)(1 - p(\tau_{k+1} \leq T_i | M))
\]

\[
p^K_{i+1}(k|M) = p^K_i(k|M)(1 - p(\tau_{k+1} \leq T_i | M)) + p^K_i(k - 1|M)p(\tau_{k+1} \leq T_i | M)
\]

and

\[
p^K_{i+1}(k + 1|M) = p^K_i(k|M)p(\tau_{k+1} \leq T_i | M) \quad \text{for} \quad k = 1, \ldots, K
\]

For \( K = 0 \), \( p^0(0|M) = 1 \). After computing \( p^K_i(k|M) \), for \( k = 0, 1, \ldots, K \), the unconditional portfolio number of defaults distribution \( p^K_i(k) \) is calculated by

\[
p^K_i(k) = \int_{-\infty}^{\infty} p^K_i(k|M)g(M)dM
\]

where \( g(\cdot) \) is the probability density function of \( M \). The integral can be computed using numerical integration. Note that in the above computation, conditional on \( M \), the probabilities of default for the different assets in the portfolio are independent. The conditional probabilities of default for the different asset \( k \), \( p(\tau_{k+1} \leq T_i | M) \), are different as the assets’ characteristics are different.

Let \( p(l_i) \) be the portfolio loss probability. Suppose this is a discrete distribution, then we can write the expected tranche loss as

\[
E(e_i) = \sum_{l_i \leq H} (l_i - L)p(l_i) + \sum_{l_i \geq H} (H - L)p(l_i)
\]
From \( E(e_i) \), Eqs. (4)–(6) are then used to calculate the spread \( s \) of each tranche.

Suppose the collateral portfolio is a large homogeneous portfolio made up of small similar assets. Homogeneity is with respect to the terms of \( I_k, R_k, \) and \( \lambda_k \), resulting in the same \( I, R, \) and \( \lambda \). Then, instead of the above algorithm, \( p^K_i(k|M) \) can be calculated simply as a binomial function:

\[
p^K_i(k|M) = \binom{K}{k} \times \left( \frac{N^{-1}(p(\tau \leq T_i)) - aM}{\sqrt{1 - a^2}} \right)^k \times \left( 1 - N^{-1}(p(\tau \leq T_i)) - aM \right)^{K-k}
\]

where all \( a_i \)'s equal \( a \), and \( \tau \) denotes the default time of any one of the assets. In this case, we treat each credit entity in the portfolio as identical. The unconditional portfolio number of defaults distribution \( p^K_i(k) \) is then similarly computed using Eq. (12). Under homogeneity, the portfolio loss \( l_i \) can be simplified from Eq. (2) to \( I(1-R) \) times the number of defaults. Hence in this case, the probability distribution of portfolio loss \( p(l_i) \) is represented by the distribution of the number of defaults. The expected portfolio loss is then \( I(1-R) \) times the expected number of default by a certain time.

### 2.2. Monte Carlo Method

The Monte Carlo method takes into account the default correlation using the Copula function. The following Gaussian Copula is most commonly used.

\[
C(u_1, u_2, \ldots, u_K) = N(N^{-1}(u_1), N^{-1}(u_2), \ldots, N^{-1}(u_K))
\]

where \( u_k \) is equal to \( p(\tau_k \leq T_i) \) in Eq. (1), and \( N(v_1, v_2, \ldots, v_K) \) on the right side of Eq. (15) denotes a multivariate normal probability distribution function with mean zero and correlation matrix \( \rho_{i,j} \), where \( i, j = 1, 2, \ldots, K \). As in typical applications, we employ a constant correlation matrix with a single parameter \( \rho \in [0,1] \) for all \( \rho_{i,j} \). If Eq. (15) is specialized to the single-factor analytical model, the value of \( \rho \) in Eq. (15) would be related to \( a_k \) in Eq. (7) by the following formulae: \( \rho_{k,j} = a_k a_j, k \neq j, \) and \( \rho_{k,k} = 1, k = j \).
Based on Eq. (15) with a constant correlation matrix, the Monte Carlo method calculates \( e_i \) using the following algorithm. Perform \( N \) number of simulations each of which takes the following steps:

1. Employ the multivariate normal distribution \( N(v_1, v_2, \ldots, v_K) \) to generate for a given \( \rho \), the \( K \) random variables. Calculate the default time of each asset by

\[
\tau_k = -\frac{\ln(1 - \Phi(v_k))}{\lambda_k}, \quad k = 1, 2, \ldots, K
\]  

(16)

where \( \Phi(\cdot) \) denotes the univariate standard normal probability distribution function.

2. From the computed \( \tau_k \) for \( k = 1, 2, \ldots, K \), we can determine if for asset \( k \), default has occurred by time \( T_i \), whether \( 1_{[\tau_k \leq T_i]} \) takes value 1 in the event of default by time \( T_i \), or 0 otherwise in the event of no-default.

Then we can calculate the portfolio loss \( l_i \) at each payment date \( T_i \), \( i = 1, 2, L \), \( n \) according to \( l_i = \sum_{k=1}^{K} I_k (1 - R_k) 1_{[\tau_k \leq T_i]} \) in Eq. (2).

3. Next calculate the tranche loss \( e_i \) at each payment date \( T_i \), \( i = 1, 2, L \), \( n \) according to Eq. (3).

4. Then calculate \( DL = \sum_{i=0}^{n} B_i (e_i - e_{i-1}) \) and \( Q \equiv PL/s = \sum_{i=1}^{n} \Delta t_i B_i (H - L - e_i) \).

Finally, the \( N \) number of simulations each involving terms DL and \( Q \) in step (4) are averaged to obtain the spread \( s \) by the following:

\[
s = \frac{\sum_{q=1}^{N} DL(q)}{\sum_{q=1}^{N} Q(q)}
\]  

(17)

where \( q \) denotes the \( q \)th simulation. The expected portfolio loss at each time \( T_i \) can also be computed by averaging across the portfolio loss values at each step (2). The number of simulations \( N \) may be 50,000 or less depending on the complication of the model and the allocated computing time.

### 3. METHODOLOGY OF CALCULATING DEFAULT DELTA SENSITIVITY

In Monte Carlo, the brute force method is used for calculating the sensitivity of the price of the \([L, H]\) CDO tranche to default intensity \( \lambda_k \). The approach is described as follows. First, \( \lambda_k \) for the \( k \)th asset is increased by a small
amount $\Delta \lambda_k$ to re-calculate price of the tranche. Second, the ratio of the price difference to $\Delta \lambda_k$ is calculated as the default delta sensitivity. The mathematical formula is written as

$$\frac{\partial V}{\partial \lambda_k} = \frac{V(\lambda_k + \Delta \lambda_k) - V(\lambda_k)}{\Delta \lambda_k} \quad (18)$$

and

$$V = \sum_{i=0}^{n} B_i (E(e_i) - E(e_{i-1})) - \sum_{i=1}^{n} s \Delta t_i B_i (H - L - E(e_i)) \quad (19)$$

where $V$ is the market value of the tranche to the CDO issuer or the protection buyer.

In the analytic method, the analytic methodology proposed in Andersen and Sidenius (2004) is used for calculating the default delta sensitivity. The methodology is described below.

From Eq. (19), it can be observed that

$$\frac{\partial V}{\partial \lambda_k} = \sum_{i=0}^{n} B_i \left( \frac{\partial E(e_i)}{\partial \lambda_k} - \frac{\partial E(e_{i-1})}{\partial \lambda_k} \right) - \sum_{i=1}^{n} s \Delta t_i B_i \left( H - L - \frac{\partial E(e_i)}{\partial \lambda_k} \right) \quad (20)$$

According to Eq. (13), $\frac{\partial E(e_i)}{\partial \lambda_k}$ is equal to

$$\frac{\partial E(e_i)}{\partial \lambda_k} = \sum_{l_i \leq H} (l_i - L) \frac{\partial p(l_i)}{\partial \lambda_k} + \sum_{l_i \geq H} (H - L) \frac{\partial p(l_i)}{\partial \lambda_k} \quad (21)$$

From Eq. (12), by assuming $\lambda_k$ and $M$ are independent, $\frac{\partial p(l_i)}{\partial \lambda_k}$ is calculated by

$$\frac{\partial p(l_i)}{\partial \lambda_k} = \int_{-\infty}^{\infty} \frac{\partial p_i^k(k|M)}{\partial \lambda_k} g(M) dM \quad (22)$$

where we have used the result that the probability distribution of portfolio loss $p(l_i)$ is equivalent to the unconditional probability distribution of the number of defaults. According to Eqs. (9)–(11), $\frac{\partial p_i^k(k|M)}{\partial \lambda_k}$ is calculated by

$$\frac{\partial p_i^k(0|M)}{\partial \lambda_k} = -p_i^{k-1}(0|M) \frac{\partial p(\tau_k \leq T_i|M)}{\partial \lambda_k} \quad (23)$$
\[ \frac{\partial p^K(k|M)}{\partial \lambda_k} = \frac{\partial p(\tau_k \leq T_i|M)}{\partial \lambda_k} \left( -p_i^{K-1}(k|M) + p_i^{K-1}(k - 1|M) \right) \] (24)

and

\[ \frac{\partial p^K(k|M)}{\partial \lambda_k} = \frac{\partial p(\tau_k \leq T_i|M)}{\partial \lambda_k} p_i^{K-1}(k - 1|M) \quad \text{for } k = 1, \ldots, K \] (25)

From Eqs. (1) and (8), we can derive

\[ \frac{\partial p(\tau_k \leq T_i|M)}{\partial \lambda_k} = e^{-0.5((N^{-1}(p(\tau_k \leq T_i))-a_k M)/\sqrt{1-a_k^2})^2} \frac{1}{\sqrt{1-a_k^2}} e^{0.5(N^{-1}(p(\tau_k \leq T_i))^2) T_i e^{-\lambda_k T_i}} \] (26)

From Eqs. (21)–(26), the default delta sensitivity in Eq. (20) can be calculated.

4. EMPIRICAL RESULTS

The CDO data studied in Peixoto (2004) are employed in our study. The collateral portfolio of CDO is composed of 100 loans each with equal face value. The maturity of the CDO is 5 years. The default intensity or the hazard rate of each loan is 0.03. The recovery rate of each asset is 0.4. The premium and default loss is paid quarterly. The risk-free interest rate is 5% with continuous compounding.

As illustrated in Table 2, the attachment points are as following: equity tranche, [0, 3%]; mezzanine tranche, [3%, 14%] and senior tranche, [14%, 100%]. The expected loss, the spread, and default delta sensitivity in each tranche are analyzed.

4.1. Expected Loss (EL)

In the first set of empirical results, the portfolio loss and the loss distributed in each tranche are analyzed with respect to different parameters.

Fig. 2(a) and (b) show the portfolio EL with different maturity \( T \) and default correlation \( \rho \) in the analytic and the Monte Carlo method. It can be observed that both methods give close results, while Monte Carlo method generates the EL surface that is smoother than that of the analytic method.
For a fixed value of $r$, the portfolio EL increases with maturity $T$ as more defaults are likely to happen at larger $T$. When $T$ is fixed, the portfolio EL increases is less sensitive to increase in $r$. The difference between the maximum and minimum values of the portfolio EL for different values of $r$ is within 20 basis points or 0.2% of portfolio value.

Fig. 3(a) and (b) show that the EL’s allocated to the equity tranche in the analytic and the Monte Carlo methods agree closely. Equity EL increases with $T$ for fixed $r$ due to the larger default probability at larger $T$. When $T$ is fixed, equity EL decreases with the increase of $r$. The result can be explained as follows. As $r$ increases, there is a higher probability that either many obligors default together, resulting in larger losses, or many do not default together resulting in smaller overall losses. The latter obviously has a more weighted impact on the equity tranche that takes the first loss, resulting in an overall lower expected loss.

Fig. 4(a) and (b) are produced from the analytical method and show the portfolio loss distribution of $r=0$ and 0.9 with $T=5$. For large losses in the range 0.19–0.58, it can be observed from Fig. 4(b) that the probability of occurrence in the case $r=0.9$ is much higher than that in the case $r=0$. For small losses in the range 0.08–0.11, it can be observed from Fig. 4(a) that the probability of occurrence in the case $r=0.9$ is much smaller than that in the case $r=0$. Thus, higher default correlation $r$ leads to higher chances of large portfolio loss and lower chances of small portfolio loss. Contagion effect or high default correlation is therefore risky from the point of view of the tranche buyers. The larger losses make the tranche with higher seniority suffer from more loss. For higher $r$, the lower probability of the smaller portfolio losses hitting mainly the equity tranche means also that expected loss on the latter is lower. This concurs with the result expressed in Fig. 3(a) and (b).

**Table 2.** The Characteristics of CDO Used in This Study.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maturity</td>
<td>5 years</td>
</tr>
<tr>
<td>Hazard rate</td>
<td>0.03</td>
</tr>
<tr>
<td>Recovery rate</td>
<td>0.4</td>
</tr>
<tr>
<td>Risk-free rate</td>
<td>5% with continuous compounding</td>
</tr>
<tr>
<td>Payment frequency</td>
<td>Quarterly payment</td>
</tr>
<tr>
<td>Equity tranche</td>
<td>[0, 3%]</td>
</tr>
<tr>
<td>Mezzanine tranche</td>
<td>[3%, 14%]</td>
</tr>
<tr>
<td>Senior tranche</td>
<td>[14%, 100%]</td>
</tr>
</tbody>
</table>
Fig. 2. The Portfolio Expected Loss (EL) with Different Maturity and Default Correlation in the Analytic Method (a) and Monte Carlo Method (b).
Fig. 3. The Equity Tranche Expected Loss (EL) with Different Maturity and Default Correlation in the Analytic Method (a) and Monte Carlo Method (b).
Fig. 4. Portfolio Loss Distribution in the Entire Loss Range (a) and Large Losses (b). Large Loss is Defined as Loss over 0.186. Two Cases of Default Correlation $\rho = 0$ and 0.9 are Considered. Time Horizon is $T = 5$ Years.

Fig. 5(a) and (b) illustrate the loss distributions of the equity and senior tranches respectively for cases $\rho = 0$ and 0.9. These are computed from the analytical method. Fig. 5(a) shows that in equity tranche the probability of losing the tranche notional amount for case $\rho = 0.9$ is less than that for case $\rho = 0$ once the loss amount gets into the non-trivial range above 0.01. This results in smaller equity EL for large $\rho$ as discussed in the last two sets of figures. Fig. 5(b) shows that in the senior tranche the probability of large loss is much higher in the case $\rho = 0.9$ compared to the case $\rho = 0$. 
Fig. 6(a) and (b) show EL allocated to the mezzanine tranche in the analytic and the Monte Carlo methods. Mezzanine EL increases as $T$ increases. For long maturity $T = [4, 5]$, mezzanine EL decreases with the increase of $\rho$, having the same characteristics as that of equity. However, for short maturity $T = [1, 4]$, mezzanine EL firstly increases and then decreases with the increase of $\rho$. The latter decrease is due to the rapid increase in the probability of loss for senior tranche at high levels of $\rho$, in which the loss impact on the mezzanine tranche would be reduced. In effect, larger values of $\rho$ reduce the chance of absorbing loss in the mezzanine tranche.

Fig. 7(a) and (b) show EL allocated to the senior tranche in the analytic and the Monte Carlo methods. The largest EL occurs for higher $T$ and $\rho$. 

Fig. 5. The Portfolio Loss Distribution of Equity Tranche (a) and Senior Tranche (b). Two Cases of Default Correlation $\rho = 0$ and 0.9 are Considered.
Fig. 6. The Mezzanine Tranche Expected Loss (EL) with Different Maturity and Default Correlation in the Analytic Method (a) and Monte Carlo Method (b).
Fig. 7. The Senior Tranche Expected Loss (EL) with Different Maturity and Default Correlation in the Analytic Method (a) and Monte Carlo Method (b).
values, for example when $T = 5$ and $\rho = 0.9$. Senior tranche does not absorb loss at smaller values of $T$ and $\rho$. This is consistent with the characteristic of senior tranche that it is the last tranche to take losses in the portfolio. When either $\rho$ or $T$ increases, senior EL increases.

Table 3 summarizes the expected losses of the analytic and the Monte Carlo methods for the case of $T = 5$. It can be observed that the difference in expected losses between the two methods is within 20 basis points.

### 4.2. Tranche Spread and Default Delta Sensitivity

Tranche spread and default delta sensitivity in each tranche are analyzed in the second set of empirical results.

Fig. 8(a) and (b) show the spread of equity tranche with different $T$ and $\rho$ in the analytic and Monte Carlo methods. For a fixed value of $T$ equity spread decreases with the increase of $\rho$, due to the decreased expected loss. When $\rho$ is fixed equity spread increase is not sensitive to increase in $T$.

Fig. 9(a) and (b) show the default delta sensitivity in equity tranche for the analytic and the Monte Carlo methods, respectively. For calculating sensitivity, the spreads of equity, mezzanine, and senior tranches are arbitrarily set as 1,000, 500, and 1 bp, respectively. In the Monte Carlo method, $\Delta \lambda_k$ is set as 10 bp. The Monte Carlo results do not converge as fast as the analytical ones in calculating sensitivity. Thus Monte Carlo method is not robust in the calculation of default delta sensitivity. Fig. 9(a) shows that the sensitivity is highest at $T = 1$ and $\rho = 0$ in the equity tranche. This is consistent with the fact that equity is mostly sensitive to small losses occurring at an early time. The results show that when $T$ or $\rho$ is small, the value of delta sensitivity decreases with increase of $\rho$ or $T$. When $T$ or $\rho$ is large the value of delta sensitivity increases initially and then later decreases with increase of $\rho$ or $T$.

Fig. 10(a) and (b) show the mezzanine tranche spread at different $T$ and $\rho$. It can be observed that there the largest spread occurs at $T = 5$ and $\rho = 0$, corresponding to the maximum values of mezzanine EL in Fig. 6(a) and (b). When maturity is large where $T = [4, 5]$, the mezzanine spread decreases with the increase of $\rho$, having the same characteristic of the equity tranche. When maturity is small where $T = [1, 4]$, the mezzanine spread increases initially and then decreases with the increase of $\rho$, showing the same characteristic as the mezzanine EL in Fig. 6(a) and (b). For a fixed value of $\rho$, mezzanine spread increases with $T$.
Table 3. Portfolio Expected Loss (EL), Equity EL, Mezzanine EL, and Senior EL with Different Values of $\rho$ for $T=5$.

<table>
<thead>
<tr>
<th>Default Correlation $\rho$</th>
<th>Portfolio EL</th>
<th>Equity EL</th>
<th>Mezzanine EL</th>
<th>Senior EL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Analytic</td>
<td>Monte Carlo</td>
<td>Difference</td>
<td>Analytic</td>
</tr>
<tr>
<td>0</td>
<td>0.0837</td>
<td>0.0836</td>
<td>0.0001</td>
<td>0.0300</td>
</tr>
<tr>
<td>0.1</td>
<td>0.0838</td>
<td>0.0831</td>
<td>0.0008</td>
<td>0.0290</td>
</tr>
<tr>
<td>0.2</td>
<td>0.0839</td>
<td>0.0829</td>
<td>0.0010</td>
<td>0.0270</td>
</tr>
<tr>
<td>0.3</td>
<td>0.0830</td>
<td>0.0811</td>
<td>0.0019</td>
<td>0.0248</td>
</tr>
<tr>
<td>0.4</td>
<td>0.0835</td>
<td>0.0839</td>
<td>-0.0003</td>
<td>0.0225</td>
</tr>
<tr>
<td>0.5</td>
<td>0.0857</td>
<td>0.0876</td>
<td>-0.0019</td>
<td>0.0201</td>
</tr>
<tr>
<td>0.6</td>
<td>0.0829</td>
<td>0.0812</td>
<td>0.0016</td>
<td>0.0178</td>
</tr>
<tr>
<td>0.7</td>
<td>0.0851</td>
<td>0.0859</td>
<td>-0.0009</td>
<td>0.0153</td>
</tr>
<tr>
<td>0.8</td>
<td>0.0833</td>
<td>0.0828</td>
<td>0.0005</td>
<td>0.0127</td>
</tr>
<tr>
<td>0.9</td>
<td>0.0836</td>
<td>0.0831</td>
<td>0.0005</td>
<td>0.0300</td>
</tr>
</tbody>
</table>
Fig. 8. The Equity Tranche Spread with Different Maturity and Default Correlation in the Analytic Method (a) and Monte Carlo Method (b).
Fig. 9. The Delta Sensitivity of Equity Tranche with Different Maturity and Default Correlation in the Analytic Method (a) and Monte Carlo Method (b).
Fig. 10. The Mezzanine Tranche Spread with Different Maturity and Default Correlation in the Analytic Method (a) and Monte Carlo Method (b).
Fig. 11(a) and (b) show the default delta sensitivity in the mezzanine tranche for the analytic and the Monte Carlo methods. The largest value of sensitivity occurs at $T = 5$ and $\rho = 0$. The same relationship between the mezzanine spread (EL) and $\rho$ can be applied here. For maturity in interval $T = [4, 5]$, the mezzanine sensitivity decreases with the increases of $\rho$. When $T$ is small, the mezzanine sensitivity increases initially and then decreases with the increase of $\rho$. For a fixed value of $\rho$, mezzanine default delta sensitivity increases with $T$.

Figs. 12(a), (b), 13(a), and (b), respectively, illustrates the spread and the default delta sensitivity of the senior tranche in the analytic and the Monte Carlo methods. The largest values of spread and default delta sensitivity for the senior tranche occur at $T = 5$ and $\rho = 0.9$, corresponding to the maximum values of the senior EL. When $T$ or $\rho$ is fixed, both the spread and sensitivity increase with the increase of $\rho$ or $T$.

Table 4 compares the spread difference between the analytic and the Monte Carlo methods for the case of $T = 5$ and across various values of $\rho$. It can be observed that the spread difference between the two methods falls within 30 bp.

Table 5 compares the difference of the default delta sensitivity values between the analytic and the Monte Carlo methods for the case of $T = 5$ and across various values of $\rho$. It can be observed that the difference between the analytic and the Monte Carlo results is generally small at approximately less than 10 bp. However, but for a few cases, the difference is larger as the Monte Carlo result is sensitive to the choice of $\Delta \lambda$.

The spread and default delta sensitivity are further examined by using different values of recovery rates $R$ and default intensity or hazard rate $\lambda$, with maturity and default correlation set at $T = 5$ and $\rho = 0.4$.

Fig. 14(a)–(c), respectively, illustrate the spreads of the equity, mezzanine, and senior tranches with respect to different values of $R$ and $\lambda$. All the figures show that spread increases with increase of $\lambda$ when $R$ is fixed, and decreases with $R$ when $\lambda$ is fixed.

Fig. 15(a)–(c), respectively, illustrate the default delta sensitivities of the equity, mezzanine, and senior tranches with respect to different values of $R$ and $\lambda$. Fig. 15(a) shows that the delta sensitivity decreases with the increase of $\lambda$ and slightly increases with the increase of $R$ in the equity tranche. Fig. 15(b) shows that the delta sensitivity for the mezzanine tranche is non-monotone. For small values of $R$ and $\lambda$, the delta sensitivity decreases with increase in $\lambda$ and increases with increase in $R$, having the same characteristic of the equity tranche. For large values of $R$ and $\lambda$, the delta sensitivity decreases with the increases of $R$ and $\lambda$. Fig. 15(c) shows that in the
Fig. 11. The Delta Sensitivity of Mezzanine Tranche with Different Maturity and Default Correlation in the Analytic Method (a) and Monte Carlo Method (b).
Fig. 12. The Senior Tranche Spread with Different Maturity and Default Correlation in the Analytic Method (a) and Monte Carlo Method (b).
Fig. 13. The Delta Sensitivity of Senior Tranche with Different Maturity and Default Correlation in the Analytic Method (a) and Monte Carlo Method (b).
Table 4. Spreads of Equity, Mezzanine, and Senior Tranches with Different Values of $\rho$ for $T=5$.

<table>
<thead>
<tr>
<th>Default Correlation $\rho$</th>
<th>Equity</th>
<th>Mezzanine</th>
<th>Senior</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Analytic</td>
<td>Monte Carlo</td>
<td>Difference</td>
</tr>
<tr>
<td>0</td>
<td>1.1065</td>
<td>1.1036</td>
<td>0.0029</td>
</tr>
<tr>
<td>0.1</td>
<td>0.7372</td>
<td>0.7375</td>
<td>0.0003</td>
</tr>
<tr>
<td>0.2</td>
<td>0.5347</td>
<td>0.5327</td>
<td>0.0020</td>
</tr>
<tr>
<td>0.3</td>
<td>0.4033</td>
<td>0.4009</td>
<td>0.0024</td>
</tr>
<tr>
<td>0.4</td>
<td>0.3210</td>
<td>0.3185</td>
<td>0.0025</td>
</tr>
<tr>
<td>0.5</td>
<td>0.2493</td>
<td>0.2470</td>
<td>0.0023</td>
</tr>
<tr>
<td>0.6</td>
<td>0.1973</td>
<td>0.1944</td>
<td>0.0028</td>
</tr>
<tr>
<td>0.7</td>
<td>0.1557</td>
<td>0.1537</td>
<td>0.0019</td>
</tr>
<tr>
<td>0.8</td>
<td>0.1152</td>
<td>0.1156</td>
<td>0.0003</td>
</tr>
<tr>
<td>0.9</td>
<td>0.0808</td>
<td>0.0794</td>
<td>0.0015</td>
</tr>
<tr>
<td>Default correlation $\rho$</td>
<td>Equity</td>
<td>Mezzanine</td>
<td>Senior</td>
</tr>
<tr>
<td>---------------------------</td>
<td>--------</td>
<td>----------</td>
<td>--------</td>
</tr>
<tr>
<td></td>
<td>Analytic</td>
<td>Monte Carlo</td>
<td>Difference</td>
</tr>
<tr>
<td>0</td>
<td>0.0014</td>
<td>0.0014</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.1</td>
<td>0.0025</td>
<td>0.0028</td>
<td>-0.0003</td>
</tr>
<tr>
<td>0.2</td>
<td>0.0033</td>
<td>0.0035</td>
<td>-0.0002</td>
</tr>
<tr>
<td>0.3</td>
<td>0.0036</td>
<td>0.0033</td>
<td>0.0003</td>
</tr>
<tr>
<td>0.4</td>
<td>0.0038</td>
<td>0.0037</td>
<td>0.0001</td>
</tr>
<tr>
<td>0.5</td>
<td>0.0037</td>
<td>0.0040</td>
<td>-0.0003</td>
</tr>
<tr>
<td>0.6</td>
<td>0.0036</td>
<td>0.0034</td>
<td>0.0002</td>
</tr>
<tr>
<td>0.7</td>
<td>0.0033</td>
<td>0.0031</td>
<td>0.0002</td>
</tr>
<tr>
<td>0.8</td>
<td>0.0030</td>
<td>0.0030</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.9</td>
<td>0.0026</td>
<td>0.0027</td>
<td>-0.0001</td>
</tr>
</tbody>
</table>
senior tranche, the delta sensitivity increases with \( \lambda \) and decreases with increase in \( R \).

### 4.3. Implied Correlation and Base Correlation

The implied correlation for each tranche is calculated as the correlation that makes the spread of the tranche equal to its market price. It is sometimes referred as “compound correlation” in the CDO literature. The implied correlation is usually calculated based on trial and error. One major disadvantage of the implied correlation is that it exhibits a “smile.” For overcoming this problem, the base correlation is proposed by JP Morgan (see McGinty & Ahluwalia, 2004). The base correlation is calculated by defining a series of hypothetical equity tranches.

The first equity tranche remains unchanged at detachment points [0\%, 3\%]. The mezzanine tranche is now replaced conceptually by a hypothetical
equity tranche at detachment points [0\%, 14\%] that combines the original equity tranche [0\%, 3\%] and mezzanine tranche [3\%, 14\%]. The base correlation for the new tranche at “mezzanine” level is calculated as the correlation that makes the spread of this hypothetical tranche equal to its market price that would be the sum of the market prices of the original equity [0\%, 3\%] tranche and the original mezzanine [3\%, 14\%] tranche. For computing the model price, the expected losses of the new hypothetical tranche is the sum of the expected losses in the original equity [0\%, 3\%] tranche and the original mezzanine [3\%, 14\%] tranche.

In the same way, the senior tranche is now replaced conceptually by a hypothetical equity tranche at detachment points [0\%, 100\%] that combines the original equity tranche [0\%, 3\%], mezzanine tranche [3\%, 14\%], and senior tranche [14\%, 100\%]. The base correlation for the new tranche at “senior” level is calculated as the correlation that makes the spread of this

Fig. 15. Default Delta Sensitivity with Respect to Different Default Intensities and Recovery Rates in the Equity Tranche (a), Mezzanine Tranche (b), and Senior Tranche (c).

Empirical Study of Pricing and Hedging Collateralized Debt Obligation
hypothetical tranche equal to its market price that would be the sum of the market prices of the original equity [0%, 3%] tranche, the original mezzanine [3%, 14%] tranche, and the senior tranche [14%, 100%]. For computing the model price, the expected losses of the new hypothetical tranche is the sum of the expected losses in the original equity [0%, 3%] tranche, the original mezzanine [3%, 14%] tranche, and the original senior tranche [14%, 100%].

Fig. 16 shows the implied correlation and the base correlation calculated in the equity, mezzanine, and senior tranches. It can be observed that the implied correlation is larger for equity and senior tranches and smaller for mezzanine tranche, exhibiting a “smile” characteristic. In contrast, the base correlation does not display the smile though it increases slightly with the seniority of tranches.

4.4. Extending to Lévy Processes

From the empirical results it is clear that the spreads in the various tranches are sensitive to the default probabilities. In particular, in the analytic model, the default probabilities are represented by \(X_k\) which follows a distribution, for example, in the single-factor Gaussian Copula model. With a single factor approach, one can extend the default modeling to encompass more
complicated situations with modeling correlated defaults or introducing fat-tailed distribution to $X_k$. Examples of the latter include the variance gamma, the normal inverse Gaussian (NIG), the Meixner, and other distributions. These generically belong to the class of Lévy processes described in Sato (2000). NIG processes are also discussed in detail in Rydberg (1996). Using the NIG process, for example, could lead to a more accurate pricing of all the tranches within a CDO structure. This is because by more accurately modeling the default probabilities at different loss levels, the spread in each tranche is more accurately priced. An NIG process could be used as follows.

Following Eq. (7), suppose there is a single (common) factor model of $X_k$ comprising a common factor $M$ and an individual factor $Z_k$ where

$$X_k = a_k M + \sqrt{1-a_k^2} Z_k, \quad k = 1, 2, \ldots, K$$

but where $M$ and $Z_k$ now follow independent NIG processes. In particular, the density of the NIG $(X_k; \alpha, \beta, \delta, \mu)$ is given by

$$f(X_k) = \frac{\alpha \delta}{\pi} \exp\left(\delta \sqrt{\alpha^2 - \beta^2 + \beta(X_k - \mu)}\right) \sqrt{\delta^2 + (X_k - \mu)^2} K_1\left(\sqrt{\delta^2 + (X_k - \mu)^2}\right)$$

where $\alpha^2 > \beta^2 > 0$, and $K_1(\omega) = \frac{1}{2} \int_0^\infty \exp(-(1/2)\omega(\theta + \theta^{-1})) d\theta$ is modified Bessel function of the third kind. Then the probability distribution of default can be modeled by the distribution of the NIG process as described. Simulations can be performed according to the density function.

An immediate outcome of Lévy process modeling of the default processes would be the more accurate pricing of individual tranches within a CDO. Fatter tails allocated to the probabilities of default modeling would provide for higher default intensities at the equity tranche and also at the senior tranche. This would imply that compared to the Gaussian Copula method, an NIG method is likely to produce higher theoretical spreads for equity and senior tranches, and lower spread for the mezzanine tranche. By matching to the market price, this would in turn imply that the implied correlations for the equity and the senior tranche under Lévy processes would be lower than those in the Gaussian process. In the latter, the implied correlation has to work harder and be pumped up in order to reflect a higher market price due to the higher default probabilities. Since the latter are captured by the Lévy process, the implied correlation becomes flatter. Indeed, many research undertaken at this time all attempt to bring about
a flat implied correlation curve using fat-tailed processes. In this sense, the correlation bias or smile can be explained away.

5. CONCLUSIONS

This paper studies the pricing of CDO using Monte Carlo and analytic methods. The portfolio loss, the expected loss in each CDO tranche, the tranche spread, and the default delta sensitivity are analyzed with respect to maturity, default correlation, default intensity, and recovery rate. The results are summarized as follows.

5.1. Maturity

The portfolio loss and the loss distributed in each tranche increases with the increase in time to maturity $T$, due to the higher probability of default for larger $T$. The spread of equity tranche is not sensitive to $T$. However, in the mezzanine and senior tranches, the spreads increase with a larger $T$. As for default delta sensitivity, the equity tranche showed mixed characteristics with respect to $T$. Both these sensitivities of the mezzanine and senior tranches increase with increase in $T$.

5.2. Default Correlation

The portfolio expected loss EL is not sensitive to the default correlation $\rho$ once the intensity $\lambda$ is fixed. Equity tranche EL appears to decrease with increase of $\rho$. In contrast, senior tranche EL increases with increase of $\rho$; and mezzanine EL displays both possibilities. Similar to EL, the spread as well as default correlation sensitivity of equity tranche decreases and that of senior tranche increases with the increase of $\rho$. Mezzanine tranches show mixed results. For sensitivity, both equity and mezzanine tranches have mixed results with respect to $\rho$, while senior tranche sensitivity increases with increase of $\rho$.

5.3. Intensity

The spreads in all tranches increase with increase of $\lambda$. Equity tranche’s default delta sensitivity decreases and that of senior tranche’s increases with increase of $\lambda$; mezzanine tranche shows mixed results.
5.4. Recovery

Contrary to the case of intensity, recovery rate has an reverse relationship with the spreads of all tranches. The spreads in all tranches decrease with increase of recovery rate $R$. The sensitivity of equity increases and that of senior tranche decreases with increase of $R$; mezzanine tranche shows mixed results.

The analysis of default correlation shows that the implied default correlation has a “smile” characteristic, while the base correlation increases slightly with the seniority of tranches.

Our results also show that the Monte Carlo method is slower in terms of computational time than the analytic method. Monte Carlo does not appear to be a satisfactory approach for calculating default delta sensitivity as the sensitivity values computed under Monte Carlo vary widely.

Considering the disadvantages of the current Monte Carlo methods, future work should explore improved Monte Carlo methods. The performance of the analytic approach can also be further improved in future work. The likelihood and pathwise methods used in Joshi and Kainth (2004) can be explored for calculating the default delta sensitivity of CDO.

NOTE

1. This assumption effectively puts any loss at the beginning of the CDO as zero, since any loss within $[T_0, T_1]$ is paid at $T_1$.

REFERENCES


THE SKEWED $t$ DISTRIBUTION FOR PORTFOLIO CREDIT RISK

Wenbo Hu and Alec N. Kercheval

ABSTRACT

Portfolio credit derivatives, such as basket credit default swaps (basket CDS), require for their pricing an estimation of the dependence structure of defaults, which is known to exhibit tail dependence as reflected in observed default contagion. A popular model with this property is the (Student’s) $t$-copula; unfortunately there is no fast method to calibrate the degree of freedom parameter.

In this paper, within the framework of Schönbucher’s copula-based trigger-variable model for basket CDS pricing, we propose instead to calibrate the full multivariate $t$ distribution. We describe a version of the expectation-maximization algorithm that provides very fast calibration speeds compared to the current copula-based alternatives.

The algorithm generalizes easily to the more flexible skewed $t$ distributions. To our knowledge, we are the first to use the skewed $t$ distribution in this context.

1. INTRODUCTION

For portfolio risk modeling and basket derivative pricing, it is essential to understand the dependence structure of prices, default times, or other
asset-related variables. This structure is completely described by the second moments (the covariance matrix) for jointly normal variables, so practitioners often use the covariance matrix as a simple proxy for multivariate dependence. However, it is widely acknowledged that prices, returns, and other financial variables are not normally distributed. They have fat tails, and exhibit “tail dependence” (see Section 4), in which correlations are observed to rise during extreme events. Therefore there is a need for practical uses of more general multivariate distributions to model joint price behavior. This raises the question of how to choose these distributions, and, once chosen, how to efficiently calibrate them to data. In this paper, we look at the multivariate (Student’s) $t$ distribution, which has become a popular choice because of its heavy tails and non-zero tail dependence, and its generalization, the skewed $t$ distribution, described, for example, by Demarta and McNeil (2005) – see Section 2 below.

It has become popular and useful to isolate the dependence structure of a distribution from the individual marginal distributions by looking at its copula (see Section 3). Copulas that come from known distributions inherit their names (e.g., we have the Gaussian copulas, the $t$ copulas, etc.).

There are now many financial applications of copulas. For example, Di Clemente and Romano (2003b) used copulas to minimize expected shortfall (ES) in modeling operational risk. Di Clemente and Romano (2004) applied the same framework in the portfolio optimization of credit default swaps (CDS). Masala, Menzietti, and Micocci (2004) used the $t$ copula and a transition matrix with a gamma-distributed hazard rate and a $b$-distributed recovery rate to compute the efficient frontier for credit portfolios by minimizing ES.

The success of copulas greatly depends both on good algorithms for calibrating the copula itself, and on the availability of a fast algorithm to calculate the cumulative distribution functions (CDF) and quantiles of the corresponding one-dimensional marginal distributions.

The calibration of a $t$ copula is very fast if we fix the degree of freedom parameter $\nu$, which in turn is optimized by maximizing a log likelihood; however, the latter is slow. Detailed algorithms for calibrating $t$ copulas can be found in the work of many researchers, such as Di Clemente and Romano (2003a), Demarta and McNeil (2005), Mashal and Naldi (2002), and Galiani (2003).

The calibration of a $t$ copula is (by definition) separate from the calibration of marginal distributions. It is generally suggested to use the empirical distributions to fit the margins, but empirical distributions tend to have poor performance in the tails. A hybrid of the parametric and non-parametric
methods considers the use of the empirical distribution in the center and a generalized Pareto distribution (GPD) in the tails. Some use a Gaussian distribution in the center. To model multivariate losses, Di Clemente and Romano (2003a) used a $t$ copula and Gaussian distribution in the center and left tail and a GPD in the right tail for the margins. We will be able to avoid these issues because we can effectively calibrate the full distribution directly by using $t$ or skewed $t$ distributions.

In this paper, the primary application we have in mind is portfolio credit risk – specifically, the pricing of multiname credit derivatives such as $k$th-to-default basket credit default swaps (basket CDS).

For this problem, the most important issue is the correlation structure among the default obligors as described by the copula of their default times. Unfortunately, defaults are rarely observed, so it is difficult to calibrate their correlations directly. In this paper, we follow Cherubini, Luciano, and Vecchiato (2004) and use the distribution of daily equity prices to proxy the dependence structure of default times (see Section 6.2 below).

Several groups have discussed the pricing of basket CDS and CDO via copulas, such as Galiani (2003), Mashal and Naldi (2002), and Meneguzzo and Vecchiato (2003), among others. However, in this paper, we find that calibrating the full joint distribution is much faster than calibrating the copula separately, because of the availability of the expectation-maximization (EM) algorithm discussed below.

In Hu (2005), we looked at the large family of generalized hyperbolic distributions to model multivariate equity returns by using the EM algorithm (see Section 2). We showed that the skewed $t$ has better performance and faster convergence than other generalized hyperbolic distributions. Furthermore, for the $t$ distribution, we have greatly simplified formulas and an even faster algorithm. For the $t$ copula, there is still no good method to calibrate the degree of freedom $v$ except to find it by direct search. The calibration of a $t$ copula takes days while the calibration of a skewed $t$ or $t$ distribution via the EM algorithm takes minutes. To our knowledge, we are the first to directly calibrate the skewed $t$ or $t$ distributions to price basket CDS.

This paper is organized as follows. In Section 2, we introduce the skewed $t$ distribution from the normal mean-variance mixture family and provide a version of the EM algorithm to calibrate it, including the limiting $t$ distribution. We give an introduction to copulas in Section 3, and review rank correlation and tail dependence in Section 4.

In Section 5, we follow Rutkowski (1999) to review the reduced form approach to single name credit risk. In Section 6, we follow Schönbucher (2003) to provide our model setup for calculating default probabilities for
the $k$th to default using a copula-based trigger-variable method. There we also discuss the calibration problem.

In Section 7, we apply all the previous ideas to describe a method for pricing basket CDS. We illustrate how selecting model copulas with different tail dependence coefficients (TDCs) influences the relative probabilities of first and last to default (FTD and LTD). We then argue that calibrating the skewed $t$ distribution is the best and fastest approach, among the common alternatives.

## 2. SKEWED $t$ DISTRIBUTIONS AND THE EM ALGORITHM

### 2.1. Skewed $t$ and $t$ Distributions

**Definition 1 Inverse Gamma Distribution.** The random variable $X$ has an inverse gamma distribution, written $X \sim \text{InverseGamma}(\alpha, \beta)$, if its probability density function is

$$f(x) = \frac{\beta^\alpha x^{-\alpha-1}e^{-\beta/x}}{\Gamma(\alpha)}, \quad x > 0, \quad \alpha > 0, \quad \beta > 0$$

where $\Gamma$ is the usual gamma function.

We have the following standard formulas:

- $E(X) = \frac{\beta}{\alpha - 1}, \quad \text{if } \alpha > 1$  
- $\text{Var}(X) = \frac{\beta^2}{(\alpha - 1)^2(\alpha - 2)}, \quad \text{if } \alpha > 2$
- $E(\log(X)) = \log(\beta) - \psi(\alpha)$

where

$$\psi(\alpha) = \frac{d\log(\Gamma(x))}{dx}$$  

is the digamma function.

The skewed $t$ distribution is a subfamily of the generalized hyperbolic distributions – see McNeil, Frey, and Embrechts (2005), who suggested the name “skewed $t$.” It can be represented as a normal mean–variance mixture, where the mixture variable is inverse gamma distributed.
**Definition 2 Normal Mean–Variance Mixture Representation of Skewed \( t \) Distribution.** Let \( \mu \) and \( \gamma \) be parameter vectors in \( \mathbb{R}^d \), \( \Sigma \) be a \( d \times d \) real positive semidefinite matrix, and \( \nu > 2 \). The \( d \)-dimensional skewed \( t \) distributed random vector \( X \), which is denoted by

\[
X \sim \text{Skewed } T_{d}(\nu, \mu, \Sigma, \gamma)
\]

is a multivariate normal mean–variance mixture variable with distribution given by

\[
X \overset{d}{=} \mu + W\gamma + \sqrt{W}Z
\]

where \( Z \sim \mathcal{N}_d(0, \Sigma) \), the multivariate normal with mean 0 and covariance \( \Sigma \); \( W \sim \text{InverseGamma}(\nu/2, \nu/2) \); and \( W \) is independent of \( Z \).

Here, \( \mu \) are location parameters, \( \gamma \) are skewness parameters, and \( \nu \) is the degree of freedom.

From the definition, we can see that

\[
X|W \sim N_d(\mu + W\gamma, W\Sigma)
\]

This is also why it is called a normal mean–variance mixture distribution.

We can get the following moment formulas easily from the mixture definition:

\[
E(X) = \mu + E(W)\gamma
\]

\[
\text{COV}(X) = E(W)\Sigma + \text{var}(W)\gamma\gamma'
\]

when the mixture variable \( W \) has finite variance \( \text{var}(W) \).

**Definition 3.** Setting \( \gamma \) equal to zero in Definition 2 defines the multivariate \( t \) distribution,

\[
X \overset{d}{=} \mu + \sqrt{W}Z
\]

For convenience, we next give the density functions of these distributions. Denoting by \( K_\lambda(x) \), \( x > 0 \), the modified Bessel function of the third kind, with index \( \lambda \) is:

\[
K_\lambda(x) = \frac{1}{2} \int_{0}^{\infty} y^{\lambda-1} e^{-x(y^{1/2} + y^{1/2})} dy
\]
The following formula may be computed using Eq. (7), and is given in McNeil et al. (2005).

**Proposition 1 Skewed $t$ Distribution.** Let $X$ be skewed $t$ distributed, and define

$$
\rho(x) = (x - \mu)'\Sigma^{-1}(x - \mu)
$$

(11)

Then the joint density function of $X$ is given by

$$
f(x) = c \frac{K_{(v+d)/2} \left( \sqrt{(v + \rho(x))(\gamma'\Sigma^{-1}\gamma)} \right) e^{(x-\mu)'\Sigma^{-1}\gamma}}{\sqrt{(v + \rho(x))(\gamma'\Sigma^{-1}\gamma)^{-\frac{v+d}{2}}(1 + (\rho(x)/\nu))^{\frac{v+d}{2}}}}
$$

(12)

where the normalizing constant is

$$
c = \frac{2^{1-(v+d)/2}}{\Gamma(v/2)(\pi v)^{d/2} |\Sigma|^{1/2}}
$$

The mean and covariance of a skewed $t$ distributed random vector $X$ are

$$
E(X) = \mu + \gamma \frac{v}{v - 2}
$$

(13)

$$
\text{COV}(X) = \frac{v}{v - 2} \Sigma + \gamma^{\prime} \frac{2v^2}{(v - 2)^2(v - 4)}
$$

(14)

where the covariance matrix is only defined when $v > 4$, and the expectation only when $v > 2$.

Furthermore, in the limit as $\gamma \rightarrow 0$ we get the joint density function of the $t$ distribution:

$$
f(x) = \frac{\Gamma((v + d)/2)}{\Gamma(v/2)(\pi v)^{d/2} |\Sigma|^{1/2}} \left( 1 + \frac{\rho(x)}{\nu} \right)^{-\frac{v+d}{2}}
$$

(15)

The mean and covariance of a $t$ distributed random vector $X$ are

$$
E(X) = \mu
$$

(16)

$$
\text{COV}(X) = \frac{v}{v - 2} \Sigma
$$

(17)
2.2. Calibration of $t$ and Skewed $t$ Distributions Using the EM Algorithm

The mean–variance representation of the skewed $t$ distribution has a great advantage: the so-called EM algorithm can be applied to such a representation. See McNeil et al. (2005) for a general discussion of this algorithm for calibrating generalized hyperbolic distributions.

The EM algorithm is a two-step iterative process in which (the E-step) an expected log likelihood function is calculated using current parameter values, and then (the M-step) this function is maximized to produce updated parameter values. After each E and M step, the log likelihood is increased, and the method converges to a maximum log likelihood estimate of the distribution parameters.

What helps this along is that the skewed $t$ distribution can be represented as a conditional normal distribution, so most of the parameters ($\Sigma$, $\mu$, $\gamma$) can be calibrated, conditional on $W$, like a Gaussian distribution. We give a brief summary of our version of the EM algorithms for skewed $t$ and $t$ distributions here. Detailed derivations, along with comparisons to other versions, can be found in Hu (2005).

To explain the idea, suppose we have i.i.d. data $x_1, \ldots, x_n \in \mathbb{R}^d$ that we want to fit to a skewed $t$ distribution.

We seek parameters $\theta = (\nu, \mu, \Sigma, \gamma)$ to maximize the log likelihood

$$\log L(\theta; x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} \log f(x_i; \theta)$$

where $f(\cdot; \theta)$ denotes the skewed $t$ density function.

The method is motivated by the observation that if the latent variables $w_1, \ldots, w_n$ were observable, our optimization would be straightforward. We define the augmented log-likelihood function

$$\log \tilde{L}(\theta; x_1, \ldots, x_n, w_1, \ldots, w_n) = \sum_{i=1}^{n} \log f_{X,W}(x_i, w_i; \theta)$$

$$= \sum_{i=1}^{n} \log f_{X|W}(x_i|w_i; \mu, \Sigma, \gamma)$$

$$+ \sum_{i=1}^{n} \log h_W(w_i; \nu)$$

where $f_{X|W}(\cdot|w; \mu, \Sigma, \gamma)$ is the conditional normal $N(\mu + \omega \gamma, \sigma^2 \Sigma)$ and $h_W(\cdot; \nu)$ is the density of InverseGamma ($\nu/2, \nu/2$).
These two terms could be maximized separately if the latent variables were observable. Since they are not, the method is instead to maximize the expected value of the augmented log-likelihood \( \bar{L} \) conditional on the data and on an estimate of the parameters \( \theta \). We must condition on the parameters because the distribution of the latent variables depends on the parameters. This produces an updated guess for the parameters, which we then use to repeat the process until convergence.

To be more explicit, suppose we have a step \( k \) parameter estimate \( \theta^{[k]} \).

We carry out the following steps.

- **E-step:** Compute an objective function

  \[
  Q(\theta; \theta^{[k]}) = E(\log \bar{L}(\theta; x_1, \ldots, x_n, W_1, \ldots, W_n) | x_1, \ldots, x_n; \theta^{[k]})
  \]

  This can be done analytically and requires formulas for quantities like \( E(W_i | x_i, \theta^{[k]}) \), \( E(1/W_i | x_i, \theta^{[k]}) \), and \( E(\log W_i | x_i, \theta^{[k]}) \), which can all be explicitly derived from the definitions.

- **M-step:** Maximize \( Q \) to find \( \theta^{[k+1]} \).

  Using our explicit formulas for the skewed \( t \) distribution, we can compute the expectation and the subsequent maximum \( \theta \) explicitly. Below we summarize the resulting formulas needed for directly implementing this algorithm.

We will use a superscript in square brackets to denote the iteration counter. Given, at the \( k \)th step, parameter estimates \( \nu^{[k]} \), \( \Sigma^{[k]} \), \( \mu^{[k]} \), and \( \gamma^{[k]} \), let, for \( i = 1, \ldots, n 

\[
\rho_i^{[k]} = (x_i - \mu^{[k]})'(\Sigma^{[k]})^{-1}(x_i - \mu^{[k]})
\]

Define the auxiliary variables \( \theta_i^{[k]} \), \( \eta_i^{[k]} \), and \( \zeta_i^{[k]} \) by

\[
\theta_i^{[k]} = \left( \frac{\rho_i^{[k]} + v^{[k]}}{\gamma^{[k]} \Sigma^{[k]^{-1}} \gamma^{[k]}} \right)^{-1/2} \frac{K_{(v+d+2)/2}}{K_{(v+d)/2}} \left( \sqrt{\left( \rho_i^{[k]} + v^{[k]} \right) \left( \gamma^{[k]} \Sigma^{[k]^{-1}} \gamma^{[k]} \right)} \right)
\]

(18)

\[
\eta_i^{[k]} = \left( \frac{\rho_i^{[k]} + v^{[k]}}{\gamma^{[k]} \Sigma^{[k]^{-1}} \gamma^{[k]}} \right)^{1/2} \frac{K_{(v+d-2)/2}}{K_{(v+d)/2}} \left( \sqrt{\left( \rho_i^{[k]} + v^{[k]} \right) \left( \gamma^{[k]} \Sigma^{[k]^{-1}} \gamma^{[k]} \right)} \right)
\]

(19)
\[ \xi_i^{[k]} = \frac{1}{2} \log \left( \frac{\rho_i^{[k]} + v^{[k]}}{\gamma^{[k]}} \right) \]

\[ + \left( \frac{\partial^2 K_{(v+d)/2}}{\partial \zeta} \right) \left( \frac{\sqrt{(\rho_i^{[k]} + v^{[k]})(\gamma^{[k]})}}{\partial \zeta} \right) \bigg|_{\zeta=0} \]

\[ K_{(v+d)/2} \left( \sqrt{(\rho_i^{[k]} + v^{[k]})(\gamma^{[k]})} \right) \]

In the special case of the multivariate t distributions, we have simpler forms for above formulas:

\[ \theta_i^{[k]} = \frac{v^{[k]} + d}{\rho_i^{[k]} + v^{[k]}} \]

\[ \eta_i^{[k]} = \frac{\rho_i^{[k]} + v^{[k]}}{\gamma^{[k]}} \]

\[ \xi_i^{[k]} = \log \left( \frac{\rho_i^{[k]} + v^{[k]}}{2} \right) - \psi \left( \frac{d + v^{[k]}}{2} \right) \]

Let us denote

\[ \bar{\theta} = \frac{1}{n} \sum_{i=1}^{n} \theta_i, \quad \bar{\eta} = \frac{1}{n} \sum_{i=1}^{n} \eta_i, \quad \bar{\zeta} = \frac{1}{n} \sum_{i=1}^{n} \xi_i \]

**Algorithm 1.** EM algorithm for calibrating the t and skewed t distributions

1. Set the iteration counter \( k = 1 \). Select staring values for, \( \nu^{[1]} \), \( \gamma^{[1]} \), \( \mu^{[1]} \), and \( \Sigma^{[1]} \). Reasonable starting value for mean and dispersion matrix are the sample mean and sample covariance matrix.
2. Calculate \( \theta_i^{[k]} \), \( \eta_i^{[k]} \), and \( \xi_i^{[k]} \) and their averages \( \bar{\theta}, \bar{\eta}, \) and \( \bar{\zeta} \).
3. Update \( \gamma, \mu, \) and \( \Sigma \) according to

\[ \gamma^{[k+1]} = \frac{n^{-1} \sum_{i=1}^{n} \theta_i^{[k]} (\bar{x} - x_i)}{\bar{\theta}^{[k]} \bar{\eta}^{[k]} - 1} \]

\[ \mu^{[k+1]} = \frac{n^{-1} \sum_{i=1}^{n} \theta_i^{[k]} x_i - \gamma^{[k+1]} \bar{\eta}^{[k]} - 1}{\bar{\theta}^{[k]}} \]
\[
\Sigma^{[k+1]} = \frac{1}{n} \sum_{i=1}^{n} \theta_i^{[k]} (x_i - \mu^{[k+1]}) (x_i - \mu^{[k+1]})' - \tilde{\eta}^{[k]} \sigma^{[k+1]} \sigma^{[k+1]}' \tag{27}
\]

4. Compute \( \psi^{[k+1]} \) by numerically solving the equation
\[
-\psi\left(\frac{\psi}{2}\right) + \log\left(\frac{\psi}{2}\right) + 1 - \xi^{[k]} - \tilde{\theta}^{[k]} = 0 \tag{28}
\]

5. Set counter \( k := k + 1 \) and go back to step 2 unless the relative increment of log likelihood is small and in this case, we terminate the iteration.

The result of this algorithm is an estimate of the maximum likelihood parameter values for the given data.

### 3. COPULAS

Copulas are used to describe the dependence structure of a multivariate distribution, which is well discussed in Nelsen (1999). One of the definitions can be found in Li (1999), the first one to use copulas to price portfolio credit risk.

**Definition 4 Copula Functions.** \( U \) is a uniform random variable if it has a uniform distribution on the interval \([0, 1]\).

For \( d \) uniform random variables \( U_1, U_2, \ldots, U_d \), the joint distribution function \( C \), defined as
\[
C(u_1, u_2, \ldots, u_d) = P[U_1 \leq u_1, U_2 \leq u_2, \ldots, U_d \leq u_d]
\]
is called a copula function.

**Proposition 2 Sklar’s Theorem.** Let \( F \) be a joint distribution function with margins \( F_1, F_2, \ldots, F_d \), then there exists a copula \( C \) such that for all \( (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d \),
\[
F(x_1, x_2, \ldots, x_d) = C(F_1(x_1), F_2(x_2), \ldots, F_d(x_d)) \tag{29}
\]

If \( F_1, F_2, \ldots, F_d \) are continuous, then \( C \) is unique. Conversely, if \( C \) is a copula and \( F_1, F_2, \ldots, F_d \) are distribution functions, then the function \( F \) defined by Eq. (29) is a joint distribution function with margins \( F_1, F_2, \ldots, F_d \).
Corollary 1. If $F_1, F_2, \ldots, F_m$ are continuous, then, for any $(u_1, \ldots, u_m) \in [0, 1]^m$ we have

$$C(u_1, u_2, \ldots, u_m) = F(F_1^{-1}(u_1), F_2^{-1}(u_2), \ldots, F_m^{-1}(u_m))$$  \hspace{1cm} (30)$$

where $F_i^{-1}(u_i)$ denotes the inverse of the CDF, namely, for $u_i \in [0, 1]$, $F_i^{-1}(u_i) = \inf\{x : F_i(x) \geq u_i\}$.

The name copula means a function that couples a joint distribution function to its marginal distributions. If $X_1, X_2, \ldots, X_d$, are random variables with distributions $F_1, F_2, \ldots, F_d$, respectively, and a joint distribution $F$, then the corresponding copula $C$ is also called the copula of $X_1, X_2, \ldots, X_d$, and $(U_1, U_2, \ldots, U_d) = (F_1(X_1), F_2(X_2), \ldots, F_d(X_d))$ also has copula $C$. We will use this property to price basket CDS later.

We often assume the marginal distributions to be empirical distributions. Suppose that the sample data is $x_i = (x_{i,1}, \ldots, x_{i,d})$, where $i = 1, \ldots, n$, then we may take the empirical estimator of $j$th marginal distribution function to be

$$\hat{F}_j(x) = \frac{\sum_{i=1}^{n} I\{x_{ij} \leq x\}}{n + 1}$$  \hspace{1cm} (31)$$

(Demarta and McNeil (2005) suggested dividing by $n + 1$ to keep the estimation away from the boundary 1.) By using different copulas and empirical or other margins, we can create a rich family of multivariate distributions. It is not required that the margins and joint distribution be the same type of distribution.

Two types of copulas are widely used: Archimedean copulas and elliptical copulas. Archimedean copulas form a rich family of examples of bivariate copulas, including the well-known Frank, Gumbel, and Clayton copulas. These have only one parameter and are easy to calibrate. However, the usefulness of Archimedean copulas of more than two variables is quite limited: they have only one or two parameters, and enforce a lot of symmetry in the dependence structure, such as bivariate exchangeability, that is unrealistic for a portfolio of heterogeneous firms. Therefore, we now restrict attention to the elliptical copulas, which are created from multivariate elliptical distributions, such as the Gaussian and $t$ distributions, and their immediate generalizations, such as the skewed $t$ copula.

Definition 5 Multivariate Gaussian Copula. Let $R$ be a positive semi-definite matrix with $\text{diag}(R) = 1$ and let $\Phi_R$ be the standardized
multivariate normal distribution function with correlation matrix \( R \). Then the multivariate Gaussian copula is defined as

\[
C(u_1, u_2, \ldots, u_m; R) = \Phi_R(\Phi^{-1}(u_1), \Phi^{-1}(u_2), \ldots, \Phi^{-1}(u_m)) \tag{32}
\]

where \( \Phi^{-1}(u) \) denotes the inverse of the standard univariate normal CDF.

**Definition 6 Multivariate \( t \) Copula.** Let \( R \) be a positive semidefinite matrix with \( \text{diag}(R) = 1 \) and let \( T_{R,v} \) be the standardized multivariate \( t \) distribution function with correlation matrix \( R \) and \( v \) degrees of freedom. Then the multivariate \( t \) copula is defined as

\[
C(u_1, u_2, \ldots, u_m; R, v) = T_{R,v}(T^{-1}_v(u_1), T^{-1}_v(u_2), \ldots, T^{-1}_v(u_m)) \tag{33}
\]

where \( T^{-1}_v(u) \) denotes the inverse of standard univariate \( t \) cumulative distribution function.

### 4. MEASURES OF DEPENDENCE

All dependence information is contained in the copula of a distribution. However, it is helpful to have real-valued measures of the dependence of two variables. The most familiar example of this is Pearson’s linear correlation coefficient; however, this does not have the desirable properties we will see below.

#### 4.1. Rank Correlation

**Definition 7 Kendall’s \( \tau \).** Kendall’s \( \tau \) rank correlation for the bivariate random vector \((X, Y)\) is defined as

\[
\tau(X, Y) = P((X - \hat{X})(Y - \hat{Y}) > 0) - P((X - \hat{X})(Y - \hat{Y}) < 0) \tag{34}
\]

where \((\hat{X}, \hat{Y})\) is an independent copy of \((X, Y)\).

As suggested by Meneguzzo and Vecciato (2003), the sample consistent estimator of Kendall’s \( \tau \) is given by

\[
\hat{\tau} = \frac{\sum_{i,j=1}^{n} \text{sign}[(x_i - x_j)(y_i - y_j)]}{n(n - 1)/2} \tag{35}
\]
where \( \text{sign}(x) = 1 \) if \( x \geq 0 \), otherwise \( \text{sign}(x) = 0 \), and \( n \) is the number of observations.

In the case of elliptical distributions, Lindskog, McNeil, and Schmock (2003) showed that

\[
\tau(X, Y) = \frac{2}{\pi} \arcsin(\rho) \tag{36}
\]

where \( \rho \) is Pearson’s linear correlation coefficient between random variables \( X \) and \( Y \). However, Kendall’s \( \tau \) is more useful in discussions of dependence structure because it depends in general only on the copula of \((X, Y)\) (Nelsen, 1999):

\[
\tau(X, Y) = 4 \int_{[0,1]^2} C(u, v) dC(u, v) - 1 \tag{37}
\]

It has nothing to do with the marginal distributions. Sometimes, we may need the following formula

\[
\tau(X, Y) = 1 - 4 \int_{[0,1]^2} C_u(u, v) C_v(u, v) dudv \tag{38}
\]

where \( C_u \) denotes the partial derivative of \( C(u, v) \) with respect to \( u \) and \( C_v \) denotes the partial derivative of \( C(u, v) \) with respect to \( v \).

**Proposition 3 Copula of Transformations.** (Nelsen, 1999). Let \( X \) and \( Y \) be continuous random variables with copula \( C_{XY} \). If both \( \alpha(X) \) and \( \beta(Y) \) are strictly increasing on \( \text{Ran}X \) and \( \text{Ran}Y \), respectively, then \( C_{\alpha(X)\beta(Y)} = C_{XY} \). If both \( \alpha(X) \) and \( \beta(Y) \) are strictly decreasing on \( \text{Ran}X \) and \( \text{Ran}Y \), respectively, then \( C_{\alpha(X)\beta(Y)}(u, v) = u + v - 1 + C_{XY}(1 - u, 1 - v) \).

**Corollary 2 Invariance of Kendall’s \( \tau \) under Monotone Transformation.** Let \( X \) and \( Y \) be continuous random variables with copula \( C_{XY} \). If both \( \alpha(X) \) and \( \beta(Y) \) are strictly increasing or strictly decreasing on \( \text{Ran}X \) and \( \text{Ran}Y \), respectively, then \( \tau_{\alpha(X)\beta(Y)} = \tau_{XY} \).

**Proof.** We just need to show the second part. If both \( \alpha(X) \) and \( \beta(Y) \) are strictly decreasing, then \( C_{\alpha(X)\beta(Y)}(u, v) = u + v - 1 + C_{XY}(1 - u, 1 - v) \).

From Eq. (38), we have

\[
\tau_{\alpha(X)\beta(Y)} = 1 - 4 \int_{[0,1]^2} (1 - C_1(1 - u, 1 - v))(1 - C_2(1 - u, 1 - v)) dudv
\]
where $C_i$ denotes the partial derivative with respect to $i$th variable to avoid confusion. By replacing $(1-u)$ by $x$ and $(1-v)$ by $y$, we have

$$
\tau_{\alpha(X)\beta(Y)} = 1 - 4 \int_{[0,1]^2} (1 - C_1(x, y))(1 - C_2(x, y)) \, dx \, dy
$$

Since

$$
\int_{[0,1]^2} C_1(x, y) \, dx \, dy = \int_{[0,1]} y \, dy = 0.5
$$

we have $\tau_{\alpha(X)\beta(Y)} = \tau_{XY}$.

These results are the foundation of modeling of default correlation in the pricing of portfolio credit risk. From now on, when we talk about correlation, we will mean Kendall’s $\tau$ rank correlation.

### 4.2. Tail Dependence

Corresponding to the heavy tail property in univariate distributions, tail dependence is used to model the co-occurrence of extreme events. For credit risk, this is the phenomenon of default contagion. Realistic portfolio credit risk models should exhibit positive tail dependence, as defined next.

**Definition 8 Tail Dependence Coefficient.** Let $(X_1, X_2)$ be a bivariate vector of continuous random variables with marginal distribution functions $F_1$ and $F_2$. The level of upper tail dependence $\lambda_U$ and lower tail dependence $\lambda_L$ are given, respectively, by

$$
\lambda_U = \lim_{u \uparrow 1} P[X_2 > F_2^{-1}(u) | X_1 > F_1^{-1}(u)]
$$

$$
\lambda_L = \lim_{u \downarrow 0} P[X_2 \leq F_2^{-1}(u) | X_1 \leq F_1^{-1}(u)]
$$

If $\lambda_U > 0$, then the two random variables $(X_1, X_2)$ are said to be **asymptotically dependent in the upper tail**. If $\lambda_U = 0$, then $(X_1, X_2)$ are **asymptotically independent in the upper tail**. Similarly for $\lambda_L$ and the lower tail.
Joe (1997) gave the copula version of TDC,

\[
\lambda_U = \lim_{u \uparrow 1} \frac{1 - 2u + C(u, u)}{1 - u} \tag{41}
\]

\[
\lambda_L = \lim_{u \downarrow 0} \frac{C(u, u)}{u} \tag{42}
\]

For elliptical copulas, \( \lambda_U = \lambda_L \), denoted simply by \( \lambda \). Embrechts, Lindskog, and McNeil (2003) showed that for a Gaussian copula, \( \lambda = 0 \), and for a \( t \) copula,

\[
\lambda = 2 - 2t_{v+1} \left( \sqrt{v + 1} \frac{\sqrt{1 - \rho}}{\sqrt{1 + \rho}} \right) \tag{43}
\]

where \( \rho \) is the Pearson correlation coefficient. We can see that \( \lambda \) is an increasing function of \( \rho \) and a decreasing function of the degree of freedom \( v \).

The \( t \) copula is a tail dependent copula. We can see the difference of the tail dependence between Gaussian copulas and \( t \) copulas from Fig. 1.

**5. SINGLE NAME CREDIT RISK**

Before looking at the dependence structure of defaults for a portfolio, we first review the so-called reduced form approach to single firm credit risk,
sometimes called stochastic intensity modeling. We follow the approach of Rutkowski (1999).

5.1. Defaultable Bond Pricing

Suppose that $t$ is the default time of a firm. Let $H_t = I_{t \leq T}$, and $\mathcal{H}_t = \sigma(H_s : s \leq t)$ denote the default time information filtration. We denote by $F$ the right-continuous CDF of $\tau$, that is, $F(t) = P(\tau \leq t)$.

**Definition 9 Hazard Function.** The function $\Gamma : \mathbb{R}^+ \to \mathbb{R}^+$ given by

$$
\Gamma(t) = -\log(1 - F(t)), \quad \forall t \in \mathbb{R}^+
$$

is called the hazard function. If $F$ is absolutely continuous, that is, $F(t) = \int_0^t f(u)du$, where $f$ is the probability density function of $\tau$, then so is $\Gamma(t)$, and we define the intensity function

$$
\lambda(t) = \Gamma'(t)
$$

It is easy to check that

$$
F(t) = 1 - e^{-\int_0^t \lambda(u)du}
$$

and

$$
f(t) = \lambda(t)S(t)
$$

where $S(t) = 1 - F(t) = e^{-\int_0^t \lambda(u)du}$ is called the survival function.

For simplicity, we suppose the risk free short interest rate $r(t)$ is a non-negative deterministic function, so that the price at time $t$ of a unit of default free zero-coupon bond with maturity $T$ equals $B(t, T) = e^{-\int_t^T r(u)du}$.

Suppose now we have a defaultable zero-coupon bond that pays $c$ at maturity $T$ if there is no default, or pays a recovery amount $h(t)$ if there is a default at time $\tau < T$. The time $t$ present value of the bond’s payoff, therefore, is

$$
Y_t = I_{\{t < \tau \leq T\}}h(\tau)e^{-\int_t^\tau r(u)du} + I_{\{\tau > T\}}c e^{-\int_t^T r(u)du}
$$

When the only information is contained in the default filtration $\mathcal{H}_t$, we have the following pricing formula.
Proposition 4. Rutkowski (1999). Assume that $t \leq T$, and $Y_t$ is defined as above. If $\Gamma(t)$ is absolutely continuous, then

$$E(Y_t|\mathcal{H}_t) = I_{(\tau > t)} \left( \int_t^T h(u) \hat{\lambda}(u)e^{-\int_u^T \hat{r}(v)dv}du + c e^{-\int_t^T \hat{r}(u)du} \right)$$

(47)

where $\hat{r}(v) = r(v) + \hat{\lambda}(v)$.

The first term is the price of the default payment and the second is the price of the survival payment. Note that in the first term, we have used Eq. (46) to express the probability density function of $\tau$. In the case of zero recovery, the formula tells us that a defaultable bond can be valued as if it were default free by replacing the interest rate by the sum of the interest rate and a default intensity, which can be interpreted as a credit spread. We use this proposition to price basket CDS.

5.2. Credit Default Swaps

A CDS is a contract that provides insurance against the risk of default of a particular company. The buyer of a CDS contract obtains the right to sell a particular bond issued by the company for its par value once a default occurs. The buyer pays to the seller a periodic payment, at time $t_1, \ldots, t_n$, as a fraction $q$ of the nominal value $M$, until the maturity of the contract $T = t_n$ or until a default at time $\tau < T$ occurs. If a default occurs, the buyer still needs to pay the accrued payment from the last payment time to the default time. There are $1/\theta$ payments a year (for semiannual payments, $\theta = 1/2$), and every payment is $\theta q M$.

5.3. Valuation of Credit Default Swaps

Set the current time $t_0 = 0$. Let us suppose the only information available is the default information, interest rates are deterministic, the recovery rate $R$ is a constant, and the expectation operator $E(\cdot)$ is relative to a risk neutral measure. We use Proposition 4 to get the premium leg (PL), accrued payment (AP), and default leg (DL). PL is the present value of periodic payments and AP is the present value of the accumulated amount from last payment to default time. The DL is the present value of
the net gain to the buyer in case of default. We have

\[
PL = M \theta q \sum_{i=1}^{n} E(B(0, t_i)I\{\tau > t_i\})
\]

\[
= M \theta q \sum_{i=1}^{n} B(0, t_i)e^{-\int_{0}^{t_i} \lambda(u)du}
\]

(48)

\[
AP = M \theta q \sum_{i=1}^{n} E\left(\frac{\tau - t_{i-1}}{t_i - t_{i-1}} B(0, \tau)I\{t_{i-1} < \tau \leq t_i\}\right)
\]

\[
= M \theta q \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \frac{u - t_{i-1}}{t_i - t_{i-1}} B(0, u)\lambda(u)e^{-\int_{0}^{u} \lambda(s)ds} du
\]

(49)

\[
DL = M(1 - R)E(B(0, \tau)I\{\tau \leq T_1\})
\]

\[
= M(1 - R) \int_{0}^{T_1} B(0, u)\lambda(u)e^{-\int_{0}^{u} \lambda(s)ds} du
\]

(50)

The spread price \(q^*\) is the value of \(q\) such that the value of the CDS is zero,

\[
PL(q^*) + AP(q^*) = DL
\]

(51)

5.4. Calibration of Default Intensity: Illustration

As Hull (2002) points out, the CDS market is so liquid that we can use CDS spread data to calibrate the default intensity using Eq. (51).

In Table 1, we have credit default spread prices for five companies on 07/02/2004 from GFI (http://www.gfigroup.com). The spread price is quoted in basis points. It is the annualized payment made by the buyer of the CDS per dollar of nominal value. The mid price is the average of bid price and ask price.

We denote the maturities of the CDS contracts as \((T_1, \ldots, T_5) = (1, 2, 3, 4, 5)\). It is usually assumed that the default intensity is a step function, with step size of 1 year, expressed in the following form (where \(T_0 = 0\)),

\[
\lambda(t) = \sum_{i=1}^{5} c_i I_{(T_{i-1}, T_i)}(t)
\]

(52)
We can get \( c_1 \) by using the 1-year CDS spread price first. Knowing \( c_1 \), we can estimate \( c_2 \) using the 2-year CDS spread price. Following this procedure, we can estimate all the constants \( c_i \) for the default intensity.

In our calibration, we assume a recovery rate \( R \) of 0.4, a constant risk-free interest rate of 0.045, and semiannual payments \((\theta = 1/2)\). In this setting, we can get PL, AP, and DL explicitly. The calibrated default intensity is shown in Table 2.

### 6. PORTFOLIO CREDIT RISK

#### 6.1. Setup

Our setup for portfolio credit risk is to use default trigger variables for the survival functions (Schönbucher & Schübert, 2001), as a means of introducing default dependencies through a specified copula.

Suppose we are standing at time \( t = 0 \).

**Model Setup and Assumptions.** Suppose there are \( d \) firms. For each obligor \( 1 \leq i \leq d \), we define

---

**Table 1.** Credit Default Swap Mid Price Quote (Where Year 1, \ldots, Year 5 Mean Maturities).

<table>
<thead>
<tr>
<th>Company</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
<th>Year 4</th>
<th>Year 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT&amp;T</td>
<td>144</td>
<td>144</td>
<td>208</td>
<td>272</td>
<td>330</td>
</tr>
<tr>
<td>Bell South</td>
<td>12</td>
<td>18</td>
<td>24</td>
<td>33</td>
<td>43</td>
</tr>
<tr>
<td>Century Tel</td>
<td>59</td>
<td>76</td>
<td>92</td>
<td>108</td>
<td>136</td>
</tr>
<tr>
<td>SBC</td>
<td>15</td>
<td>23</td>
<td>31</td>
<td>39</td>
<td>47.5</td>
</tr>
<tr>
<td>Sprint</td>
<td>57</td>
<td>61</td>
<td>66</td>
<td>83</td>
<td>100</td>
</tr>
</tbody>
</table>

**Table 2.** Calibrated Default Intensity.

<table>
<thead>
<tr>
<th>Company</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
<th>Year 4</th>
<th>Year 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT&amp;T</td>
<td>0.0237</td>
<td>0.0237</td>
<td>0.0599</td>
<td>0.0893</td>
<td>0.1198</td>
</tr>
<tr>
<td>Bell South</td>
<td>0.0020</td>
<td>0.0040</td>
<td>0.0061</td>
<td>0.0105</td>
<td>0.0149</td>
</tr>
<tr>
<td>Century Tel</td>
<td>0.0097</td>
<td>0.0155</td>
<td>0.0210</td>
<td>0.0271</td>
<td>0.0469</td>
</tr>
<tr>
<td>SBC</td>
<td>0.0025</td>
<td>0.0052</td>
<td>0.0080</td>
<td>0.0109</td>
<td>0.0144</td>
</tr>
<tr>
<td>Sprint</td>
<td>0.0094</td>
<td>0.0108</td>
<td>0.0127</td>
<td>0.0235</td>
<td>0.0304</td>
</tr>
</tbody>
</table>
1. The default intensity \( \lambda^i(t) \): A deterministic function. We usually assume it to be a step function.

2. The survival function \( S^i(t) \):
\[
S^i(t) := \exp\left(-\int_0^t \lambda^i(u) du\right)
\] (53)

3. The default trigger variables \( U^i \): Uniform random variables on \([0, 1]\). The \( d \)-dimensional vector \( U := (U_1, U_2, \ldots, U_d) \) is distributed according to the \( d \)-dimensional copula \( C \) (see Definition 4).

4. The time of default \( \tau_i \) of obligor \( i \), where \( i = 1, \ldots, d \),
\[
\tau_i := \inf\{t : S^i(t) \leq 1 - U_i\}
\] (54)

The copula \( C \) of \( U \) is also called the survival copula of \( 1 - U \). (See Georges, Lamy, Nicolas, Quibel, and Roncalli (2001) for more details about survival copulas.)

From Eq. (54), we can see that the default time \( \tau_i \) is an increasing function of the uniform random variable \( U_i \), so the rank correlation Kendall’s \( \tau \) between default times is the same as the Kendall’s \( \tau \) between the uniform random variables, and the copula of \( \tau \) equals the copula of \( U \). Equivalently, the copula of \( 1 - U \) is the survival copula of \( \tau \).

Define the default function, \( F^i(t) = 1 - S^i(t) \).

**Theorem 1 Joint Default Probabilities.** The joint default probabilities of \( (\tau_1, \tau_2, \ldots, \tau_d) \) are given by
\[
P[\tau_1 \leq T_1, \tau_2 \leq T_2, \ldots, \tau_d \leq T_d] = C(F^1(T_1), \ldots, F^d(T_d))
\] (55)

**Proof.** From the definition of default in Eq. (54), we have
\[
P[\tau_1 \leq T_1, \ldots, \tau_d \leq T_d] = P[1 - U_1 \geq S^1(T_1), \ldots, 1 - U_d \geq S^d(T_d)]
\]
By the definition of the copula \( C \) of \( U \) we have
\[
P[\tau_1 \leq T_1, \ldots, \tau_d \leq T_d] = C(F^1(T_1), \ldots, F^d(T_d))
\]

6.2. Calibration

In the preceding setup, two kinds of quantities need to be calibrated: the default intensities \( \lambda^i(t) \) and the default time copula \( C \). Calibration of the
default intensities can be accomplished using the single name credit default spreads visible in the market, as described below in Section 7. However, calibration of the default time copula $C$ is difficult. Indeed, it is a central and fundamental problem for portfolio credit risk modeling to properly calibrate correlations of default times. The trouble is that data are scarce – for example, a given basket of blue chips may not have any defaults at all in recent history. In contrast, calibration using market prices of basket CDS is hampered by the lack of a liquid market with observable prices. Even if frequently traded basket CDS prices were observable, we would need many different basket combinations in order to extract full correlation information among all the names. Therefore, in the modeling process we need to choose some way of proxying the required data. McNeil et al. (2005) report that asset price correlations are commonly used as a proxy for default time correlations. This is also the approach taken by Cherubini et al. (2004), who remark that it is consistent with most market practice.

From the perspective of Merton-style value threshold models of default, it makes sense to use firm value correlations, since downward value co-movements will be associated with co-defaults. However, firm values are frequently not available, so asset prices can be used instead – even if, as Schönbucher (2003) points out, liquidity effects may lead to higher correlations for asset prices than for firm values.

Another way to simplify this calibration problem is to restrict to a family of copulas with only a small number of parameters, such as Archimedean copulas. Because this introduces too much symmetry among the assets, we choose instead to use asset price correlations as a proxy for default time correlations in this paper. This specific choice does not affect our conclusions, which apply to calibrating the copula of any asset-specific data set chosen to represent default time dependence.

A good choice of copula family for calibration is the $t$-copula, because it naturally incorporates default contagion through tail dependence, which is not present in the Gaussian copula. An even better choice is the skewed $t$-copula, for which the upper and lower tail dependence need not be equal.

When applying this copula approach, a direct calibration of the $t$-copula or skewed $t$-copula is time-consuming because there is no fast method of finding the degree of freedom $\nu$ except by looping. Instead, we will show that it is much faster to find the copula by calibrating the full multivariate distribution and extracting the implied copula, as in Eq. (33). This may seem counterintuitive, since the full distribution also contains the marginals as well as the dependence structure. However, for calibrating the full distribution function, we have at our disposal the fast EM algorithm; we
know of no corresponding algorithm for the copula alone. Moreover, we will see that the marginals are needed anyway to construct uniform variates. If they are not provided as a by-product of calibrating the full distribution, they need to be separately estimated.

7. PRICING OF BASKET CREDIT DEFAULT SWAPS: ELLIPTICAL COPULAS VERSUS THE SKEWED $t$ DISTRIBUTION

7.1. Basket CDS Contracts

We now address the problem of basket CDS pricing. For ease of illustration we will look at a 5-year basket CDS, where the basket contains the five firms used in Section 5.4; other maturities and basket sizes are treated in the same way. All the settings are the same as the single CDS except that the default event is triggered by the $k$th default in the basket, where $k$ is the seniority level of this structure, specified in the contract. The seller of the basket CDS will face the default payment upon the $k$th default, and the buyer will pay the spread price until $k$th default or until maturity $T$. Let $(\tau^1, \ldots, \tau^5)$ denote the default order.

The premium leg, accrued payment, and default leg are

$$PL = M q \sum_{i=1}^{n} E(B(0, t_i)1\{\tau^k > t_i\})$$

$$AP = M q \sum_{i=1}^{n} E\left(\frac{\tau_k - t_i}{t_i - t_{i-1}} B(0, \tau^k)1\{t_{i-1} < \tau^k \leq t_i\}\right)$$

$$DL = M(1 - R)E(B(0, \tau)1_{[\tau^k \leq T]})$$

The spread price $q^*$ is the $q$ such that the value of CDS is zero, that is,

$$PL(q^*) + AP(q^*) = DL$$

7.2. Pricing Method

To solve this equation, we now need the distribution of $\tau^k$, the time of the $k$th default in the basket, so that we can evaluate the foregoing expectations.
To do this, we need all the preceding tools of this paper. Here is a summary of the steps:

1. Select firm-specific critical variables $X$ whose dependence structure will proxy for the dependence structure of default times. (In the study below we use equity prices.)
2. Calibrate the copula $C$ of $X$ from a selected parametric family of copulas or distributions, such as the $t$ copula or the skewed $t$ distribution. In the distribution case, use the EM algorithm.
3. Separately, calibrate deterministic default intensities from single name CDS spread quotes, as in Section 5.
4. Use the default intensities to calculate survival functions $S^i(t)$ for each of the firms, using Eq. (53).
5. Using the copula $C$, develop the distribution of $k$th-to-default times by Monte Carlo sampling of many scenarios, as follows. In each scenario, choose a sample value of $U$ from the copula $C$. Use Eq. (54) to determine the default time for each firm in this scenario. Order these times from first to last to define $\tau^1, \ldots, \tau^5$. By repeating this simulation over many scenarios, we can develop a simulated unconditional distribution of each of the $k$th-to-default times $\tau^k$.
6. Use these distributions to compute the expectations in Eq. (59) in order to solve for the basket CDS spread price $q^*$.

7.3. The Distribution of $k$th-to-Default Times

Before describing our empirical results for this basket CDS pricing method, we elaborate a little on item 5 above, and examine via some experiments how the distributions depend on the choice of copula, comparing four different commonly used bivariate copulas: Gaussian, $t$, Clayton, and Gumbel.

To simplify the picture, we assume there are two idealized firms, with Kendall’s $\tau = 0.5$ for all copulas. We take a 5-year horizon and set the default intensity of the first firm to be a constant 0.05 and 0.03 for the second firm. We want to look at FTD and LTD probabilities at different times before maturity.

7.3.1. Algorithm

We calculated the $k$th to default probabilities using the following procedure:

1. Use Matlab™ copula toolbox 1.0 to simulate Gaussian, $t$, Clayton, and Gumbel copulas uniform variables $u_{ij}$ with the same Kendall’s $\tau$ correlation, where $i = 1, 2$, $j = 1, \ldots, n$, and $n$ is the number of samples.
2. From Eq. (54), we get \( \tau_{i,j} \) and sort according to column. The \( k \)th row is a series of \( k \)th to default times \( \tau^k_i \).

3. Divide the interval from year 0 to year 5 into 500 small sub-intervals. Count the number of \( \tau^k_i \) values that fall into each sub-interval and divide by the number of samples to get the default probabilities for each small sub-interval, and hence an approximate probability density function.

In the following, we illustrate results for FTD and LTD using \( n = 1,000,000 \) samples.

7.3.2. Empirical Probabilities of Last to Default and First to Default

First, we recall that the \( t \)-copula is both upper and lower tail dependent; the Clayton copula is lower tail dependent, but upper tail independent; the Gumbel copula is the reverse; and the Gaussian copula is tail independent in both tails.

We can see from Fig. 2 that a copula function with lower tail dependence (Clayton copula) leads to the highest default probabilities for LTD, while a copula function with upper tail dependence (Gumbel copula) leads to the lowest default probabilities. The tail dependent \( t \)-copula leads to higher default probabilities than tail independent Gaussian copula.

![Fig. 2. Default Probabilities of Last to Default (LTD).](image-url)
Default events tend happen when the uniform random variables $U$ are small (close to 0). Since the LTD requires that both uniform variables in the basket are small, a lower tail dependent copula will lead to higher LTD probabilities than a copula without lower tail dependence.

In Fig. 3, we see that the Clayton copula with only lower tail dependence leads to the lowest FTD probabilities, while the Gumbel copula with only upper tail dependence leads to the highest FTD probabilities. These results illustrate the sometimes unexpected relationships between tail dependence and FTD probabilities.

### 7.4. Empirical Basket CDS Pricing Comparison

We now use the method described in Section 7.2 to compare two approaches to the calibration of the copula $C$. The first approach, popular in the literature, is to directly calibrate a $t$ copula. Since this copula has tail dependence, it provides a way to introduce default contagion explicitly into the model. In order to get uniform variates, we will still need to specify marginal distributions, which we will take to be the empirical distributions.
The second approach is to calibrate the skewed $t$ distribution using the EM algorithm described earlier. Calibrating the full distribution frees us from the need to separately estimate the marginals. Also, the skewed $t$ distribution, has heavier tails than the $t$ distribution, and does not suffer from the bivariate exchangeability of the $t$ copula, which some argue is an unrealistic symmetry in the dependence structure of defaults.

In this experiment, we use for our critical variables the equity prices for the same five underlying stocks as used above: AT&T, Bell South, Century Tel, SBC, Sprint. We obtained the adjusted daily closing prices from http://www.finance.yahoo.com for the period 07/02/1998 to 07/02/2004.

7.4.1. Copula Approach
We first use the empirical distribution to model the marginal distributions and transform the equity prices into uniform variables. Then we can calibrate the $t$ copula using those variates. For comparison, we also calibrate a Gaussian copula.

If we fix in advance the degree of freedom $v$, the calibration of the $t$ copula is fast – see Di Clemente and Romano (2003a), Demarta and McNeil (2005), and Galiani (2003). However, we know of no appropriate method to calibrate the degree of freedom $v$. With this data, we find the degree of freedom to be $7.406$, which is found by maximizing the log likelihood using direct search, looping $v$ from 2.001 to 20 with step size 0.001. Each step takes about 5 s, and the full calibration takes about 24 h (2005 vintage laptop running Windows XP).

The maximum log likelihood for the Gaussian copula was 936.90, while for the $t$ copula it was 1043.94, substantially better. After calibration, we follow the remaining steps of Section 7.2 and report the results in the table below.

Demarta and McNeil (2005) also suggest using the skewed $t$ copula, but we were not able to calibrate it directly for this study.

7.4.2. Distribution Approach
We calibrate the multivariate $t$ and skewed $t$ distributions using the EM algorithm described in Section 2. The calibration is fast compared to the copula calibration: with the same data and equipment, it takes less than 1 min, compared to 24 h for the looping search of $v$. The calibrated degree of freedom for both $t$ and skewed $t$ is 4.31. The log likelihood for skewed $t$ and $t$ are almost the same: 18420.58 and 18420.20, respectively.

Spread prices for the $k$th to default basket CDS are reported in Table 3. We can see that lower tail dependent $t$ copula, compared to the Gaussian,
leads to higher default probability for LTD and lower probability for FTD, thus leads to higher spread price for LTD and lower spread price for FTD. The $t$ distribution has almost the same log likelihood and almost the same spread price of $k$th to default as the skewed $t$ distribution. Both distributions lead to higher spread price for LTD and lower spread price for FTD.

The calibration of the $t$ distribution is a superior approach, both because there is no extra requirement to assume a form for the marginals, and because the EM algorithm has tremendous speed advantages. Basket CDS or collateralized debt obligations usually have a large number of securities. For example, a synthetic CDO called EuroStoxx50 issued on May 18, 2001 has 50 single name CDS on 50 credits that belong to the DJ EuroStoxx50 equity index. In this case, the calibration of a $t$ copula will be extremely slow.

### Table 3. Spread Price for $k$th to Default Using Different Models.

<table>
<thead>
<tr>
<th>Model</th>
<th>FTD</th>
<th>2TD</th>
<th>3TD</th>
<th>4TD</th>
<th>LTD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian copula</td>
<td>525.6</td>
<td>141.7</td>
<td>40.4</td>
<td>10.9</td>
<td>2.2</td>
</tr>
<tr>
<td>$t$ copula</td>
<td>506.1</td>
<td>143.2</td>
<td>46.9</td>
<td>15.1</td>
<td>3.9</td>
</tr>
<tr>
<td>$t$ distribution</td>
<td>498.4</td>
<td>143.2</td>
<td>48.7</td>
<td>16.8</td>
<td>4.5</td>
</tr>
<tr>
<td>Skewed $t$ distribution</td>
<td>499.5</td>
<td>143.9</td>
<td>49.3</td>
<td>16.8</td>
<td>4.5</td>
</tr>
</tbody>
</table>

8. SUMMARY AND CONCLUDING REMARKS

We follow Rukowski’s single name credit risk modeling and Schönbucher and Schubert’s portfolio credit risk copula approach to price basket CDS. The $t$ copula is widely used in the pricing of basket CDS for its lower tail dependence. However, we need to specify the marginal distributions first and calibrate the marginal distributions and copula separately. In addition, there is no good (fast) method to calibrate the degree of freedom $\nu$.

Instead, we suggest using the fast EM algorithm for $t$ distribution and skewed $t$ distribution calibration, where all the parameters are calibrated together. To our knowledge, we are the first to suggest calibrating the full multivariate distribution to price basket CDS with this trigger-variable approach.

As compared to the Gaussian copula, the $t$ copula leads to higher default probabilities and spread prices of basket LTD credit default swaps, and lower default probabilities and spread prices for FTD, because of the introduction of tail dependence to model default contagion.
Both the $t$ distribution and the skewed $t$ distribution lead to yet higher spread prices of basket LTD credit default swaps and lower spread prices for FTD than the $t$ copula. This is suggestive of a higher tail dependence of default times than is reflected in the pure copula approach. Because default contagion has shown itself to be pronounced during extreme events, we suspect that this is a more useful model of real default outcomes.

We feel the skewed $t$ distribution has potential to become a powerful tool for quantitative analysts doing rich-cheap analysis of credit derivatives.

REFERENCES


This page intentionally left blank
CREDIT RISK DEPENDENCE MODELING WITH DYNAMIC COPULA: AN APPLICATION TO CDO TRANCHEs

Daniel Totouom and Margaret Armstrong

ABSTRACT

We have developed a new family of Archimedean copula processes for modeling the dynamic dependence between default times in a large portfolio of names and for pricing synthetic CDO tranches. After presenting a general procedure for constructing these processes, we focus on a specific one with lower tail dependence as in the Clayton copula. Using CDS data as on July 2005, we show that the base correlations given by this model at the standard detachment points are very similar to those quoted in the market for a maturity of 5 years.

1. INTRODUCTION

Base correlation, first developed by JP Morgan (McGinty and Ahulwalia, 2004), has become the industry standard for pricing CDO tranches. The base correlation model is based on The Homogeneous Large Pool Gaussian
Copula Model, which is a simplified version of the Gaussian copula widely used in the market. This model is not new, it is a simple methodology that is almost identical to the original CreditMetrics model (Gupton, Finger, & Bhatia, 1997). It is a simplified form of earlier one-factor models (Vasicek, 1987). It provides a mapping between CDO tranche prices or spreads and a single factor correlation surface.

Unfortunately it does not link prices/spreads at different times, which is needed for pricing different maturities and more importantly for forward starting CDOs. Ideally we would like a mathematically consistent model of the dependence structure between default times (as in factor copulas) that reproduces market prices and spreads (as base correlations do).

Over the past 5 years the factor copulas first proposed by Li (2000) have been widely used for pricing CDOs (see Andersen & Sidenius, 2005; Andersen, Sidenius, & Basu, 2003; Gregory & Laurent, 2003; Hull & White, 2003; Burtschell, Gregory, & Laurent, 2005a). Their strong points are that the pricing is semi-analytic and that the dependence structure between default times can be specified independent of the marginal credit curves. But as the CDS market became more liquid, it became clear that a flat correlation model did not price CDO tranches correctly (see Burtschell, Gregory, & Laurent, 2005b, for an example). Tests by Burtschell et al. (2005a) showed that the Clayton copula gave better results than other copulas, notably the Gaussian and Student’s $t$. Why is this?

Factor copulas based on the normal distribution (or Student’s $t$) have symmetric upper and lower tails. They are effectively saying that defaults occur in the same way in bull and bear markets. In tough times, one default tends to trigger others, which is not the case in normal times. The classic “ice cream cone” shape of the Clayton copula with its lower tail dependence (Fig. 1, left) captures this insight; the symmetric Gaussian (normal distribution) copula (Fig. 1, right) does not. The Clayton copula belongs to a special family of copulas called Archimedean copulas. While books have been written about their statistical properties (Nelsen, 1999; Joe, 1997), very little work has been done on stochastic processes based on them. In this chapter we present a family of dynamic Archimedean copula processes suitable for pricing synthetic CDO tranches.

This chapter is organized as follows. In the next section, after giving an overview of Archimedean copulas we introduce the new family of dynamic copula processes. In Section 3, we present a specific copula process related to the Clayton, which is lower tail dependent but not upper tail dependent. In Section 4 this model is used to price standard CDO tranches assuming a bullet exposure at maturity (5 years) and a large but not necessarily
homogeneous portfolio. Using market data (Anonymous, 2005) we show that a correlation skew similar to that observed in the market in July 2005, can be obtained with a suitable set of parameter values. In fact a wide range of correlation skews (both convex and concave) can be obtained, depending on the parameter values. The conclusions follow in the last section.

2. DYNAMIC ARCHIMEDEAN COPULA PROCESSES

Copulas express the dependence structure between two or more variables $X_1, \ldots, X_N$ separately from their marginal distributions. The original variables $X_1, \ldots, X_N$ are replaced their cumulative distribution functions $V_1 = F_1(X_1), \ldots, V_N = F_N(X_N)$ which are uniform on [0, 1]. In our case they will represent the latent variables of default probabilities of $N$ names in a credit portfolio. Archimedean copulas are a special type of copula that are defined via a generator, $f$:

$$ C(v_1, \ldots, v_N) = f^{-1}[f(v_1) + \cdots + f(v_N)] $$  \hspace{1cm} (1.a)

The copula function $C$ represents a joint distribution function for random variables $V_i$:

$$ C(v_1, \ldots, v_N) = \text{Probability}(V_1 < v_1, \ldots, V_N < v_N) $$  \hspace{1cm} (1.b)

Fig. 1. Clayton Copula with Parameter $\theta = 5$ (Left) and Gaussian Copula with $\rho = 0.87$ (Right).
While many bivariate Archimedean copulas are known, few multivariate ones exist because their generators have to be Laplace transforms. Table 1 lists selected multivariate Archimedean copulas with their Laplace transforms. For example, the Clayton copula corresponds to a gamma distribution. Burtschell et al. (2005a) showed that the Clayton copula was useful for modeling the correlation smile at a fixed point in time. The question was how to develop a dynamic continuous time stochastic process whose values at any given time have a given Archimedean copula (in our case, one with lower tail dependence). Our approach is based on an observation found in Rogge and Schonbucher (2003): let $Y$ be a positive random variable whose Laplace transform is $\varphi(s)$ and let $U_i$ be $n$ uniform random variables on $[0, 1]$ that are mutually independent and also independent of $Y$. Then the $n$ random variables $V_i$ defined by

$$V_i = \varphi\left(\frac{-\ln(U_i)}{Y}\right) \quad \text{for } i = 1, \ldots, n$$

(2)

are uniform on $[0, 1]$, and their cumulative distribution function is given as

$$\text{Prob}(V_1 \leq v_1, \ldots, V_N \leq v_N) = \varphi\left(\sum_{i=1}^{N} \varphi^{-1}(v_i)\right)$$

(3)

Consequently their multivariate copula is the Archimedean copula having $\varphi^{-1}$ as its generator (see Rogge and Schonbucher, 2003, for details). This

<table>
<thead>
<tr>
<th>Copula Name</th>
<th>Generator (Laplace Transform)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clayton</td>
<td>$\varphi(s) = (1 + s)^{-1/\theta}$, $\theta &gt; 0$</td>
</tr>
<tr>
<td>Gumbel</td>
<td>$\varphi(s) = \exp(-s^{1/\theta})$, $\theta &gt; 1$</td>
</tr>
<tr>
<td>Frank</td>
<td>$\varphi(s) = -\frac{1}{\theta} \ln[1 - \exp(-s)[1 - \exp(-\theta)]]$, $\theta \neq 0$</td>
</tr>
<tr>
<td>LTE</td>
<td>$\varphi(s) = (1 + s^{1/\delta})^{-1/\theta}$</td>
</tr>
<tr>
<td>LTF</td>
<td>$\varphi(s) = (1 + \delta^{-1}\ln(1 + s))^{-1/\theta}$, $\delta &gt; 0, \theta \geq 1$</td>
</tr>
<tr>
<td>LTG</td>
<td>$\varphi(s) = \exp[1 - \frac{1}{\delta^{\frac{1}{\theta}}}\ln(1 + s)]^{1/\theta}$, $\delta &gt; 0, \theta \geq 1$</td>
</tr>
<tr>
<td>LTH</td>
<td>$\varphi(s) = 1 - [1 - \exp(-s^{1/\delta})]^{-1/\theta}$, $\delta &gt; 0, \theta \geq 1$</td>
</tr>
<tr>
<td>LTI</td>
<td>$\varphi(s) = 1 - [1 - (1 + s)^{-1/\delta}]^{-1/\theta}$, $\delta &gt; 0, \theta \geq 1$</td>
</tr>
</tbody>
</table>

The names LTE, LTF, LTG, LTH, and LTI are derived from the naming system used by Joe (1997).
provides a fast and efficient method for simulating realizations, one that is not mentioned by Nelsen (1999).\textsuperscript{2} At this point we diverge from their approach. We let $Y(t)$ be a stochastic process that represents the state of the economy, and so the $V_i$ become stochastic processes, $V_i(t)$. Provided the $U_i(t)$ are mutually independent and independent of $Y(t)$, then the static copula of the $V_i(t)$ is the Archimedean copula given in (3). $U_i(t)$ can be interpreted as the prior probability of default, which is then updated by the current state of the economy $Y(t)$. So $V_i(t)$ conditional on the realization of $Y(t)$ is the conditional prior probability of default, since the Laplace transform computes the expectation depending on the distribution of $Y(t)$.

The specific properties of $V_i(t)$ and $V_i(t + \delta_t)$, and of $V_i(t)$ and $V_j(t + \delta_t)$ for $i \neq j$ depend on the way the $U_i(t)$ are constructed (see Totouom & Armstrong, 2005, for details).

### 3. SPECIFIC DYNAMIC ARCHIMEDEAN COPULA PROCESS

First we construct a new type of compound gamma process $Y(t)$ conditional on an underlying gamma process $\alpha(t)$. As usual $\alpha(0) = 0$. For $t > 0$, its increments are independent gammas:\textsuperscript{3}

$$\alpha(t + \delta t) - \alpha(t) \equiv \Gamma(a_1 \times \delta t, a_2) \quad (4)$$

The parameters $a_1$ and $a_2$ are constant over time. For $t > 0$, $\alpha(t)$ has the gamma distribution: $\Gamma(a_1 t, a_2)$.

The values of $Y(t)$ are drawn from the gamma distribution: $\Gamma(\alpha(t), \beta(t))$ where $\beta(t)$ is a strictly positive, deterministic function of time. There are two obvious choices: $\beta(t) = 1$ and $\beta(t) = 1/t$. While the first one leads to a Levy process, the second does not. To the best of our knowledge, this process has not been studied before. In the next section we compute the Laplace transform of $Y(t)$, and hence its moments. The mean and variance of the two processes given in Table 2 will be used later when calibrating the model to market data.

#### Table 2. Moments of the Processes $\alpha(t)$ and $Y(t)$.  

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha(t)$</td>
<td>$a_1 a_2 t$</td>
<td>$a_1 a_2 t$</td>
</tr>
<tr>
<td>$Y(t)$</td>
<td>$a_1 a_2 t \beta(t)$</td>
<td>$a_1 a_2 t \beta(t)[1 + a_2 \beta(t)]$</td>
</tr>
</tbody>
</table>
3.1. Laplace Transform of $Y(t)$

As the process $Y(t)$ is always positive, its Laplace transform is given by:

$$\varphi_s = E \left[ \exp[-s \times Y(t)] \right] = E \left[ E_{\alpha(t)} \left[ \exp[-s \times Y(t)] \right] \right]$$  \hspace{1cm} (5)

We first compute the conditional Laplace transform of $Y(t)$ given $\alpha(t)$.

$$\varphi_s(\alpha(t)) = (1 + s \beta(t))^{-a(t)} = \exp[-\lambda a(t)]$$  \hspace{1cm} (6)

Deconditioning over all values of $\alpha(t)$ gives the Laplace transform of $Y(t)$:

$$\varphi_s = (1 + a_2 \ln(1 + s \beta(t)))^{-a_1 t}$$

$$= \exp[-a_1 t \ln(1 + a_2 \ln(1 + s \beta(t)))]$$  \hspace{1cm} (7)

For any given time $t$, the associated static copula is not a standard Clayton copula but it has the same type of lower tail dependence (Fig. 1). For a better name we call it an extended Clayton copula. The shape can be interpreted as follows. When $Y(t)$ takes low values, the values of the $V_i(t)$ will be low and hence correlated. If one of the names defaults, others are likely to follow suit. Conversely, when $Y(t)$ takes high values, the $V_i(t)$ will be weakly correlated. So if one name defaults the others are unlikely to do so. So this dynamic copula process effectively reproduces what one would intuitively expect.

3.2. Simulating $V_i(t)$

A simple three-step procedure is used for simulating $V_i(t)$

(a) Simulate the process $\alpha(t)$

- Initialize $\alpha(0)$ to 0
- For any $t > 0$ and $\delta t > 0$, simulate an increment

$$\alpha(t + \delta t) - \alpha(t) \equiv \Gamma(a_1 \times \delta t, a_2)$$

- Compute $\alpha(t + \delta t)$

(b) Simulate the compound gamma process $Y(t)$

- At time $t > 0$, draw a value of $Y(t)$ with the conditional gamma distribution

$$\Gamma(\alpha(t), \beta(t))$$
• The values at different times, \( Y(t_1) \) and \( Y(t_2) \), are drawn conditional on the values of the underlying process, \( a(t_1) \) and \( a(t_2) \), but otherwise independent of each other. This adds random noise around \( a(t) \).

(c) Simulate the \( U_i(t) \) then deduce the \( V_i(t) \)
• For each of the \( N \) realizations of \( Y(t) \) simulate \( NU_i(t) \) where \( N \) is the number of names in the portfolio.

(d) In a case of a bullet exposure, the default time can be estimated as in a classical static copula, otherwise a barrier has to be calibrated such that the cumulative probability for \( V_i(t) \) crossing the barrier equals the cumulative default probability of the name \( i \). The default then happened when the barrier threshold is breached.

3.3. Asymptotic Loss Distribution with Bullet Exposure

Assume that the credit portfolio consists of \( N \) underlying credits whose notional are \( P_i = P/N \), with fixed recovery rates \( R_i = R \), \((i = 1, \ldots, N)\). The aggregate loss from today to time \( t \) is a fixed sum of random variables:

\[
\text{Loss}_N(t) = \sum_{i=1}^{N} (1 - R_i)P_i 1_{[\tau_i < t]} = \frac{(1 - R)P}{N} \sum_{i=1}^{N} 1_{[\tau_i < t]} \tag{8}
\]

where \( 1_{[\tau_i < t]} \) is the indicator function for the default of the \( i \)th name. Its Laplace transform is

\[
E\{\exp(-s \text{Loss}_N(t))\} = E\left\{\exp\left(-\frac{sP(1 - R)}{N} \sum_{i=1}^{N} 1_{[\tau_i < t]}\right)\right\}
\]

\[
= E\left\{\prod_{i=1}^{N} \{1 - 1_{[\tau_i < t]} + 1_{[\tau_i < t]} \eta^{1/N}\}\right\} \tag{9}
\]

Letting \( \eta = \exp(-sN(1 - R)) > 0 \)

\[
E\{\exp(-s\text{Loss}_N(t))\} = E\{[(1 - \eta^{1/N}) \exp(-Y_i \varphi_i^{-1}(\text{PD}(t))) + 1]^N\}
\]

We now compute its limit as \( N \) the number of names tends to infinity. Since

\[
\lim_{N \to +\infty} \{N(\eta^{1/N} - 1)\} = \lim_{n \to +\infty} \left\{\left(\frac{\eta^{1/N} - 1}{1/N}\right)\right\} = \left. \frac{\partial \eta^x}{\partial x}\right|_{x=0} = \ln(\eta) = -sN(1 - R)
\]
we obtain

$$\text{Loss}_\mathcal{N}(t) \approx P(1 - R) \exp(-Y_i \varphi_i^{-1}(PD(t)))$$

(10)

### 3.4. Evolution of $V_i(t)$ Over Time

In this section we compute the bivariate distribution function of $V_i(t)$ and $V_i(t + \delta t)$ for two different cases. To simplify the notation, let

$$K_{V_i(t), V_i(t + \delta t)}(a, b) = \text{Prob}(V_i(t) < a, V_i(t + \delta t) < b)$$

$$H_{U_i(t), U_i(t + \delta t)}(a, b) = \text{Prob}(U_i(t) < a, U_i(t + \delta t) < b)$$

(11)

Note that these can also be viewed as integral transforms. In the multi-period case, we extend this notation in the obvious way:

$$K_{V_i(t), V_i(t + k\delta t), ..., V_i(t + n\delta t)}(v_i, ..., v_i^{t + k\delta t}, ..., v_i^{t + n\delta t})$$

$$= \text{Prob}(V_i(t) < v_i, ..., V_i(t + k\delta t) < v_i^{t + k\delta t}, ..., V_i(t + n\delta t) < v_i^{t + n\delta t})$$

By conditioning on the values of $Y(t)$ and $Y(t + \delta t)$ and noting that

$$\varphi\left(-\frac{\ln(U_i)}{Y_i}\right) \leq V_i \leq \exp\left(-Y_i \varphi^{-1}(v_i)\right)$$

it is easy to show that

$$\text{Prob}(V_i(t) < w, V_i(t + \delta t) < z| Y(t), Y(t + \delta t))$$

$$= H_{U_i(t), U_i(t + \delta t)}(e^{-\varphi^{-1}(w)Y_i(t)}, e^{-\varphi^{-1}(z)Y_i(t + \delta t)})$$

(12)

Hence

$$K_{V_i(t), V_i(t + \delta t)}(w, z) = E[H_{U_i(t), U_i(t + \delta t)}(e^{-\varphi^{-1}(w)Y_i(t)}, e^{-\varphi^{-1}(z)Y_i(t + \delta t)})]$$

(13)

**Case 1. One-time step analysis**

We assume $U_i(t)$ and $U_i(t + \delta t)$ are independent and that $Y(t)$ is a stochastic process with independent identically distributed increments. Because of the independence and because the process $U(t)$ is uniform on $[0, 1]$

$$H_{U_i(t), U_i(t + \delta t)}(a, b) = \text{Prob}(U_i(t) < a) \times \text{Prob}(U_i(t + \delta t) < b) = a \times b$$
If we let $1_{\delta t>0}$ be an indicator function that takes the value 1 if $\delta t>0$, and 0 otherwise, then

$$H_{U_i(t), U_i(t+\delta t)}(u_i, u_i^{t+\delta t}) = 1_{\delta t>0} \times u_i^{t+\delta t} + (1 - 1_{\delta t>0}) \times \text{Min}(u_i', u_i^{t+\delta t})$$

Consequently,

$$K_{V_i(t), V_i(t+\delta t)}(w, z) = E\left[1_{\delta t>0} \times e^{-[\varphi_t^{-1}(w)Y(t) + \varphi_{t+\delta t}^{-1}(z)Y(t+\delta t)]} + (1 - 1_{\delta t>0}) \times e^{-Y(t)\text{Max}[\varphi_t^{-1}(w), \varphi_{t+\delta t}^{-1}(z)]}\right]$$

(14)

Since $Y(t)$ is a stochastic process with independent identically distributed increments

$$K_{V_i(t), V_i(t+\delta t)}(w, z) = \left\{\begin{array}{ll}
1_{\delta t>0} \times E\left[e^{-[\varphi_t^{-1}(w) + \varphi_{t+\delta t}^{-1}(z)]Y(t)}\right] \times E\left[e^{-\varphi_{t+\delta t}^{-1}(z)[Y(t+\delta t) - Y(t)]}\right] + \\
(1 - 1_{\delta t>0}) \times E\left[e^{-Y(t)\text{Max}[\varphi_t^{-1}(w), \varphi_{t+\delta t}^{-1}(z)]}\right]
\end{array}\right\}$$

(15)

Since $\varphi_t$ is a decreasing function, this simplifies to

$$K_{V_i(t), V_i(t+\delta t)}(w, z) = \left\{\begin{array}{ll}
1_{\delta t>0} \times \varphi_t(\varphi_t^{-1}(w) + \varphi_{t+\delta t}^{-1}(z)) \times \varphi_{t+\delta t}(\varphi_{t+\delta t}^{-1}(z)) + \\
(1 - 1_{\delta t>0}) \times \text{Min}(w, z)
\end{array}\right\}$$

(16)

If $\delta t=0$

$$K_{V_i(t), V_i(t+0)}(w, z) = \text{Min}[w, z]$$

As $\delta t \to 0$

$$K_{V_i(t), V_i(t+\delta t)}(w, z) \to \varphi_t(\varphi_t^{-1}(w) + \varphi_t^{-1}(z)) \leq K_{V_i(t), V_i(t+0)}(w, z) = \text{Min}[w, z]$$

(17)

Fig. 2 shows how the joint probability that $V_i(t)$ is less than $w$ and $V_i(t+\delta t)$ is less than $z$, evolves as a function of $t$, especially as $\delta t \to 0$. Note the discontinuity at zero.
Case 2. Multi-time step analysis

As before, we assume $U_i(t)$ and $U_i(t + \delta t)$ are independent and that $Y(t)$ is a stochastic process with independent identically distributed increments. Because of the independence,

$$
H_{U_i(t), ..., U_i(t + k\delta t), ..., U_i(t + n\delta t)}(u_i^t, \ldots, u_i^{t+k\delta t}, \ldots, u_i^{t+n\delta t})
= 1_{\delta t > 0} \times \prod_{k=0}^{n} u_i^{t+k\delta t} + (1 - 1_{\delta t > 0}) \times \text{Min}[u_i^t, \ldots, u_i^{t+k\delta t}, \ldots, u_i^{t+n\delta t}]
$$

$$
K_{V_i(t), ..., V_i(t + k\delta t), ..., V_i(t + n\delta t)}(v_i^t, \ldots, v_i^{t+k\delta t}, \ldots, v_i^{t+n\delta t})
= E \left[ 1_{\delta t > 0} \times \exp \left( - \sum_{k=0}^{n} \varphi_{t+k\delta t}^{-1}(v_i^{t+k\delta t}) Y(t + k\delta t) \right) \right] + \left( 1 - 1_{\delta t > 0} \right) \times \exp(-Y(t)\text{Max}[\varphi_{t}^{-1}(v_i^t), \ldots, \varphi_{t+n\delta t}^{-1}(v_i^{t+n\delta t})])
$$

Remark. Since

$$
\sum_{k=0}^{n} \varphi_{t+k\delta t}^{-1}(v_i^{t+k\delta t}) Y(t + k\delta t) = Y(t) \times \sum_{k=0}^{n} \varphi_{t+k\delta t}^{-1}(v_i^{t+k\delta t})
+ \sum_{m=1}^{n} \{ (Y(t + k\delta t) - Y(t)) \times \varphi_{t+k\delta t}^{-1}(v_i^{t+k\delta t}) \}
$$
we obtain

\[ K_{V_i(t), V_i(t+k\delta t), \ldots, V_i(t+n\delta t)}(v_i^t, \ldots, v_i^{t+k\delta t}, \ldots, v_i^{t+n\delta t}) \]

\[ = E \left[ 1_{\delta t > 0} \times \varphi_i \left( \sum_{k=0}^{n} \varphi_{t+k\delta t}(v_i^{t+k\delta t}) \right) \prod_{m=1}^{n} \varphi_{k\delta t}(\varphi_{t+k\delta t}(v_i^{t+k\delta t})) \right] + (1 - 1_{\delta t > 0}) \times \text{Min}[v_i^t, \ldots, v_i^{t+k\delta t}, \ldots, v_i^{t+n\delta t}] \] 

(19)

4. PRICING OF A CORRELATION PRODUCT: CDO

Pricing synthetic CDOs involves computing aggregate loss distributions over different time horizons. So CDO tranche premiums depend upon the individual credit risk of names in the underlying portfolio and the dependence structure between default times.

4.1. Notation and Definitions

\( i = 1, \ldots, N \): Single name credits in the base portfolio for CDO pricing

\( \tau_1, \tau_2, \ldots, \tau_N \): Default times

\( \text{LGD}_i \): Loss given default on the name \( i \)

\( \text{PD}_i(t) \): Defines a cumulative default probability of on the name \( i \) at time \( t \)

\( N_i \): Nominal of the name \( i \)

The aggregated loss in the portfolio at time \( t \) is given by:

\[ \text{Loss}(t) = \sum_{i=1}^{N} N_i \text{LGD}_i 1_{[\tau_i < t]} \]

(20)

If \( K_u \) and \( K_d \) are the upper and lower detachment points, the loss in the tranche \([K_d, K_u]\) at time \( t \) is

\[ \text{Loss}_t(K_d, K_u) = \text{Min}[K_u, \text{Loss}(t)] - \text{Min}[K_d, \text{Loss}(t)] \]

(21)

The Expected Loss (EL) in the base tranche \([0, K]\) at time \( t \) is just:

\[ \text{EL} = E[\text{Min}[K, \text{Loss}(t)]] \]

(22)
Having analytic expressions for the expected loss makes it easy to compute the Greeks for the portfolio.

4.2. Data Source

We used market data (Anonymous, 2005) as of July 22, 2005 (see Table 3). Fig. 3 shows the base correlation as a function of the detachment point. The next step was to compute the cleanspreads and the default probabilities for the 125 names, for different horizons: 1, 3, 5, 7, or 10 year(s), from the second spreadsheet. A constant loss given default of 60% was assumed on all names.

The cumulative default probability at any time horizon was computed as follows:

\[
\text{cleanSpread}_i(\text{Horizon}) = \frac{\text{Spread}_i(\text{Horizon})}{\text{LGD}_i} \tag{23}
\]

\[
\text{PD}_i(\text{Horizon}) = 1 - \exp\left[-\frac{\text{cleanSpread}_i(\text{Horizon})}{10,000} \times \text{Horizon}\right]
\]

The average 5-year default probability in the portfolio is 4.30%. Table 4 gives the summary statistics of default probabilities at a 5-year horizon.

Table 3. Attachment and Detachment Points for the Market Data for July 22, 2005.

<table>
<thead>
<tr>
<th>Attachment (%)</th>
<th>Detachment (%)</th>
<th>Correlation (%)</th>
<th>26 July</th>
<th>25 July</th>
<th>18 July</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>3</td>
<td>12.08</td>
<td>12.3</td>
<td>11.4</td>
<td>11.6</td>
</tr>
<tr>
<td>0</td>
<td>7</td>
<td>33.8</td>
<td>32.9</td>
<td>32.3</td>
<td>33.6</td>
</tr>
<tr>
<td>0</td>
<td>10</td>
<td>44.14</td>
<td>43.2</td>
<td>42.5</td>
<td>44.1</td>
</tr>
<tr>
<td>0</td>
<td>15</td>
<td>58.14</td>
<td>56.2</td>
<td>55.4</td>
<td>57.2</td>
</tr>
<tr>
<td>0</td>
<td>30</td>
<td>79.78</td>
<td>80.2</td>
<td>78.8</td>
<td>80.5</td>
</tr>
<tr>
<td>$\text{Index not Extracted}$</td>
<td>$\text{Index EL}$</td>
<td>$\text{Index (bps)}$</td>
<td>$\text{Index (bps)}$</td>
<td>$\text{Index (bps)}$</td>
<td>$\text{Index (bps)}$</td>
</tr>
<tr>
<td>$\text{$125,000$}$</td>
<td>$\text{$2,933.69$}$</td>
<td>53</td>
<td>53</td>
<td>53</td>
<td>56</td>
</tr>
</tbody>
</table>

Extracted from Excel file on Wilmott website, together with the correlation expressed as a percentage.
4.3. Calibrating the Parameters for Monte Carlo Pricing of the CDO

A simple iterative procedure was used to calibrate the parameters of the gamma distribution. The base correlation was computed by running Monte Carlo simulations of the portfolio and comparing this with the market base correlation. Typically 10,000 simulations were carried out. For simplicity all the exposures are bullet. Further work will be needed to improve the calibration procedure.

The risk-free rate shown in Fig. 4 was used within the model and within the Gaussian copula model to obtain the base correlation but the risk-free rate has little or very few impact on the pricing since the asset and the liability of the CDO are floating rate instrument, only default matters. This choice has no impact on the base correlation. So the result is the same as if we assumed that it was zero as does JP Morgan (see McGinty & Ahulwalia, 2004).

![Table 4. Summary Statistics of 125 5-Year Default Probabilities.](image)

<table>
<thead>
<tr>
<th>Minimum</th>
<th>0.09%</th>
<th>Median</th>
<th>3.05%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum</td>
<td>36.80%</td>
<td>Mode</td>
<td>2.26%</td>
</tr>
<tr>
<td>Mean</td>
<td>4.30%</td>
<td>Skewness</td>
<td>4.30</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>4.78%</td>
<td>Kurtosis</td>
<td>22.13</td>
</tr>
</tbody>
</table>
The parameters were calibrated for a maturity of 5 years because this is the most liquid. There is no unique optimum. Three possible sets of values are shown in Table 5. The resulting base correlations are compared to those for the market, at the standard detachment points. As different parameter values give comparable base correlations for this maturity, other maturities should be used to choose the most appropriate set overall. Table 6 shows the term structure for the equity tranche for the standard maturities (5, 7, and 10 years) for the same sets of parameters.

Fig. 5 presents the base correlation as a function of the detachment point, for the four maturities (3, 5, 7, and 10 years) for different values of the second parameter $a_2$. Note how the convexity of the curve changes with the maturity. The model can produce convex curves as well as concave ones.

<table>
<thead>
<tr>
<th>Detachment Points</th>
<th>Market</th>
<th>Set 1</th>
<th>Set 2</th>
<th>Set 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>3%</td>
<td>12%</td>
<td>16.6%</td>
<td>15.5%</td>
<td>17.9%</td>
</tr>
<tr>
<td>7%</td>
<td>34%</td>
<td>28.7%</td>
<td>28.3%</td>
<td>30.6%</td>
</tr>
<tr>
<td>10%</td>
<td>44%</td>
<td>38.9%</td>
<td>38.8%</td>
<td>41.0%</td>
</tr>
<tr>
<td>15%</td>
<td>58%</td>
<td>52.9%</td>
<td>52.8%</td>
<td>55.2%</td>
</tr>
<tr>
<td>30%</td>
<td>80%</td>
<td>80.7%</td>
<td>80.7%</td>
<td>82.7%</td>
</tr>
</tbody>
</table>

$\begin{bmatrix}
0.5 & 5 \\
0.5 & 90
\end{bmatrix}$

Table 5. The Base Correlations Computed from the Model Using Three Sets of Parameter Estimates for the Process $\alpha(t)$, Together with the Market Values. The Parameters are Shown Below.
Table 6. The Term Structure (i.e., the Base Correlation for the Equity Tranche for the Same Parameters).

<table>
<thead>
<tr>
<th>Maturity</th>
<th>Set 1</th>
<th>Set 2</th>
<th>Set 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 years</td>
<td>46.1%</td>
<td>44.1%</td>
<td>49.2%</td>
</tr>
<tr>
<td>5 years</td>
<td>16.6%</td>
<td>15.5%</td>
<td>17.9%</td>
</tr>
<tr>
<td>7 years</td>
<td>9.3%</td>
<td>8.6%</td>
<td>9.9%</td>
</tr>
<tr>
<td>10 years</td>
<td>6.1%</td>
<td>5.9%</td>
<td>6.6%</td>
</tr>
</tbody>
</table>

Fig. 5. The Base Correlation as a Function of the Detachment Point for the Four Standard Maturities: 3 Years (Top Left), 5 Years (Top Right), 7 Years (Lower Left), and 10 Years (Lower Right). Note That the Change in the Convexity with Maturity.
Fig. 6. The Term Structure of the Equity Tranche for Different Maturities; on the Left, for a Fixed Value of the First Parameter $a_1$, on the Right, for a Fixed Value of the Second Parameter $a_2$.

Fig. 6 illustrates the impact of these two ratios on the terms structure. In both cases, increasing one parameter for a fixed value of the other one leads to a decrease in the base correlation of the equity tranche.

5. CONCLUSIONS

In this chapter we have chosen to model default probabilities dependency explicitly rather than intensities, and have developed a new class of dynamic copula processes, based on the well-known relation between Archimedean copulas and Laplace transforms:

$$V_i = \varphi\left(\frac{-\ln(U_i)}{Y}\right) \quad \text{for } i = 1, \ldots, n$$

Replacing the random variables $Y$ and $U_i$, by suitably chosen processes $Y(t)$ and $U_i(t)$, provides a simple way of constructing and simulating a wide range of dynamic copula processes. This framework based on conditional independence effectively overcomes the difficulties of constructing multivariate copulas that have been well documented in the literature on copulas (Nelsen, 1999; Joe, 1997). The difficulties in the standard multivariate copulas construction is the complexity of the computation.

After presenting the procedure for simulating this class of copula processes (Section 2), we focus on a particular case: where $Y(t)$ is a new type of compound gamma process, because this gives rise to a dynamic
process in which the copulas have lower tail dependence but not upper tail dependence. As we use $Y(t)$ to represent the current economic climate, this means that defaults are correlated in unfavorable times but not during normal times, as one would intuitively expect. The $U_i(t)$ can be interpreted as the prior probability of default, which is updated given the state of the economy to obtain the posterior probability of default $V_i(t)$.

In Section 4, we use market data to calibrate the model. We show that the model reproduces the base correlations observed at that time. We have also studied the types of term structure given by the model. One advantage of this approach compared to those based on default intensities is that it provides a simple way of computing base correlations without having to specify or calibrate the marginal densities, but its primary strong point is that it provides a mathematically consistent framework for modeling the structure of defaults over different time horizons.

NOTES

1. To avoid confusion, note that in this chapter we use $\varphi$ to denote a Laplace transform, and $f$ for the generator of an Archimedean copula, whereas Nelsen (1999) uses $\varphi$ for the generator of an Archimedean copula. The function $f$ must be a continuous and strictly decreasing.


3. To make it simpler to use standard software to do the simulations, we have changed to the standard notation: if the random variable $X$ has the gamma distribution $\Gamma(a, b)$, its density is

$$f(X) = \frac{X^{a-1}}{\Gamma(a)b^a} e^{-X/b} \quad X \geq 0$$

REFERENCES


PERTURBED GAUSSIAN COPULA

Jean-Pierre Fouque and Xianwen Zhou

ABSTRACT

Gaussian copula is by far the most popular copula used in the financial industry in default dependency modeling. However, it has a major drawback – it does not exhibit tail dependence, a very important property for copula. The essence of tail dependence is the interdependence when extreme events occur, say, defaults of corporate bonds. In this paper, we show that some tail dependence can be restored by introducing stochastic volatility on a Gaussian copula. Using perturbation methods we then derive an approximate copula – called perturbed Gaussian copula in this paper.

A copula is a joint distribution function of uniform random variables. Sklar’s Theorem states that for any multivariate distribution, the univariate marginal distributions and the dependence structure can be separated. The dependence structure is completely determined by the copula. It then implies that one can “borrow” the dependence structure, namely the copula, of one set of dependent random variables and exchange the marginal distributions for a totally different set of marginal distributions.

An important property of copula is its invariance under monotonic transformation. More precisely, if $g_i$ is strictly increasing for each $i$, then $(g_1(X_1), g_2(X_2), \ldots, g_n(X_n))$ have the same copula as $(X_1, X_2, \ldots, X_n)$.
From the above discussion, it is not hard to see that copula comes in default dependency modeling very naturally. For a much detailed coverage on copula, including the precise form of Sklar’s Theorem, as well as modeling default dependency by way of copula, the readers are referred to Schonbucher (2003).

Let \((Z_1, \ldots, Z_n)\) be a normal random vector with standard normal marginals and correlation matrix \(R\), and \(\Phi(\cdot)\) be the standard normal cumulative distribution function. Then the joint distribution function of \((\Phi(Z_1), \ldots, \Phi(Z_n))\) is called the Gaussian copula with correlation matrix \(R\).

Gaussian copula is by far the most popular copula used in the financial industry in default dependency modeling. This is basically because of two reasons. First, it is easy to simulate and second, it requires the “right” number of parameters – equal to the number of correlation coefficients among the underlying names. However, Gaussian copula does not exhibit any tail dependence, a very important property for copula (we refer to Carmona, 2004, for a detailed analysis of tail dependence). Tail dependence, which is roughly the interdependence when extreme events occur, is a desirable feature of a copula when modeling for instance defaults of corporate bonds. In fact, the lack of it is considered as a major drawback of Gaussian copula.

On the other hand, by introducing stochastic volatility into the classic Black-Scholes model, Fouque, Papanicolaou, and Sircar (2000), by way of singular perturbation method, gave a satisfactory answer to the “smile curve” problem of implied volatilities in the financial market, leading to a pricing formula which is in the form of a robust simple correction to the classic Black-Scholes constant volatility formula. Furthermore, an application of this perturbation method to defaultable bond pricing has been studied by Fouque, Sircar, and Solna (2006). By fitting real market data, they concluded that the method works fairly well. An extension to multi-name first passage models is proposed by Fouque, Wignall, and Zhou (2008).

In this paper, we will show the effect of stochastic volatility on a Gaussian copula. Specifically, in Section 1, we first set up the stochastic volatility model and state the objective – the transition density functions. Then by singular perturbation, we obtain approximate transition density functions. In order to make them true probability density functions, we introduce the transformation \(1 + \tanh(\cdot)\). In Section 2, we study this new class of approximate copula density functions, first analytically and then numerically. Section 3 concludes this paper.
1. ASYMPTOTICS

1.1. Model Setup

A two-dimensional Gaussian copula can be generated by a pair of correlated Brownian motions. We propose to “create” additional correlation through a common process driving their diffusion coefficients. For that, we start with a process $(X_t^{(1)}, X_t^{(2)}, Y_t)$ defined on the complete probability space $(\Omega, \mathcal{F}, P)$ and which follows the dynamics:

\[
\begin{align*}
  dX_t^{(1)} &= f_1(Y_t) dW_t^{(1)} \\
  dX_t^{(2)} &= f_2(Y_t) dW_t^{(2)} \\
  dY_t &= \frac{1}{\varepsilon} (m - Y_t) dt + \frac{\sqrt{2}}{\sqrt{\varepsilon}} dW_t^{(Y)}
\end{align*}
\]

where $W_t^{(1)}, W_t^{(2)},$ and $W_t^{(Y)}$ are standard Brownian motions correlated as follows:

\[
\begin{align*}
  d\langle W^{(1)}, W^{(2)} \rangle_t &= \rho dt, \\
  d\langle W^{(1)}, W^{(Y)} \rangle_t &= \rho_{1Y} dt, \\
  d\langle W^{(2)}, W^{(Y)} \rangle_t &= \rho_{2Y} dt
\end{align*}
\]

with $-1 \leq \rho, \rho_{1Y}, \rho_{2Y} \leq 1$ and making the correlation matrix symmetric positive definite, $\varepsilon$ and $\nu$ are positive constant numbers with $\varepsilon \ll 1$. The $f_i$’s are real functions for $i = 1, 2,$ and are assumed here to be bounded above and below away from 0. It is worth noting that $f_i$’s are not explicit functions of $t$. They depend on $t$ only through $Y_t$.

Observe that $Y_t$ is a mean-reverting process and $1/\varepsilon$ is the rate of mean-reversion so that $Y_t$ is fast mean-reverting. Furthermore, $Y_t$ admits the unique invariant normal distribution $\mathcal{N}(m, \nu^2)$.

For a fixed time $T > 0,$ our objective is to find, for $t < T,$ the joint distribution

\[
\mathbb{P} \left\{ X_T^{(1)} \leq \xi_1, X_T^{(2)} \leq \xi_2 | X_t = x, Y_t = y \right\}
\]

and the two marginal distributions

\[
\mathbb{P} \left\{ X_T^{(1)} \leq \xi_1 | X_t = x, Y_t = y \right\}, \quad \mathbb{P} \left\{ X_T^{(2)} \leq \xi_2 | X_t = x, Y_t = y \right\}
\]
where $X_t \equiv (X^{(1)}_t, X^{(2)}_t)$, $x \equiv (x_1, x_2)$, and $\xi_1, \xi_2$ are two arbitrary numbers. Equivalently, we need to find the following three transition densities:

$$u^e \equiv \mathbb{P}\left\{ X^{(1)}_T \in d\xi_1, X^{(2)}_T \in d\xi_2 | X_t = x, Y_t = y \right\}$$

$$v^e_1 \equiv \mathbb{P}\left\{ X^{(1)}_T \in dx_1 | X_t = x, Y_t = y \right\}$$

$$v^e_2 \equiv \mathbb{P}\left\{ X^{(2)}_T \in dx_2 | X_t = x, Y_t = y \right\}$$

where we show the dependence on the small parameter $\varepsilon$. Indeed $v^e_1$ and $v^e_2$ can be obtained from $u^e$ by integration.

### 1.2. PDE Representation

Let us consider $u^e$ first. In terms of partial differential equation (PDE), $u^e$ satisfies the following Kolmogorov backward equation

$$\mathcal{L}^e u^e(t, x_1, x_2, y) = 0$$

$$u^e(T, x_1, x_2, y) = \delta(\xi_1; x_1)\delta(\xi_2; x_2)$$

where $\delta(\xi_i; x_i)$ is the Dirac delta function of $x_i$ with spike at $x_i = \xi_i$ for $i = 1, 2$, and operator $\mathcal{L}^e$ has the following decomposition:

$$\mathcal{L}^e = \frac{1}{\varepsilon} \mathcal{L}_0 + \frac{1}{\sqrt{\varepsilon}} \mathcal{L}_1 + \mathcal{L}_2$$

with the notations:

$$\mathcal{L}_0 = (m - y) \frac{\partial}{\partial y} + v^2 \frac{\partial^2}{\partial y^2}$$

$$\mathcal{L}_1 = v\sqrt{2}\rho_1 f_1(y) \frac{\partial^2}{\partial x_1^2} + v\sqrt{2}\rho_2 f_2(y) \frac{\partial^2}{\partial x_2^2}$$

$$\mathcal{L}_2 = \frac{\partial}{\partial t} + \frac{1}{2} f_1^2(y) \frac{\partial^2}{\partial x_1^2} + \frac{1}{2} f_2^2(y) \frac{\partial^2}{\partial x_2^2} + \rho f_1(y)f_2(y) \frac{\partial^2}{\partial x_1 \partial x_2}$$

As in Fouque et al. (2000), we expand the solution $u^e$ in powers of $\sqrt{\varepsilon}$:

$$u^e = u_0 + \sqrt{\varepsilon} u_1 + \varepsilon u_2 + \varepsilon^{3/2} u_3 + \cdots$$
In the following, we will determine the first few terms appearing on the right-hand side of the above expansion. Specifically, we will retain

\[ \tilde{u} \equiv u_0 + \sqrt{\varepsilon}u_1 \]  

as an approximation to \( u^\varepsilon \) (later we will propose another approximation in order to restore positiveness).

1.3. Leading Order Term \( u_0 \)

Following Fouque et al. (2000), the leading order term \( u_0 \), which is independent of variable \( y \), is characterized by:

\[ \langle \mathcal{L}_2 \rangle u_0(t, x_1, x_2) = 0 \]

\[ u_0(T, x_1, x_2) = \delta(\xi_1; x_1)\delta(\xi_2; x_2) \]  

where \( \langle \cdot \rangle \) denotes the average with respect to the invariant distribution \( \mathcal{N}(m, \nu^2) \) of \( Y_t \), that is,

\[ \langle g \rangle \equiv \int_{-\infty}^{\infty} g(y) \frac{1}{\nu \sqrt{2\pi}} \exp \left\{ -\frac{(y - m)^2}{2\nu^2} \right\} dy \]

for a general function \( g \) of \( y \).

We define the effective volatilities \( \bar{\sigma}_1 \) and \( \bar{\sigma}_2 \) and the effective correlation \( \bar{\rho} \) by:

\[ \bar{\sigma}_1 \equiv \sqrt{\langle f_1^2 \rangle}, \quad \bar{\sigma}_2 \equiv \sqrt{\langle f_2^2 \rangle}, \quad \bar{\rho} \equiv \frac{\rho \langle f_1 f_2 \rangle}{\bar{\sigma}_1 \bar{\sigma}_2} \]  

(6)

Using Eq. (3) and the notations (6), Eq. (5) becomes

\[ \frac{\partial u_0}{\partial t} + \frac{1}{2} \bar{\sigma}_1^2 \frac{\partial^2 u_0}{\partial x_1^2} + \frac{1}{2} \bar{\sigma}_2^2 \frac{\partial^2 u_0}{\partial x_2^2} + \bar{\rho} \bar{\sigma}_1 \bar{\sigma}_2 \frac{\partial^2 u_0}{\partial x_1 \partial x_2} = 0 \]

\[ u_0(T, x_1, x_2) = \delta(\xi_1; x_1)\delta(\xi_2; x_2) \]

It can be verified that \( u_0 \) is the transition density of two correlated scaled Brownian motions with instantaneous correlation \( \bar{\rho} \) and scale factors \( \bar{\sigma}_1 \) and \( \bar{\sigma}_2 \), respectively. That is,

\[ u_0(t, x_1, x_2) = \frac{1}{2\pi \bar{\sigma}_1 \bar{\sigma}_2 (T - t)^{1/2} \sqrt{1 - \bar{\rho}^2}} \times \exp \left\{ -\frac{1}{2(1 - \bar{\rho}^2)} \left[ \frac{(\xi_1 - x_1)^2}{\bar{\sigma}_1^2 (T - t)} - 2\bar{\rho} \frac{(\xi_1 - x_1)(\xi_2 - x_2)}{\bar{\sigma}_1 \bar{\sigma}_2 (T - t)} + \frac{(\xi_2 - x_2)^2}{\bar{\sigma}_2^2 (T - t)} \right] \right\} \]  

(7)
1.4. Correction Term $\sqrt{\delta} u_1$

Again, similar to Fouque et al. (2000), the correction term $u_1$, which is also independent of variable $y$, is characterized by:

$$\langle \mathcal{L}_2 \rangle u_1(t, x_1, x_2) = \mathcal{A} u_0$$
$$u_1(T, x_1, x_2) = 0$$ (8)

where the operator $\mathcal{A}$ is defined by

$$\mathcal{A} = \langle \mathcal{L}_1 \mathcal{L}_0^{-1} (\mathcal{L}_2 - \langle \mathcal{L}_2 \rangle) \rangle$$

and the inverse $\mathcal{L}_0^{-1}$ acts on the centered quantity $\mathcal{L}_2 - \langle \mathcal{L}_2 \rangle$.

From the definition of $\mathcal{L}_2$ given in Eq. (3), it is straightforward to obtain that

$$\mathcal{L}_2 - \langle \mathcal{L}_2 \rangle = \frac{1}{2} (f_1^2(y) - \langle f_1^2 \rangle) \frac{\partial^2}{\partial x_1^2} + \frac{1}{2} (f_2^2(y) - \langle f_2^2 \rangle) \frac{\partial^2}{\partial x_2^2} + \rho (f_1(y) f_2(y)$$

$$- \langle f_1 f_2 \rangle) \frac{\partial^2}{\partial x_1 \partial x_2}$$

Let us denote by $\phi_1(y)$, $\phi_2(y)$, and $\phi_{12}(y)$, the solutions of the following Poisson equations, respectively

$$\mathcal{L}_0 \phi_1(y) = f_1^2(y) - \langle f_1^2 \rangle$$
$$\mathcal{L}_0 \phi_2(y) = f_2^2(y) - \langle f_2^2 \rangle$$
$$\mathcal{L}_0 \phi_{12}(y) = f_1(y) f_2(y) - \langle f_1 f_2 \rangle$$

Their existence (with at most polynomial growth at infinity) is guaranteed by the centering property of the right-hand sides and the Fredholm alternative for the infinitesimal generator $\mathcal{L}_0$. They are defined up to additive constants in $y$, which will play no role after applying the operator $\mathcal{L}_1$, which takes derivatives with respect to $y$. It then follows that

$$\mathcal{L}_0^{-1} (\mathcal{L}_2 - \langle \mathcal{L}_2 \rangle) = \frac{1}{2} \phi_1(y) \frac{\partial^2}{\partial x_1^2} + \frac{1}{2} \phi_2(y) \frac{\partial^2}{\partial x_2^2} + \rho \phi_{12}(y) \frac{\partial^2}{\partial x_1 \partial x_2}$$
Now, by definition of $L_1$ given in Eq. (2), we have

$$L_1 L_0^{-1} (L_2 - \langle L_2 \rangle)$$

$$= \sqrt{2} \rho_1 y f_1(y) \left[ \frac{1}{2} \phi'_1(y) \frac{\partial^3}{\partial x_1^3} + \frac{1}{2} \phi'_2(y) \frac{\partial^3}{\partial x_1 \partial x_2^2} + \rho \phi'_{12}(y) \frac{\partial^3}{\partial x_1 \partial x_2} \right]$$

$$+ \sqrt{2} \rho_2 y f_2(y) \left[ \frac{1}{2} \phi'_1(y) \frac{\partial^3}{\partial x_1^3} + \frac{1}{2} \phi'_2(y) \frac{\partial^3}{\partial x_2^3} + \rho \phi'_{12}(y) \frac{\partial^3}{\partial x_1 \partial x_2} \right]$$

Therefore, the operator $\sqrt{\epsilon} A$ can be written

$$\sqrt{\epsilon} A = R_1 \frac{\partial^3}{\partial x_1} + R_2 \frac{\partial^3}{\partial x_2^3} + R_{12} \frac{\partial^3}{\partial x_1 \partial x_2^2} + R_{21} \frac{\partial^3}{\partial x_1^2 \partial x_2}$$

where the constant parameters $R_1$, $R_2$, $R_{12}$, and $R_{21}$ are defined as follows:

$$R_1 \equiv \frac{\sqrt{2} \rho_1 y \sqrt{\epsilon}}{\sqrt{2}} (f_1 \phi'_1)$$

$$R_2 \equiv \frac{\sqrt{2} \rho_2 y \sqrt{\epsilon}}{\sqrt{2}} (f_2 \phi'_2)$$

$$R_{12} \equiv \frac{\sqrt{2} \rho_1 y \sqrt{\epsilon}}{\sqrt{2}} (f_1 \phi'_2) + \sqrt{2} \epsilon \rho_2 y (f_2 \phi'_{12})$$

$$R_{21} \equiv \frac{\sqrt{2} \rho_2 y \sqrt{\epsilon}}{\sqrt{2}} (f_2 \phi'_1) + \sqrt{2} \epsilon \rho_1 y (f_1 \phi'_{12})$$

Note that they are all small of order $\sqrt{\epsilon}$.

It can be checked directly that $u_1$ is given explicitly by

$$u_1 = -(T - t) A u_0$$

and therefore

$$\sqrt{\epsilon} u_1 = -(T - t) \left[ R_1 \frac{\partial^3}{\partial x_1} + R_2 \frac{\partial^3}{\partial x_2^3} + R_{12} \frac{\partial^3}{\partial x_1 \partial x_2^2} + R_{21} \frac{\partial^3}{\partial x_1^2 \partial x_2} \right] u_0$$

(9)

Explicit formulas for the third-order partial derivatives of $u_0$ are given in appendix.
1.5. Regularity Conditions for Density Functions

Since

\[ 1 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u_0(t, x_1, x_2; \xi_1, \xi_2) d\xi_1 d\xi_2 \]

by Lebesgue dominated convergence theorem, we then have

\[ 0 = \frac{\partial^{k_1+k_2}}{\partial x_1^{k_1} \partial x_2^{k_2}} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\partial^{k_1+k_2}}{\partial x_1^{k_1} \partial x_2^{k_2}} u_0(t, x_1, x_2; \xi_1, \xi_2) d\xi_1 d\xi_2 \]

for integers \( k_1, k_2 \geq 0 \) such that \((k_1, k_2) \neq (0, 0)\). It follows that

\[ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sqrt{\varepsilon} u_1(t, x_1, x_2; \xi_1, \xi_2) d\xi_1 d\xi_2 = 0 \]

and hence

\[ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \tilde{u}(t, x_1, x_2; \xi_1, \xi_2) d\xi_1 d\xi_2 = 1 \]

where \( \tilde{u} \equiv u_0 + \sqrt{\varepsilon} u_1 \) is the approximation introduced in (4).

In order to guarantee that our approximated transition density function is always non-negative, which is the other regularity condition for a density function, we seek a multiplicative perturbation of the form

\[ \tilde{u} = \hat{u}_0(1 + \tan h(\sqrt{\varepsilon} \hat{u}_1)) \]

where \( \hat{u}_0 \) and \( \hat{u}_1 \) are defined such that

\[ u_0 + \sqrt{\varepsilon} u_1 = \hat{u}_0(1 + \sqrt{\varepsilon} \hat{u}_1) \]

for any \( \varepsilon > 0 \). It can be easily seen that this is achieved with the choice:

\[ \hat{u}_0 = u_0, \quad \hat{u}_1 = \frac{u_1}{u_0} \]

Now instead of using \( \tilde{u} \) as our approximation for \( u^\varepsilon \), we use

\[ \tilde{u} = u_0 \left[ 1 + \tan h \left( \frac{\sqrt{\varepsilon} u_1}{u_0} \right) \right] \]

\[ = u_0 \left\{ 1 + \tan h \left( -(T-t) \frac{1}{u_0} \left[ R_1 \frac{\partial^3 u_0}{\partial x_1^3} + R_2 \frac{\partial^3 u_0}{\partial x_2^3} + R_{12} \frac{\partial^3 u_0}{\partial x_1 \partial x_2} \right] + R_{21} \frac{\partial^3 u_0}{\partial x_1 \partial x_2^2} \right) \right\} \]

(10)
Before proving that \( \tilde{u} \) given in (10) is indeed a probability density function, we clarify a definition first.

**Definition 1.** Let \( g \) be a function of \( n \) variables \((x_1, x_2, \ldots, x_n) \in \mathbb{R}^n\). The function \( g \) is called an \( n \)-dimensional even function if
\[
g(-x_1, -x_2, \ldots, -x_n) = g(x_1, x_2, \ldots, x_n)
\]
for all \((x_1, x_2, \ldots, x_n) \in \mathbb{R}^n\), and an \( n \)-dimensional odd function if
\[
g(-x_1, -x_2, \ldots, -x_n) = -g(x_1, x_2, \ldots, x_n)
\]
for all \((x_1, x_2, \ldots, x_n) \in \mathbb{R}^n\).

With this definition, we can state the following proposition.

**Proposition 1.** Let \( g(x) \) be a probability density function on \( \mathbb{R}^n \) for \( n \geq 1 \) and \( \phi(x) \) be an odd function. If \( g \) is an even function, then the function \( f \) defined by
\[
f(x) = (1 + \tan h(\phi(x)))g(x)
\]
is also a probability density function on \( \mathbb{R}^n \).

**Proof.** We need to prove that \( f \) is globally non-negative and its integral over \( \mathbb{R}^n \) is equal to 1. Observe that \( \tan h(\cdot) \) is strictly between \(-1\) and \(1\), and this together with the non-negativity of \( g \) justifies that \( f \) is always non-negative. On the other hand, \( \tan h(\cdot) \) is a (one-dimensional) odd function, and hence \( \tan h(\phi(x)) \) is an \((n\)-dimensional\) odd function. Now by change of variables \( y = -x \), we have
\[
I \equiv \int_{\mathbb{R}^n} \tan h(\phi(x))g(x)dx = \int_{\mathbb{R}^n} \tan h(\phi(-y))g(-y)dy
\]
\[
= -\int_{\mathbb{R}^n} \tan h(\phi(y))g(y)dy = -I
\]
which implies that \( I = 0 \). Therefore,
\[
\int_{\mathbb{R}^n} f(x)dx = \int_{\mathbb{R}^n} g(x)dx + I = 1 + 0 = 1
\]
The proof is complete. \( \blacksquare \)

Now observe that \( u_0 \) is a probability density function with respect to the variables \((\xi_1, \xi_2)\), and is even in \((\xi_1-x_1, \xi_2-x_2)\). In addition, \( \sqrt{\eta u_1/u_0} \) is an odd function in \((\xi_1-x_1, \xi_2-x_2)\). By Proposition 1, we know that \( \tilde{u} \) given in (10) is indeed a probability density function.
As for the approximation accuracy $|\tilde{u} - u^\varepsilon|$, we first note that
\[
\tan h(x) \approx x - \frac{x^3}{3}
\]
when $x$ is close to 0. Now for fixed $(t, x_1, x_2)$, when $\varepsilon$ is small, we have
\[
\tilde{u} = u_0 \left[ 1 + \tan h \left( \frac{\sqrt{\varepsilon} u_1}{u_0} \right) \right]
\approx u_0 \left[ 1 + \frac{\sqrt{\varepsilon} u_1}{u_0} - \frac{1}{3} \left( \frac{\sqrt{\varepsilon} u_1}{u_0} \right)^3 \right]
= u_0 + \sqrt{\varepsilon} u_1 - \varepsilon^{3/2} \left( \frac{u_1^3}{3u_0^3} \right) = \tilde{u} - \varepsilon^{3/2} \left( \frac{u_1^3}{3u_0^3} \right)
\]

Therefore, $|\tilde{u} - \tilde{u}|$ is small of order $\varepsilon^{3/2}$, while $|\tilde{u} - u^\varepsilon|$ is small of order $\varepsilon$ (see Fouque, Papanicolaou, Sircar, & Solna, 2003). Thus, $|\tilde{u} - u^\varepsilon|$ is small of the same order of $\varepsilon$ as $|\tilde{u} - \tilde{u}|$, that is, the approximation accuracy remains unchanged when replacing $\tilde{u}$ by $\tilde{u}$.

### 1.6. Marginal Transition Densities

For the marginal transition density function
\[
v^\varepsilon_i \equiv \mathbb{P}\{X^{(1)}_T \in d\xi_1 | X_t = x, Y_t = y\}
\]
the above argument goes analogously, and we obtain
\[
v^\varepsilon_i \approx \tilde{v}_1 \equiv p_1(t, x_1; T, \xi_1 | \tilde{\sigma}_1) - (T - t)R_1 \frac{\partial^3}{\partial x_1^3} p_1(t, x_1; T, \xi_1 | \tilde{\sigma}_1)
\]
where $p_1(t, x_1; T, \xi_1 | \tilde{\sigma}_1)$ is the transition density of the scaled Brownian motion with scale factor $\tilde{\sigma}_1$, that is,
\[
p_1(t, x_1; T, \xi_1 | \tilde{\sigma}_1) = \frac{1}{\sqrt{2\pi(T - t)\tilde{\sigma}_1}} \exp \left\{ - \frac{(\xi_1 - x_1)^2}{2\tilde{\sigma}_1(T - t)} \right\}
\]

A straightforward calculation shows that
\[
\frac{\partial^3 p_1}{\partial x_1^3} = \left[ - \frac{3(\xi_1 - x_1)}{\sqrt{2\pi\tilde{\sigma}_1^3(T - t)^{3/2}}} + \frac{(\xi_1 - x_1)^3}{\sqrt{2\pi\tilde{\sigma}_1^3(T - t)^{7/2}}} \right] \exp \left\{ - \frac{(\xi_1 - x_1)^2}{2\tilde{\sigma}_1^2(T - t)} \right\}
\]
Note again that
\[ \int_{-\infty}^{\infty} \tilde{v}_1(t, x_1; T, \xi_1) d\xi_1 = 1 \]

To guarantee the non-negativity of the approximated density function, we, again, use instead
\[ \tilde{v}_1 \equiv p_1 \left[ 1 + \tan h \left( -(T - t)R_1 \frac{1}{p_1} \frac{\partial^3 p_1}{\partial x_1^3} \right) \right] \]
as our approximation to \( v^\epsilon \).

By symmetry, we have
\[ v_2^\epsilon \equiv \mathbb{P} \left\{ X_T^{(2)} \in d\xi_2 | X_t = x, Y_t = y \right\} \]
\[ \approx \tilde{v}_2 \equiv p_2(t, x_2; T, \xi_2, \bar{\sigma}_2) - (T - t)R_2 \frac{\partial^3}{\partial x_2^3} p_2(t, x_2; T, \xi_2, \bar{\sigma}_2) \]
\[ \approx \tilde{v}_2 \equiv p_2 \left[ 1 + \tan h \left( -(T - t)R_2 \frac{1}{p_2} \frac{\partial^3 p_2}{\partial x_2^3} \right) \right] \]

where
\[ p_2(t, x_2; T, \xi_2, \bar{\sigma}_2) = \frac{1}{\sqrt{2\pi(T - t)|\bar{\sigma}_2^2}} \exp \left\{ -\frac{(\xi_2 - x_2)^2}{2\bar{\sigma}_2^2(T - t)} \right\} \]
\[ \frac{\partial^3 p_2}{\partial x_2^3} = \left[ - \frac{3(\xi_2 - x_2)}{\sqrt{2\pi\bar{\sigma}_2^2(T - t)^{3/2}}} + \frac{(\xi_2 - x_2)^3}{\sqrt{2\pi\bar{\sigma}_2^2(T - t)^{3/2}}} \right] \exp \left\{ -\frac{(\xi_2 - x_2)^2}{2\bar{\sigma}_2^2(T - t)} \right\} \]

and \( \tilde{v}_2 \) is our approximation to \( v_2^\epsilon \).

By exactly the same argument used for \( \tilde{u} \), one can show that \( \tilde{v}_1 \) and \( \tilde{v}_2 \) are indeed probability density functions of \( \xi_1 \) and \( \xi_2 \), respectively. Furthermore, the approximation accuracies remain unchanged when switching from \( \tilde{v}_1 \) to \( \tilde{v}_2 \), and from \( \tilde{v}_2 \) to \( \tilde{v}_2 \).

**2. Density of the Perturbed Copula**

**2.1. Approximated Copula Density**

Now suppose that conditional on \( \{X_t = x, Y_t = y\}, (X_T^{(1)}, X_T^{(2)}) \) admits the copula \( \Psi(\cdot, \cdot) \), then, by Sklar’s Theorem, its density function \( \psi(\cdot, \cdot) \) can be
represented as

\[ \psi(z_1, z_2) = \frac{u(t, x_1, x_2, y; T, \xi_1, \xi_2)}{v_1(t, x_1, y; T, \xi_1) v_2(t, x_2, y; T, \xi_2)} \]

where

\[ z_1 = \mathbb{P}(X^{(1)}_T \leq \xi_1 | X_t = x, Y_t = y) \]
\[ z_2 = \mathbb{P}(X^{(2)}_T \leq \xi_2 | X_t = x, Y_t = y) \]

Observe that if the volatility terms \( f_1(\cdot), f_2(\cdot) \) for \( (X^{(1)}_T, X^{(2)}_T) \) were constant numbers, say, the process \( \{Y_t\}_{t \leq T} \) was constant or the \( f_i \)'s were both identically constant, then \( \Psi \) would be a Gaussian copula.

Using our approximations to \( u/\sqrt{c_1}, v/\sqrt{c_1} \), and \( v/\sqrt{c_2} \), we have

\[ \psi(\xi_1, \xi_2) \approx \tilde{\psi}(\xi_1, \xi_2) = \frac{\tilde{u}(t, x_1, x_2; T, \xi_1, \xi_2)}{\tilde{v}_1(t, x_1; T, \xi_1) \tilde{v}_2(t, x_2; T, \xi_2)} \] \quad (11)

where

\[ \xi_1 = \int_{-\infty}^{\xi_1} \tilde{v}_1(t, x_1, T, \xi_1) d\xi_1 \]
\[ = \int_{-\infty}^{\xi_1} p_1(t, x_1; T, \xi_1) \]
\[ \times \left[ 1 + \tan h \left( -(T - t) R_1 \frac{1}{p_1(t, x_1; T, \xi_1)} \frac{\partial^3 p_1(t, x_1; T, \xi_1)}{\partial x_1^3} \right) \right] d\xi_1 \]
\[ \xi_2 = \int_{-\infty}^{\xi_2} \tilde{v}_2(t, x_2; T, \xi_2) d\xi_2 \]
\[ = \int_{-\infty}^{\xi_2} p_2(t, x_2; T, \xi_2) \]
\[ \times \left[ 1 + \tan h \left( -(T - t) R_2 \frac{1}{p_2(t, x_2; T, \xi_2)} \frac{\partial^3 p_2(t, x_2; T, \xi_2)}{\partial x_2^3} \right) \right] d\xi_2 \]

The function \( \tilde{u} \) is given by (10), and the marginals \( (p_1, p_2) \) and their derivatives \( \partial^3 p_1/\partial x_1^3, \partial^3 p_2/\partial x_2^3 \) are given explicitly in Section 1.6.

Before justifying that \( \tilde{\psi} \) is a probability density function defined on the unit square \([0, 1]^2\), we need the following proposition.

**Proposition 2.** Suppose function \( \Theta(x_1, x_2, \ldots, x_n) \) is an \( n \)-dimensional probability density function on \( \mathbb{R}_n \) for \( n \geq 2 \), and \( h_1(x_1), h_2(x_2), \ldots, h_n(x_n) \)
are one-dimensional strictly positive probability density functions. Then the function $c$ defined on the unit hyper-cube $[0, 1]^n$ by

$$c(z_1, z_2, \ldots, z_n) = \frac{\Theta(x_1, x_2, \ldots, x_n)}{\prod_{i=1}^{n} h_i(x_i)}$$

with $z_i \in [0, 1]$ given by

$$z_i = \int_{-\infty}^{x_i} h_i(y_i) \, dy_i$$

is a probability density function on $[0, 1]^n$. Furthermore, $c$ is a copula density function if and only if $h_1(x_1), h_2(x_2), \ldots, h_n(x_n)$ are the marginal density functions of $\Theta(x_1, x_2, \ldots, x_n)$, meaning that

$$h_i(x_i) = \int_{\mathbb{R}^{n-1}} \Theta(x_1, x_2, \ldots, x_n) \, dx_1 \, dx_2 \cdots dx_{i-1} \, dx_{i+1} \cdots dx_n$$

for every $i = 1, 2, \ldots, n$.

**Proof.** Let $H_i$ be the cumulative distribution function of $h_i$. Then $H_i$ is strictly increasing, implying the existence of its inverse function, and

$$z_i = H_i(x_i), \text{ or equivalently, } x_i = H_i^{-1}(z_i)$$

for each $i$. Since $\Theta$ is non-negative, and $h_i$'s are strictly positive, the function $c$ is non-negative. On the other hand,

$$\int_{[0,1]^n} c(z_1, z_2, \ldots, z_n) \, dz_1 \, dz_2 \cdots dz_n$$

$$= \int_{\mathbb{R}^n} c(H_1(x_1), H_2(x_2), \ldots, H_n(x_n)) \prod_{i=1}^{n} h_i(x_i) \, dx_1 \, dx_2 \cdots dx_n$$

$$= \int_{\mathbb{R}^n} \Theta(x_1, x_2, \ldots, x_n) \, dx_1 \, dx_2 \cdots dx_n = 1$$

Therefore $c(z_1, z_2, \ldots, z_n)$ is a probability density function on $[0, 1]^n$. 

*Perturbed Gaussian Copula*
Now if the additional condition is satisfied, then we have
\[
\int_{[0,1]^{n-1}} c(z_1, z_2, \ldots, z_n) dz_2 \ldots dz_n \\
= \int_{\mathbb{R}^{n-1}} c(z_1, H_2(x_2), \ldots, H_n(x_n)) \prod_{i=2}^{n} h_i(x_i) dx_2 \ldots dx_n \\
= \frac{1}{h_1(x_1)} \int_{[0,1]^{n-1}} \Theta(x_1, x_2, \ldots, x_n) dx_2 \ldots dx_n = 1
\]

This is to say that the marginal density function for the variable \(z_1\) is 1, and hence the marginal distribution for the variable \(z_1\) is uniform. Similarly, we can show that the marginal distributions for the variables \(z_2, \ldots, z_n\) are also uniform. By definition of copula, we know that function \(c\) is a copula density function. The converse can be obtained by reversing the above argument. The proof is complete. ■

Now from definition of \(\tilde{\Psi}\) given in Eq. (11), by combining the fact that \(\tilde{u}, \tilde{v}_1,\) and \(\tilde{v}_2\) are all probability density functions, one can see that \(\tilde{\Psi}\) is a density function on \([0, 1]^2\) by applying Proposition 2. However, \(\tilde{\Psi}\) is not a copula density function in general, because the additional condition required in Proposition 2 is not satisfied in general in our case, and hence \(\tilde{\Psi}\), the “copula” corresponding to density function, \(\tilde{\Psi}\), is not an exact copula in general.

Asymptotically, when \(\varepsilon\) tends to 0, for fixed \((t, x_1, x_2)\), the density \(\tilde{\Psi}\) converges to
\[
\phi(z_1, z_2) \equiv \frac{u_0(t, x_1, x_2; T, \xi_1, \xi_2)}{p_1(t, x_1; T, \xi_1)p_2(t, x_2; T, \xi_2)}
\]
with
\[
\xi_i = \int_{-\infty}^{z_i} p_i(t, x_i; T, \xi_i) d\xi_i = \Phi\left( \frac{\xi_i - x_i}{\sigma_i \sqrt{T - t}} \right)
\]
for \(i = 1, 2\), where \(\Phi(\cdot)\) denotes the univariate standard normal cumulative distribution function. One should observe that \(\phi(\cdot, \cdot)\) is the two-dimensional Gaussian copula density function with correlation parameter \(\hat{\sigma}\), and that it depends only on the parameter \(\hat{\sigma}\), independent of any other variables/parameters, including \(x_1, x_2, t, \sigma_1, \sigma_2\), etc.

As a consequence, \(\tilde{\Psi}\) converges to the Gaussian copula with correlation parameter \(\hat{\sigma}\). Since the method used in this paper is a perturbation method, we call \(\tilde{\Psi}\) a perturbed Gaussian copula.
2.2. Numerical Results

In this section, we illustrate the effectiveness of our approximation method by showing some numerical results.

In Fig. 1, we plot \( p_1(t, x_1; T, \xi_1) \), \( \tilde{v}_1(t, x_1; T, \xi_1) \), and \( \tilde{v}_1(t, x_1; T, \xi_1) \) as functions of \( \xi_1 \). Note that \( p_1(t, x_1; T, \xi_1) \) is a standard Gaussian density without any perturbation. The upper graph demonstrates the difference between \( p_1(t, x_1; T, \xi_1) \) (standard Gaussian) and \( \tilde{v}_1(t, x_1; T, \xi_1) \) (perturbed Gaussian), and the lower one between \( \tilde{v}_1(t, x_1; T, \xi_1) \) (simply perturbed Gaussian) and \( \tilde{v}_1(t, x_1; T, \xi_1) \) (perturbed Gaussian).

It can be seen from Fig. 1 that

- \( \tilde{v}_1(t, x_1; T, \xi_1) \) (simply perturbed Gaussian) takes on negative values at some places;
• \( \tilde{v}_1(t, x_1; T, \xi_1) \) (perturbed Gaussian), however, does not take on negative values, which is guaranteed by its formation;
• \( \tilde{v}_1(t, x_1; T, \xi_1) \) and \( \tilde{v}_1(t, x_1; T, \xi_1) \) are almost globally identical, which justifies the modification of the form \( 1 + \tan h(\cdot) \);
• \( \tilde{v}_1(t, x_1; T, \xi_1) \) is considerably different from \( p_1(t, x_1; T, \xi_1) \) (standard Gaussian); specifically, it shifts to the right from \( p_1(t, x_1; T, \xi_1) \);
• Despite the difference between \( \tilde{v}_1(t, x_1; T, \xi_1) \) and \( p_1(t, x_1; T, \xi_1) \), the areas under them do seem to be of the same size, which is justified by the fact that both are probability density functions and hence the overall integrals should both be one.

In Fig. 2, we plot \( \tilde{\psi}(\cdot, \cdot) \) in the lower graph and \( \phi(\cdot, \cdot, \cdot) \) in the upper graph, the Gaussian copula density that \( \tilde{\psi} \) converges to when \( \varepsilon \) tends to 0. It can be seen from Fig. 2 that the standard Gaussian copula density (upper graph) and the perturbed Gaussian copula density (lower graph) both give singularities at \((0, 0)\) and \((1, 1)\) but the perturbed one has more tail dependence at \((0, 0)\). Our numerous numerical experiments show that this picture is extremely sensitive to the choice of parameters and gives a lot of flexibility to the shape of the perturbed Gaussian copula density (the Matlab code is available on demand).

Tail dependence is a very important property for a copula, especially when this copula is to be used in modeling default correlation. The essence of tail dependence is the interdependence when extreme events occur, say, defaults of corporate bonds. The lack of tail dependence has for years been a major criticism on standard Gaussian copula.

Throughout the computation, we used the following parameters:

\[
R_1 = 0.02, \quad R_2 = 0.02, \quad R_{12} = 0.03 \\
R_{21} = 0.03, \quad \rho = 0.5, \quad T - t = 1 \\
\bar{\sigma}_1 = 0.5, \quad \bar{\sigma}_2 = 0.5, \quad x_1 = 0, \ x_2 = 0
\]

3. CONCLUSION

In summary, based on a stochastic volatility model, we derived an approximate copula function by way of singular perturbation that was introduced by Fouque et al. (2000). In the derivation, however, in order to make the candidate probability density functions globally non-negative, instead of directly using the obtained perturbation result as in
Fig. 2. Gaussian Copula and Perturbed Gaussian Copula Densities.
Fouque et al. (2000), we introduced a multiplicative modification, namely the \( 1 + \tan h(\cdot) \) form. It turns out that this modification is both necessary (to restore positiveness) and sufficient to guarantee the resulting functions to be density functions. Finally the resulting approximate copula – the so-called *perturbed Gaussian copula* in this paper – has a very desirable property compared to standard Gaussian copula: tail dependence at point \((0, 0)\). Some numerical results were provided and they strongly supported the methods described above, both the singular perturbation and the modification.

**REFERENCES**


**APPENDIX. EXPLICIT FORMULAS**

As stated in Section 1.3,

\[
\begin{align*}
    u_0(t, x_1, x_2) &= \frac{1}{2\pi \tilde{\sigma}_1 \tilde{\sigma}_2 (T - t) \sqrt{1 - \tilde{\rho}^2}} \times \\
    &\quad \exp \left\{ -\frac{1}{2(1 - \tilde{\rho}^2)} \left[ \frac{(\xi_1 - x_1)^2}{\tilde{\sigma}_1^2 (T - t)} - 2\tilde{\rho} \frac{(\xi_1 - x_1)(\xi_2 - x_2)}{\tilde{\sigma}_1 \tilde{\sigma}_2 (T - t)} + \frac{(\xi_2 - x_2)^2}{\tilde{\sigma}_2^2 (T - t)} \right] \right\}
\end{align*}
\]
By a straightforward calculation, we obtain

\[
\frac{\partial^3 u_0}{\partial x_1^3} = \exp \left\{ -\frac{1}{2(1 - \bar{\rho}^2)} \left[ \frac{(\xi_1 - x_1)^2}{\sigma_1^2(T - t)} - 2\bar{\rho} \frac{(\xi_1 - x_1)(\xi_2 - x_2)}{\sigma_1 \sigma_2(T - t)} + \frac{(\xi_2 - x_2)^2}{\sigma_2^2(T - t)} \right] \right\} 
\times \left\{ -\frac{2(\xi_1 - x_1)}{\sigma_1^2} + \frac{2\bar{\rho}(\xi_2 - x_2)}{\sigma_1 \sigma_2} \right\} \frac{3}{4\pi \sigma_1^3 \sigma_2(T - t)^3(1 - \bar{\rho}^2)^{5/2}} 
- \left[ -\frac{2(\xi_1 - x_1)}{\sigma_1^2} + \frac{2\bar{\rho}(\xi_2 - x_2)}{\sigma_1 \sigma_2} \right]^3 \frac{1}{16\pi \sigma_1 \sigma_2(T - t)^4(1 - \bar{\rho}^2)^{7/2}} 
\frac{\partial^3 u_0}{\partial x_1^2 \partial x_2} = \exp \left\{ -\frac{1}{2(1 - \bar{\rho}^2)} \left[ \frac{(\xi_1 - x_1)^2}{\sigma_1^2(T - t)} - 2\bar{\rho} \frac{(\xi_1 - x_1)(\xi_2 - x_2)}{\sigma_1 \sigma_2(T - t)} + \frac{(\xi_2 - x_2)^2}{\sigma_2^2(T - t)} \right] \right\} 
\times \left\{ -\frac{2\bar{\rho}(\xi_1 - x_1)}{\sigma_1 \sigma_2} - \frac{2(\xi_2 - x_2)}{\sigma_2^2} \right\} \frac{1}{4\pi \sigma_1^3 \sigma_2(T - t)^3(1 - \bar{\rho}^2)^{5/2}} 
- \left[ -\frac{2(\xi_1 - x_1)}{\sigma_1^2} + \frac{2\bar{\rho}(\xi_2 - x_2)}{\sigma_1 \sigma_2} \right]\frac{\bar{\rho}}{2\pi \sigma_1^2 \sigma_2^2(T - t)^3(1 - \bar{\rho}^2)^{5/2}} 
- \left[ -\frac{2(\xi_1 - x_1)}{\sigma_1^2} + \frac{2\bar{\rho}(\xi_2 - x_2)}{\sigma_1 \sigma_2} \right]^2 \frac{2\bar{\rho}(\xi_1 - x_1)}{\sigma_1 \sigma_2} - \frac{2(\xi_2 - x_2)}{\sigma_2^2} \right]\frac{1}{16\pi \sigma_1 \sigma_2(T - t)^4(1 - \bar{\rho}^2)^{7/2}} 
\right\}
\]

The partial derivatives \( \frac{\partial^3 u_0}{\partial x_1^3} \) and \( \frac{\partial^3 u_0}{\partial x_1^2 \partial x_2} \) are obtained by symmetry.
THE DETERMINANTS OF DEFAULT CORRELATIONS

Kanak Patel and Ricardo Pereira

ABSTRACT

This chapter analyses the ability of some structural models to predict corporate bankruptcy. The study extends the existing empirical work on default risk in two ways. First, it estimates the expected default probabilities (EDPs) for a sample of bankrupt companies in the USA as a function of volatility, debt ratio, and other company variables. Second, it computes default correlations using a copula function and extracts common or latent factors that drive companies’ default correlations using a factor-analytical technique. Idiosyncratic risk is observed to change significantly prior to bankruptcy and its impact on EDPs is found to be more important than that of total volatility. Information-related tests corroborate the results of prediction-orientated tests reported by other studies in the literature; however, only a weak explanatory power is found in the widely used market-to-book assets and book-to-market equity ratio. The results indicate that common factors, which capture the overall state of the economy, explain default correlations quite well.
1. INTRODUCTION

Corporate defaults exhibit two key characteristics that have profound implications for default risk management. First, default risk is correlated through time. Bankruptcies are normally the end of a process that begins with adverse economic shocks and end with financial distress. Although some bankruptcies are unexpected and, therefore, are point events, like Enron and Worldcom, investors become aware of the company’s difficulties some years prior to the bankruptcy event. Second, financial wealth of companies in the same industry, or within the same economic area, is a function of managers’ skills and common factors that introduce correlations.

Companies’ default risk is linked through sector-specific and/or macroeconomic factors. While a great deal of effort has been made by practitioners to measure and explain companies’ default correlations, academics have only recently begun to devote attention to this issue. The existing literature on default correlations can be divided into two approaches: the structural approach that models default correlations through companies’ assets values; and the reduced-form approach that models default correlations through default intensities. While financial institutions, namely banks, are aware of these relationships, their ability to model such correlations is still not fully developed. The Basel Committee on Banking and Supervision (BCBS, 1999, p. 31) states “... the factors affecting the credit worthiness of obligors sometimes behave in a related manner...” which “... requires consideration of the dependencies between the factors determining credit related losses.” While there are many different models and approaches to compute default probabilities, there is no consensus on the importance of different factors that drive default correlations. The BCBS (1999) report points out that while practitioners have been managing and studying this dependence, there is a lack of theoretical and empirical work on this issue that tests the robustness of the frameworks.

In this chapter, we concentrate our empirical investigation on the determinants of default correlation. Our analysis comprises three stages: first, we apply a set of structural models, Merton (M, 1974), Longstaff and Schwartz (LS, 1995), and Ericsson and Reneby (ER, 1998), to compute companies’ expected default probabilities (EDPs). Second, based on cross-sectional tests we analyse the effect of volatility and idiosyncratic risk on EDPs. Given that unexpected events or fraudulent defaults lead to market-wide jumps in credit spreads, which reduce the ability to diversify this risk, it
is important to examine the relationship between company’s idiosyncratic risk and bankruptcy. Third, using a factor-analytical technique, we extract common or latent factors that explain default correlations. This analysis enables us to assess the extent to which default correlation can be ascribed to the latent factors and to the systematic variables from capital and bond markets.

The results show that the set of structural models applied are able to predict bankruptcy events. Another important finding is the relevance of idiosyncratic risk (and not of total volatility) in predicting default events. This suggests that company-specific signals provide useful information to investors about the deterioration in company’s economic and financial conditions prior to bankruptcy. Factor-analytical techniques extract factors that explain around 83% of the variability of default correlations. The determinants of these factors are variables that proxy the overall state of the economy and the expectations of its evolution.

The most popular credit risk frameworks used and sold by financial institutions are the KMV (building on the Merton, 1974, model) and CreditMetrics. In the Merton model, dependence between companies’ defaults is driven by dependence between assets and threshold values. In the actuarial CreditRisk+ framework, default correlations are driven by common factors. For each pair of obligors, the asset value is assumed to follow a joint normal distribution. The efficacy of diversification within a portfolio of claims requires accurate estimates of correlations in credit events for all pairs of obligors. For example, Collateralized Bond Obligations (CBOs) and the evaluation of credit derivatives examined by Hull, Predescu, and White (2005) require estimates of the joint probability of default over different time periods and for all obligors. Default correlations can lead to a dramatic change in the tails of a portfolio’s probability density function of credit losses (PDCLs) and, consequently, in the economic capital required to cover unexpected losses. The common assumption of independence between events produces the right tails of the theoretical PDCLs to be thinner than the ones observed in practice, which implies that observed unexpected losses are higher than the ones estimated. BCBS (1999) points out that PDCLs of portfolios are skewed toward large losses and are more difficult to model. The PDCLs that result from the combination of single credit exposures depends on the assumptions made about credit correlations.

The rest of the chapter is organized as follows: Section 2 presents a brief digression on dependence measures, with an exposition of copula functions. Section 3 provides a discussion on empirical analyses of structural models.
and the variables that can account for default correlations and contagion effects. Section 4 contains our empirical work. Section 5 discusses the implications of the results and Section 6 provides some concluding remarks.

2. A BRIEF DIGRESSION ON MEASURES OF DEPENDENCE

The Pearson correlation coefficient, $\rho$, commonly used in finance as a measure of dependence between two variables, assumes that financial variables follow a multivariate normal distribution, which means that it can only be used in the elliptical world (see Embrechitz, McNeil, & Straumann (2001) for the limitations of this measure). However, the probability distribution of security returns is not normal; it has fat tails and skewness. This characteristic is crucial for credit risk management, which requires careful consideration of other dimensions of risk. One of these dimensions is the dependence structure between the variables. The copula function allows us to measure this dimension.

In this section, we briefly describe the basic concepts of copula functions. A copula function defines the dependence structure between random variables. It links univariate marginals to their multivariate distribution. Consider $p$ uniform random variables, $u_1, u_2, \ldots, u_p$. The joint distribution function of these variables is defined as

$$C(u_1, u_2, \ldots, u_p) = \text{Prob}\{U_1 \leq u_1, U_2 \leq u_2, \ldots, U_p \leq u_p\}$$ (1)

where $C$ is the copula function. Copula functions are used to relate univariate marginal distributions functions, $F_1(x_1), F_2(x_2), \ldots, F_p(x_p)$, to their joint distribution function

$$C(F_1(x_1), F_2(x_2), \ldots, F_p(x_p)) = F(x_1, x_2, \ldots, x_p)$$ (2)

For the random variable, the univariate marginal distribution can be chosen according to its features. The copula function does not constrain the choice of the marginal distribution. Sklar (1959) (cited in Frees & Valdez, 1998) proves that any multivariate distribution function, $F$, can be written in the form of Eq. (2). He also shows that if each marginal distribution function is continuous, then there is a unique copula representation.

Copula functions have been used in biological science to analyse the joint mortality pattern of groups of individuals. Li (2000) applied this concept to
default correlation between companies. Schonbucher and Schubert (2001) use a different approach, the frailty model, to study default correlations within an intensity model, which is used in biological studies to model heterogeneity via random effects.

The copula summarizes different types of dependencies even when they have been scaled by strictly monotone transformations (invariance property). The properties of bivariate copula functions, $C(u, v, \rho)$, where $u$ and $v \in (0, 1)^2$ and $\rho$ is a correlation parameter (Pearson correlation coefficient, Spearman’s Rho, Kendall’s Tau) are as follows:

(i) since $u$ and $v$ are positive numbers, $C(0, v, \rho) = C(u, 0, \rho) = 0$
(ii) the marginal distribution can be obtained by $C(1, v, \rho) = v$ or $C(u, 1, \rho) = u$
(iii) if $u$ and $v$ are independent variables, $C(u, v, \rho) = uv$
(iv) the upper and lower bound for a copula function is $\max(0, u + v - 1) \leq C(u, v) \leq \min(u, v)$

The generalization of these properties to higher dimensions is straightforward. The joint distribution function is defined by its marginals and the copula. This means that we can examine the copula function to capture the association between random variables. Both Spearman’s Rho, $\rho_S$, and Kendall’s Tau, $\tau$, can be defined in terms of the copula function as follows:

$$\rho_S = 12 \int [C(u, v) - uv] du \ dv$$

$$\tau = 4 \int C(u, v) dC(u, v) - 1$$

The nonparametric correlation measures do not depend on the marginal distributions and are not affected by nonlinear transformations like the Pearson correlation coefficient.

Mendes and Souza (2004) demonstrate that copula density functions split the joint distribution function into parameters of marginals, $\gamma$, and parameters of dependence structure, $\delta$. To fit a copula to bivariate data we maximize the log-likelihood function, $l$

$$l = (u, v, \gamma_u, \gamma_v, \delta) = \log [c(F_u(u, \gamma_u), F_v(v, \gamma_v); \delta)] + \log f_u(u; \gamma_u) + \log f_v(v; \gamma_v)$$

where $c$ is the copula density function and $f$ is the marginal density function.
Durrleman, Nikeghbali, and Roncalli (2000) present different methods for choosing the right copula. In this study, we rely on the standard measures, Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC).

3. DEFAULT RISK AND CORRELATIONS

Over the past few decades, dependence within financial markets has been extensively studied both in portfolio diversification and financial integration. Only recently, however, important events like the Asian crisis have spurred academic interest in explaining the causes and consequences of the bubble bursting in new economies and fraudulent bankruptcies of Enron and WorldCom. More recently, researchers have started to examine the dependence structure of extreme events across companies.

The structural models attempt to capture the salient features of the real economy that cause corporate defaults. These models can be divided into two sets: Merton (1974) model that considers default only at the maturity of the zero coupon bonds; and those models that allow default to occur at any time within the prediction horizon – the first-passage models. We must note that as these models work with risk-neutral measures, and their EDP can differ from the real one and is likely to be higher. This is because the drift of the real process of the asset value is normally higher than the risk-neutral process (see Appendix A). Thus, we expect to observe EDPs resulting from the first-passage models to be higher than the one resulting from the Merton-type of models.

The empirical analyses of structural models, including Jones, Mason, and Rosenfeld (1984), Huang and Huang (2003), and Eom, Helwege, and Huang (2004), report that these models tend to systematically underestimate observed yield spreads and, given the high dispersion of predicted spreads, are inaccurate. In our opinion, this does not affect the accuracy of a structural model in estimating EDPs of companies. The existing studies on observed yield spreads do not consider all the relevant components that affect yield spreads. As Fisher (1959) argues, an observed bond yield spread provides compensation to investor for credit risk and marketability risk. Several authors (see Delianedis & Geske, 2001, and Ericsson & Reneby, 2005) point out that default spread is only a small proportion of the observed yield spread. The studies by Leland (2002), Patel and Vlamis (2006), and Patel and Pereira (2007), report that EDPs from structural models are able to predict bankruptcies, in some cases up to 2 years before
the event. Evidently, the EDPs contain valuable information especially in cases when companies are close to economic/financial distress.

Recently, a number of authors have investigated the extremal dependence of risk factors to model default. Insofar as defaults are infectious, an analysis of default correlations is crucial. Li (2000) is one of the earliest studies to systematically examine default correlations. The author models default correlation between two companies as the correlation between their survival times. He uses the copula concept to define the joint distribution of survival times with given marginal distributions. Li (2000) points out that CreditMetrics uses a bivariate normal copula function with asset correlation as the correlation factor. Laurent and Gregory (2005) extend this work to several obligors. Frey and McNeil (2001) use a copula function and the notion of extremal dependence of risk factors to model default correlations in loan portfolio management. Davis and Lo (2001) study how “infectious defaults” (or contagion effects) can be introduced within the Binomial Expansion Technique developed by Moody’s. The authors investigate this issue assuming that default correlation among all firms of a CBO is equal and time independent.

Hull and White (2000) develop a method to value a vanilla credit default swap with counterparty default risk, which assumes that the dependence structure of defaults follows a multivariate normal distribution. Hull et al. (2005) extend the previous model to several obligors. They assume that default threshold has a systematic and an idiosyncratic component. The systematic component is defined as the sensitivity of the threshold to a factor (systematic), common to all firms. Default correlation is defined as the product of each company loading to the systematic variable. Zhou (2001) provides an analytical formula for computing default correlations and joint default probability for the first-passage models. However, the empirical application of this framework to portfolios of loans or bonds becomes cumbersome since it only allows pairwise comparison of obligors.

In another line of investigation, Schonbucher (2003) analyses default correlation spreads through channels other than business ties. Assuming an imperfect market, with asymmetric information, default contagion can arise from information effects, learning effects or updating of beliefs, which means that the default of one company provides information about the default risk of other companies. Collin-Dufresne, Goldstein, and Helwege (2003) study default contagion via updating of beliefs, within a reduced-form model. According to the authors, unexpected or fraudulent defaults lead to market-wide jumps in credit spreads, which reduces
investors’ ability to diversify this risk. Giesecke (2004) argues that macroeconomic variables and operational and/or financial ties can explain default correlations between companies. More specifically, default correlations between companies are due to their dependence on macroeconomic variables, which cause cyclical default correlations, and operational and financial relationships with other companies that cause default contagion effects.

Makiel and Xu (2000) find that investors price idiosyncratic risk because they cannot hold a diversifiable portfolio. Similar evidence is presented by Goyal and Santa-Clara (2003). Arguably, if investors’ ability to diversify risk is limited, idiosyncratic risk is likely to be an important determinant of default correlation in the period leading up to company’s financial distress.

4. DATA AND METHODOLOGY

The stock price and financial data on a sample of bankrupt companies in the USA used in this study is obtained from the Datastream and Osiris database. The names of bankrupt companies are collected from Moody’s Investor Service Reports (2003, 2005). A company is classified as bankrupt if it missed or delayed disbursement of interest or principal or if it entered into liquidation, receivership, or administration. Our initial sample comprised 59 bankruptcy events between 1996 and 2004, a total of 56 bankrupt companies. In order to ensure reliability of the results, we excluded thinly traded companies (when there is more than 10 days without any trade) and companies with less than 5 years of financial data. The remaining sample comprises 34 bankruptcy events and 282 yearly observations on related economic and financial variables, a total sample of 34 bankrupt companies. For the risk-free rate, we use the yield on 1-year treasury constant maturity (TCM) securities from 1990 to 2004 reported by the US government securities dealers to the Federal Reserve Bank of New York.

Our empirical methodology comprises three stages: Stage 1 involves estimation of EDPs of companies, using three structural models: Merton (1974), Longstaff and Schwartz (1995), and Ericsson and Reneby (1998). Prediction-oriented and information-related tests are employed to infer the performance of these models. Stage 2 involves an estimation of the idiosyncratic risk of companies. Stage 3 involves factor analyses of the default correlation matrix and latent factors associated with the companies.
Appendix A presents an outline of the Merton (M, 1974), Longstaff and Schwartz (LS, 1995), and Ericsson and Reneby (ER, 1998) models. Each of these models has a set of parameters that we either estimate or assume to be given. Table 1 describes the parameters and how they are computed in our analysis. Our calibration approach is not very different from the standard one employed in previous studies except that the focus here is solely on the parameters needed to compute the EDPs.

Ideally, to apply these structural models, we should have companies with simple capital structures with only the equity and zero coupon bonds. One practical approach is to assume that company’s debt can be converted to a 1-year zero coupon bond with a face value equal to its debt value. The total market value of the company, and its volatility, can be computed using an iterative procedure based on Ito’s Lemma (a similar procedure is used by KMV). For the initial estimate of the company’s volatility, \( \sigma_v \), we compute the standard deviation of daily equity returns, \( \sigma_E \), over the past 12 months. Then, using Eq. (A.4), we compute iteratively the daily market value of the company, \( V_t \), corresponding to the market value of equity, \( E_t \), until the difference in values of \( \sigma_v \) from two consecutive iterations converge to less than 10E

### Table 1. Calibration Procedure of M, LS, and ER Models.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Model</th>
<th>Estimated As</th>
</tr>
</thead>
<tbody>
<tr>
<td>Firms’ specific parameters</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( V_t ): Company’s value</td>
<td>All</td>
<td>Ito’s lemma</td>
</tr>
<tr>
<td>( \sigma_V ): Company’s volatility</td>
<td>All</td>
<td>Ito’s lemma</td>
</tr>
<tr>
<td>( F ): Debt’s face value</td>
<td>M</td>
<td>Book value of total liabilities</td>
</tr>
<tr>
<td>( T ): Years to maturity</td>
<td>All</td>
<td>Assumed 1 year</td>
</tr>
<tr>
<td>( \delta ): Payout ratio</td>
<td>ER</td>
<td>Assumed at 6%</td>
</tr>
<tr>
<td>( \tau ): Prediction horizon</td>
<td>ER</td>
<td>Assumed 1 year</td>
</tr>
<tr>
<td>( \alpha ): Bankruptcy costs</td>
<td>LS</td>
<td>Assumed at 49%</td>
</tr>
<tr>
<td>( K ): Threshold value/distress barrier</td>
<td>LS; ER</td>
<td>Debt’s face value</td>
</tr>
<tr>
<td>Interest rate parameters</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( r ): Interest rate</td>
<td>All</td>
<td>1-year TCM</td>
</tr>
<tr>
<td>( \alpha ): Mean reversion speed</td>
<td>LS</td>
<td>Vasicek’s risk-free yield curve</td>
</tr>
<tr>
<td>( \lambda ): Mean reversion level</td>
<td>LS</td>
<td>Vasicek’s risk-free yield curve</td>
</tr>
<tr>
<td>( \sigma_r ): Short rate standard deviation</td>
<td>LS</td>
<td>Assumed at 1.5%</td>
</tr>
<tr>
<td>( \rho ): Correlation coefficient between ( r ) and ( V_t )</td>
<td>LS</td>
<td>Computed</td>
</tr>
</tbody>
</table>
σ_v is then used to compute the market value of the company, V_t. We consider T and τ to equal 1 year, assuming that investors’ prediction horizons are equal to 1 year. The parameter δ captures the payments made by the company to its shareholders and bondholders, such as dividends, share repurchases, and bond coupons. According to Huang and Huang (2003), 6% can be assumed to be a reasonable estimate for this parameter.6

Several studies in the literature report that, bondholders’ recovery rate varies according to the seniority of the debt. For example, Altman (2002) finds that during the period 1985–1991, the average recovery rate for a sample of defaulted bond issues was 0.605 for secured debt, 0.523 for senior debt, 0.307 for senior subordinated debt, 0.28 for cash-pay subordinated debt, and 0.195 for non-cash-pay subordinated debt. Given this evidence, Longstaff and Schwartz (1995), Leland (2002), Huang and Huang (2003), and Eom et al. (2004), assume an average recovery rate of 51% of debt face value.

In the one-factor models (M and ER), we use the yield on 1-year TCM rate as the risk-free rate. In the two-factor LS model, we assume the interest rate is driven by the Vasicek process described in Eq. (A.1). Based on the evidence reported by Eom et al. (2004), who apply Vasicek and Nelson-Siegel models to estimate the term structure of the risk-free yield curve, we fit the Vasicek model to 1-year TCM rates assuming that σ_r=0.015 (see Appendix A and Eq. (A.2) for details of the estimation procedure). We estimate the parameters a and λ using this procedure for each year, from 1990 to 2004, with the daily observations of 1-year TCM. The correlation coefficient is computed with 1-year TCM rates and V_t for each common year.

4.2. Estimation of Idiosyncratic Risk

A widely used procedure for estimating the idiosyncratic risk involves extracting the residuals of an asset-pricing model. Obviously, the estimates are sensitive to the chosen asset-pricing model and the specified variables. Since the existing literature has tended to employ the three-factor model7 by Fama and French (1993), who use the following model:

\[ R_{it} - R_{ft} = \beta_{m,t}(R_{m,t} - R_{ft}) + \beta_{smb,t}R_{smb,t} + \beta_{hml,t}R_{hml,t} + \epsilon_{i,t} \tag{6} \]

where \( R_{it} \) is the return on company i on day t. \( R_{mt} - R_{ft} \) is the market excess return. \( R_{smb,t} \) is the return on a portfolio that captures the size effect, which
is computed as the average return for the smallest 30% of stocks minus the
average return of the largest 30% of stocks. $R_{hml,t}$ is the return on a
portfolio that captures the book-to-market equity effect (value premium),
which is computed as the average return for the 50% of stocks with the
highest book-to-market ratio minus the average return of the 50% of stocks
with the lowest book-to-market ratio. The standard deviation of $e_i$ is used
as a proxy for the idiosyncratic risk of company $i$. We fit this model using
daily observations over the previous year.

4.3. Regression Analysis of Default Correlation

Based on Eq. (5), we fit copula functions to each pair of companies’ EDPs.
To fit the copula functions, we define a minimum of 5-year common period
for each company. This reduces our sample from 25 to 24 bankruptcy
events. The estimated copula functions for each pair of companies’ EDPs
(a total of 276 copula functions) are then transformed into Kendall’s Tau
using Eq. (4) and used to construct each model default correlation matrix.
Each model default correlation matrix is then used in the factor analysis
(see Appendix B) in order to extract the common or latent factors, Fact$_{s,t}$,
that are not directly observable but explain the companies default
correlations. These common factors will then serve as dependent variables
in regression Eq. (7) below.

Our next task is to identify variables that drive default correlations so that
we can use the variables in a regression equation as the extracted factors,
Fact$_{s,t}$. We selected the following set of variables for their theoretical
robustness and empirical measurability:

(1) **Treasury Interest Rates Level.** Several authors (e.g., Longstaff &
Schwartz, 1995; Leland & Toft, 1996) argue that an increase in the
spot rate increases the drift of a company’s asset value process and
causes EDPs to fall. Since the majority of the models consider the
default threshold to be constant or deterministic, an increase in the drift
pushes the company’s value away from its threshold value and decreases
default probability. Since an increase in the level of interest rates
decreases EDPs, we should also expect to observe a decrease in default
correlations. We use the yield on the 10-year TCM securities $r_{t}^{10}$ for the
interest. In line with Collin-Dufresne, Goldstein, and Martin (2001), we
use $(r_{t}^{10})^2$ to capture potential nonlinear effects due to convexity.
(2) **Slope of the Yield Curve.** The impact of this variable on default probabilities and default correlations is controversial. In our opinion, since this variable reflects investors’ expectations about the evolution of the economy, an increase in the slope of the yield curve implies strengthening of the economy, which, consequently, would lower EDPs and default correlations. We define this variable as the difference between the 10-year and 2-year TCM yields, \( r_{10}^t - r_2^t \).

(3) **Market Volatility.** Market volatility is a critical parameter in structural models. The effect of volatility depends on the model specification. In the first-passage model, an increase in volatility increases the probability of default and increases default correlations, because the probability of a company’s value crossing the threshold at any point in time also increases. In the European type model (Merton, 1974) the effect is not obvious; it can be positive or negative. We measure market volatility, \( \sigma_{S&P} \), as the standard deviation of S and P daily returns over the past 12 months.

(4) **Equity Premium.** Equity premium can be considered to be a proxy for the overall state of the capital markets. An increase in equity premium reflects an increase in risk and, therefore, is expected to result in higher defaults and default correlations. We measure equity premium, \( R_{M,t} - r_{1m}^t \), as the difference between the value-weighted return on all NYSE, AMEX, and NASDAQ and 1-month treasury bill rate.

(5) **Default Return Spread.** This variable captures the systematic risk factor as well as the specific risk when there are unexpected events of bankruptcy or fraud. As explained by Schonbucher (2003), this variable can be interpreted as a learning or information effect variable. An increase in default return spread increases an overall uncertainty in the bond market, which causes an increase in default correlations as investors become more sensitive to bad news. We define default return spread, \( DefSpread_t \), as the difference between Moody’s AAA and BAA long-term bonds yields. 

Table 2 summarizes the expected signs of the relationship between the default correlation factors and the variables outlined above. The first four variables capture cyclical default correlation, while the last one captures the systematic component of default contagion effects. We estimate the following regression equation, the results of which will be reported later (see Table 10)

\[
\text{Fact}_{s,t} = \alpha + \beta_1 r_{10}^t + \beta_2 (r_{10}^t)^2 + \beta_3 (r_{10}^t - r_2^t) \\
+ \beta_4 \sigma_{S&P,t} + \beta_5 (R_{M,t} - r_{1m}^t) + \beta_6 \text{DefSpread}_t
\] (7)
In this section, we report and discuss the results of the three structural models obtained from prediction-oriented and information-related tests. We also analyze the importance of idiosyncratic risk in predicting bankruptcy events. We do this by extracting latent factors of companies’ default correlations based on the competing structural models. Looking ahead, in Section 5.1 we look at the predictive power of the EDPs generated by the competing structural models. In Section 5.2, we provide the summary statistics of the EDPs of the competing models. In Section 5.3, we conduct several cross-section analyses of EDPs. In Section 5.4, we conduct a logistic analysis of the EDPs and several explanatory variables, while in Section 5.5 we apply factor analysis methods to the correlation matrices of the Kendall’s Tau based on the copula analysis of the EDPs of the competing structural models.

### 5.1. Prediction-Oriented Tests

Prediction-oriented tests provide an in-sample accuracy measure. We classify the results into error type I (predicting no default when there is actually one) and type II (predicting a default when there actually is none). Since our sample comprises only bankrupt firms, we can only observe Error type I when the model fails to predict bankruptcy. The models correctly predict bankruptcy if in the final available year the EDP is above 20%. The results show misclassification of three bankruptcy events in the M and
LS models, which corresponds to 8.3% of the bankruptcy events. The ER model has the best performance with only two misclassified bankruptcy events, which corresponds to 5.5% of the sample. Overall, the structural models predict corporate bankruptcy at least 1 year in advance of the event.

5.2. Summary Statistics of EDPs

Table 3 reports the statistics of EDPs for “All years” and for up to $n-6$ previous years (it is not feasible to present all the results over the period 1990–2004). The first important observation in Table 3, also depicted in Fig. 1, is that average EDP of the M model is lower than those of LS and ER models. As mentioned earlier, the former model is a European option with one period debt, and the latter models are a kind of Barrier option

<table>
<thead>
<tr>
<th></th>
<th>Total Sample</th>
<th>$n-1$</th>
<th>$n-2$</th>
<th>$n-3$</th>
<th>$n-4$</th>
<th>$n-5$</th>
<th>$n-6$</th>
<th>All Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>M Mean</td>
<td>0.57</td>
<td>0.35</td>
<td>0.24</td>
<td>0.19</td>
<td>0.09</td>
<td>0.09</td>
<td>0.17</td>
<td></td>
</tr>
<tr>
<td>M Standard deviation</td>
<td>0.27</td>
<td>0.27</td>
<td>0.26</td>
<td>0.24</td>
<td>0.16</td>
<td>0.17</td>
<td>0.25</td>
<td></td>
</tr>
<tr>
<td>LS Mean</td>
<td>1.23</td>
<td>0.80</td>
<td>0.55</td>
<td>0.39</td>
<td>0.25</td>
<td>0.22</td>
<td>0.38</td>
<td></td>
</tr>
<tr>
<td>LS Standard deviation</td>
<td>0.66</td>
<td>0.74</td>
<td>0.61</td>
<td>0.45</td>
<td>0.61</td>
<td>0.49</td>
<td>0.60</td>
<td></td>
</tr>
<tr>
<td>ER Mean</td>
<td>1.11</td>
<td>0.74</td>
<td>0.53</td>
<td>0.47</td>
<td>0.22</td>
<td>0.21</td>
<td>0.37</td>
<td></td>
</tr>
<tr>
<td>ER Standard deviation</td>
<td>0.43</td>
<td>0.48</td>
<td>0.50</td>
<td>0.58</td>
<td>0.37</td>
<td>0.37</td>
<td>0.50</td>
<td></td>
</tr>
</tbody>
</table>

The Table 3 presents the standard statistics of the M, LS, and ER models’ EDPs.

Fig. 1. EDPs of Bankrupt Companies.
models. Focusing on the behavior of EDPs, we observe a gradual increase in EDPs up to 2 years ahead of bankruptcy and then a steep rise a year before the event.

The second important observation is that the standard deviations of EDPs of the M model (approximately 25%) are comparatively lower than those of the LS and ER models (approximately 60% and 40%, respectively). This suggests that the M model is more accurate in predicting bankruptcy than the LS and ER models (see Figs. 2–4). For the first-passage LS model, we observe a distinct clustering, however, this model appears to be the least accurate with more extreme values (see Figs. 3 and 4).

Overall, the results suggest that the first-passage LS model does not add value over and above the M model. Surprisingly, the two-factor LS model

![Fig. 2. Distribution of EDPs: Merton’s Model.](image)
has the worst performance, suggesting that the effort to capture more realistic features of the company in this model is not justified as far as EDPs are concerned.

5.3. Cross-Section Analysis of EDPs

Tables 4–6 present the results of the multivariate linear regressions of EDPs volatility and debt ratio, the two crucial parameters of structural models. In Table 4, all the coefficients of these two variables are statistically significant and have the expected signs. A high percentage of the variability of EDPs is explained by volatility debt ratio (only in the LS model, this percentage is

---

*Fig. 3. Distribution of EDPs: Longstaff and Schwartz Model.*
Table 4. Cross-Section Analysis EDPs and Volatility.

<table>
<thead>
<tr>
<th>Model</th>
<th>(a)</th>
<th>(t)-stat</th>
<th>(b)</th>
<th>(t)-stat</th>
<th>(c)</th>
<th>(t)-stat</th>
<th>(F)</th>
<th>Adj (R^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>-0.344</td>
<td>-10.9</td>
<td>0.704</td>
<td>16.3</td>
<td>0.26</td>
<td>10.9</td>
<td>150.2^a</td>
<td>0.515</td>
</tr>
<tr>
<td>LS</td>
<td>-0.706</td>
<td>-8.3</td>
<td>1.601</td>
<td>13.8</td>
<td>0.344</td>
<td>5.4</td>
<td>95.5^a</td>
<td>0.402</td>
</tr>
<tr>
<td>ER</td>
<td>-0.632</td>
<td>-9.9</td>
<td>1.474</td>
<td>16.9</td>
<td>0.333</td>
<td>6.9</td>
<td>143^a</td>
<td>0.503</td>
</tr>
</tbody>
</table>

The Table 4 reports the results of a cross-sectional multiple linear regression relating models’ EDPs to the companies’ debt ratio and volatility.

^aConfidence level at 1%.
The EDPs are more sensitive to the debt ratio than to the volatility. The first-passage models are more sensitive to debt ratio than the M model. All regressions are statistically significant at the 1% level.

Table 5. Cross-Section Analysis of EDPs and Idiosyncratic Risk.

<table>
<thead>
<tr>
<th>Model</th>
<th>$a$</th>
<th>$t$-stat</th>
<th>$b$</th>
<th>$t$-stat</th>
<th>$c$</th>
<th>$t$-stat</th>
<th>$e_i$</th>
<th>$F$</th>
<th>Adj $R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>-0.368</td>
<td>-18.1</td>
<td>0.263</td>
<td>8.7</td>
<td>0.543</td>
<td>23.7</td>
<td>471.7</td>
<td>0.77</td>
<td></td>
</tr>
<tr>
<td>LS</td>
<td>-0.845</td>
<td>-13.1</td>
<td>0.871</td>
<td>9.1</td>
<td>1.001</td>
<td>13.8</td>
<td>218.8</td>
<td>0.608</td>
<td></td>
</tr>
<tr>
<td>ER</td>
<td>-0.716</td>
<td>-15.1</td>
<td>0.837</td>
<td>11.9</td>
<td>0.834</td>
<td>15.7</td>
<td>314.3</td>
<td>0.69</td>
<td></td>
</tr>
</tbody>
</table>

The Table 5 reports the results of a cross-sectional multiple linear regression relating models’ EDPs to the companies’ debt ratio and idiosyncratic risk. *Confidence level at 1%.

Table 6. Cross-Section Analysis of EDPs Volatility and Idiosyncratic Risk.

<table>
<thead>
<tr>
<th>Model</th>
<th>$a$</th>
<th>$t$-stat</th>
<th>$b$</th>
<th>$t$-stat</th>
<th>$c$</th>
<th>$t$-stat</th>
<th>$d$</th>
<th>$t$-stat</th>
<th>$F$</th>
<th>Adj $R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>-0.415</td>
<td>-19.9</td>
<td>0.361</td>
<td>10.9</td>
<td>0.103</td>
<td>5.9</td>
<td>0.476</td>
<td>19.6</td>
<td>365.1</td>
<td>0.795</td>
</tr>
<tr>
<td>LS</td>
<td>-0.854</td>
<td>-12.2</td>
<td>0.889</td>
<td>8</td>
<td>0.02</td>
<td>0.3</td>
<td>0.988</td>
<td>12.1</td>
<td>145.4</td>
<td>0.607</td>
</tr>
<tr>
<td>ER</td>
<td>-0.749</td>
<td>-14.7</td>
<td>0.908</td>
<td>11.2</td>
<td>0.074</td>
<td>1.7</td>
<td>0.533</td>
<td>13.2</td>
<td>212.1</td>
<td>0.693</td>
</tr>
</tbody>
</table>

The Table 6 reports the results of a cross-sectional multiple linear regression relating models’ EDPs to the companies’ debt ratio, volatility, and idiosyncratic risk. *Confidence level at 1%.

below 50%). The EDPs are more sensitive to the debt ratio than to the volatility. The first-passage models are more sensitive to debt ratio than the M model. All regressions are statistically significant at the 1% level.

Table 5 reports the results of the relationship between idiosyncratic risk and EDPs. The explanatory power of the regressions increases substantially (approximately 70%). The debt ratio and idiosyncratic risk explain approximately 77% of the variability of EDP of the M model. These results confirm the significance of the idiosyncratic risk in explaining bankruptcy.

Table 6 presents the results of the tests incorporating all three variables: debt ratio, idiosyncratic risk, and volatility. Compared to the results in Table 5, the explanatory power of the regressions has not improved. It is clear, however, that idiosyncratic risk is the most important variable. The coefficient of idiosyncratic risk is statistically significant. It is surprising to
observe the coefficient of volatility becomes smaller and statistically insignificant in the LS and ER models.

5.4. Information-Related Tests

Based on Shumway (2001), we assume that the relationship between EDPs and independent variable(s) is represented by a logistic curve that asymptotically approaches one (zero) as covariates tend to positive (negative) infinity. This relationship is written as follows:

\[ P_{t-1}(Y_{it} = 1) = \frac{1}{[1 + \exp(-(z + \beta X_{i,t-1}))]} \]

where \( X_{i,t-1} \) is the vector of time varying covariates, known at the end of previous year, \( z \) denotes the constant, \( Y_{it} \) is the dependent variable, EDPs, and which equals one when a company goes bankrupt and zero otherwise. Each year that a company is alive corresponds to an observation in the estimation equation.

The logistic regressions analysis complements the prediction-oriented tests. This method, however, has several limitations: first, it assumes a dichotomous decision (0 or 1 value). Second, it does not distinguish the relative importance of Error type I and II for credit risk management. For a credit risk manager, it is more serious to have a bankrupt firm classified as nonbankrupt than a nonbankrupt firm classified as bankrupt. Third, the classification of firms as bankrupt or as nonbankrupt is somewhat subjective because it implies the definition of a cut-off value. Fourth, it is not clear which model explains the variability of companies’ default risk better. Moreover, this procedure has the limitation of considering bankruptcy as an event and not as a process.

Table 7 reports the results of logistic regressions. Columns 1–8 display univariate regressions with EDPs of the structural models and with idiosyncratic risk, debt ratio, market-to-book assets ratio (MB), book-to-market equity ratio (BE), and volatility. According to Vassalou and Xing (2004), default risk is explained by the BE ratio, while MB ratio is introduced as a proxy for companies’ growth opportunities. We use Nagelkerke \( R^2 \) as an indicator of the explanatory power.

All models are statistically significant and the coefficients have the expected signs. The M model has the highest explanatory power (around 40%), which contrasts with the results of the prediction-oriented tests reported in Chapter 5. The ER model also shows a good performance and is
<table>
<thead>
<tr>
<th>Table 7. Logistic Regressions.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
</tr>
<tr>
<td>M</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>LS</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>ER</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Idiosyncratic risk</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Debt ratio</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>MB assets</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>BM equity</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Volatility</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Constant</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Observations</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>-2logL</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Nagelkerke $R^2$</td>
</tr>
</tbody>
</table>

Two-sided p-values are in parentheses. The -2logL statistic has a $\chi^2$ distribution with $n-q$ degrees of freedom, where $q$ is the number of parameters in the model. In all logistic regressions, we cannot reject the null hypothesis, implying that the model fits the data (the $\chi^2$ statistics are corrected according to Shumway (2001) suggestions).
very close to the M model. The LS model has a comparatively lower performance. Overall, the performance of these models is better than that reported by Campbell, Hilscher, and Szilagyi (2006) at the 1-month horizon. Considering that the financial and accounting data for some failed companies suffer from observation lags of up to 2 years, these results are highly encouraging.

It is worth drawing attention to the estimated values of the coefficients of idiosyncratic risk and debt ratio. Both these variables have the expected sign and are significant at the 1% confidence level. The default risk, however, appears not to be sensitive to volatility, and the explanatory power of this coefficient is almost zero. These results show that idiosyncratic risk is an important variable. Comparing the explanatory power of idiosyncratic risk (column 4) and volatility (column 8), it is evident that the former variable has greater power in predicting bankruptcy events than the latter variable. This suggests that investors are aware of the specific circumstances responsible for a company’s deterioration and anticipate bankruptcy. The coefficient of debt ratio is significant, as expected. It is worth noting the lower explanatory power of the ER and LS models, which, given the complexity of these models, is somewhat intriguing. In contrast to the results reported by Vassalou and Xing (2004), the explanatory power of the BE ratio is almost zero, and this variable is not significant. The MB ratio has the expected sign and is significant at 5% confidence level, but its explanatory power is very low (see columns 6 and 7).

In columns 9 and 10, we use a stepwise procedure with the M/ER models and the other variables. Idiosyncratic risk and debt ratio enter in regressions and the former variable always dominates the other variables, and it is statistically more significant than the debt ratio. Finally, according to the $-2\log L$ statistic that has a $\chi^2$ distribution with $n-q$ degrees of freedom where $q$ is the number of parameters in the model, we cannot reject the null hypothesis of logistic regressions, implying that the model fits the data.

5.5. Factor Analysis of Correlation Matrix

Next we present results of the joint variability of companies default risk, that is, of the companies default correlation matrix, based on factor-analytical tests (see Appendix B for details). We compute a correlation matrix per model and fit copula functions by maximizing the log-likelihood function, as explained in Eq. (4), to each pair of EDPs. Given the restrictions outlined in Section 4 above, the sample comprises 23 companies,
which are 25 bankruptcy events (remember two companies failed twice) and 276 copula functions with matching time frames (see footnote 8). The results (based upon the Akaike and Schwartz goodness-of-fit criteria) show that all fitted copula functions belong to the normal family. Next, we construct companies’ correlation matrix with Kendall's Tau. This correlation matrix is used in the factor analysis to estimate the determinants of default correlations. We employ the principal components method to extract the factors from the correlation matrix. We retain the factors that have an eigenvalue greater than one.

Table 8 reports the results of the factor analysis for each model (see Appendix B for details). We extract five factors for the M and six factors the LS and ER models. The RMSR and the nonredundant residuals of the residual matrix are small in all models, implying a good factor solution. The five factors of the M model and the six factors of the LS and ER models, referred to as common or latent factors, explain a high percentage of the observed variance (79.5%, 86.5%, and 82.8%, respectively). This is an encouraging result for our search for the determinants of default correlation. In contrast with Zhou (2001), the results suggest that only a small percentage (21.5%, 13.5%, and 17.2% for M, LS, and ER models, respectively) of observed variance or default correlation is explained by nonretained factors. An orthogonal rotation (Varimax rotation) is

<table>
<thead>
<tr>
<th>Factors</th>
<th>M Model</th>
<th>LS Model</th>
<th>ER Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Eigenvalue</td>
<td>Cumulative %</td>
<td>Eigenvalue</td>
</tr>
<tr>
<td>1</td>
<td>7.6</td>
<td>32.8</td>
<td>7.1</td>
</tr>
<tr>
<td>2</td>
<td>5.8</td>
<td>57.9</td>
<td>4.9</td>
</tr>
<tr>
<td>3</td>
<td>2.0</td>
<td>66.5</td>
<td>2.8</td>
</tr>
<tr>
<td>4</td>
<td>1.9</td>
<td>74.8</td>
<td>2.4</td>
</tr>
<tr>
<td>5</td>
<td>1.1</td>
<td>79.5</td>
<td>1.5</td>
</tr>
<tr>
<td>6</td>
<td>1.1</td>
<td>86.5</td>
<td>1.0</td>
</tr>
<tr>
<td>RMSR</td>
<td>0.057</td>
<td>0.039</td>
<td>0.048</td>
</tr>
<tr>
<td>Nonredundant residuals</td>
<td>7.1%</td>
<td>1.2%</td>
<td>3.2%</td>
</tr>
</tbody>
</table>

Root mean square residual (RMSR) = $\sqrt{\sum_{p=1}^{\rho} \sum_{j=1}^{\rho} \text{res}^2_{ij} / (\rho(\rho - 1)/2}$, where $\rho$ is the number of companies and res gives the amount of correlation that is not explained by the retained factors. Nonredundant residuals are computed as a percentage of the number of nonredundant residuals with absolute values greater than 0.10.
performed to achieve a simpler factor structure. We use the rotated component matrix to estimate time series values for each model’s factors.

Since it is difficult to interpret and analyse the determinants of default correlations with so many factors, we use principal component analysis to reduce the factors. We use the rule of eigenvalue greater than one to retain the new factors. The initial five common factors of the M model and six common factors of LS and ER models are reduced to two common factors for each model. The two common factors explain around 90% of the total variability of its initial common factors (see Table 9).

Table 10 presents the determinants of default correlations. We estimate Eq. (7) using a stepwise estimation procedure because of multicollinearity problems. As expected, given the parameter assumptions in the stepwise procedure, all the variables and all regressions are statistically significant, at the 5% confidence level. Overall, the regressors’ explanatory power is very high, around 55%, with a maximum of 71%. Further, the signs of the estimated coefficients are generally as expected in Table 2. Market volatility and equity premium explain 56% of the variability of the common factor 1 in the M model. This factor can be interpreted as the capital market effect. The equity premium has the expected sign and market volatility has a negative effect on default correlation. Surprisingly, volatility does not explain the variability of default correlation in the first-passage models. The slope of the yield curve explains the variability of the common factor 2 in the M model, but the estimated coefficient does not have the expected sign. One possible explanation for this is that an increase at the slope of the yield curve makes it more difficult for distressed firms to renegotiate the debt, and this possibly increases the default risk and default correlations. Only the common factor 1 in the ER model has the expected sign. Common factor 2

---

**Table 9. Principal Components.**

<table>
<thead>
<tr>
<th>PCA Factors</th>
<th>M Model</th>
<th>LS Model</th>
<th>ER Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Eigenvalue</td>
<td>Percentage of variance</td>
<td>Eigenvalue</td>
</tr>
<tr>
<td>1</td>
<td>2.7</td>
<td>55.0</td>
<td>2.9</td>
</tr>
<tr>
<td>2</td>
<td>1.7</td>
<td>34.3</td>
<td>2.1</td>
</tr>
<tr>
<td>Cumulative percentage explained by PCA</td>
<td>89.3</td>
<td>88.4</td>
<td>91.2</td>
</tr>
</tbody>
</table>

The Table 9 presents the principal components of a principal component analysis (PCA) applied to the factors extracted from the defaults correlation matrix.
is explained by both the treasury interest rates and market equity premium, which allows us to interpret them as a return-driven factor. Consistent with Longstaff and Schwartz (1995) and Collin-Dufresne et al. (2001), we find that the effect of an increase in the risk-free rate is to lower EDPs and default correlations. The estimated coefficients of equity premium are of the same magnitude in all models. The default return spread, which is a proxy for default contagion, is not significant in any of the regressions, suggesting that either it does not explain default correlations, as argued by Schonbucher (2003) and Collin-Dufresne et al. (2003), or that this variable is not a good proxy. We should point out that, given the nature of this effect, it is probably better to capture this effect by a nonsystematic variable or a variable that considers the company’s business and financial ties. Convexity is not statistically significant in any of the regressions, which is consistent with the findings of Collin-Dufresne et al. (2001). The Ljung-Box test indicates that standardized residuals from the regressions are not autocorrelated; the average serial correlation of standardized residuals is 0.02, and the average Durbin-Watson statistic is 1.81.

Table 10. Determinants of Default Correlation Factors.

<table>
<thead>
<tr>
<th></th>
<th>M Model</th>
<th>LS Model</th>
<th>ER Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fact 1</td>
<td>Fact 2</td>
<td>Fact 1</td>
</tr>
<tr>
<td>Intercept</td>
<td>1.54</td>
<td>-0.81</td>
<td>-0.88</td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td>(0.027)</td>
<td>(0.011)</td>
</tr>
<tr>
<td>$r_{it}^{10}$</td>
<td>n. e.</td>
<td>n. e.</td>
<td>n. e.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$(r_{it}^{10})^2$</td>
<td>n. e.</td>
<td>n. e.</td>
<td>n. e.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r_{it}^{10} - r_{it}^2$</td>
<td>80.78</td>
<td>n. e.</td>
<td>88.10</td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td></td>
<td>(0.002)</td>
</tr>
<tr>
<td>$\sigma_{S&amp;P}$</td>
<td>-8.00</td>
<td>n. e.</td>
<td>n. e.</td>
</tr>
<tr>
<td></td>
<td>(0.030)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R_{M} - r_{it}^{1m}$</td>
<td>-2.59</td>
<td>n. e.</td>
<td>n. e.</td>
</tr>
<tr>
<td></td>
<td>(0.025)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DefSpread</td>
<td>n. e.</td>
<td>n. e.</td>
<td>n. e.</td>
</tr>
<tr>
<td>Adj. $R^2$</td>
<td>0.56</td>
<td>0.43</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.007)</td>
<td>(0.002)</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.007)</td>
<td>(0.002)</td>
</tr>
</tbody>
</table>

For each model $s$ and each default correlation factor, 1 or 2, we estimate the following regression: \( \text{Fact}_s = \alpha + \beta_1 r_{it}^{10} + \beta_2 (r_{it}^{10})^2 + \beta_3 (r_{it}^{10} - r_{it}^2) + \beta_4 \sigma_{S&P} + \beta_5 (R_{M} - r_{it}^{1m}) + \beta_6 \text{DefSpread}_t \), using a stepwise procedure. Beneath the variables, in parenthesis, we report significance values. n. e. denotes not entered in the regression.
In summary, default correlations are driven essentially by common factors that explain on average around 83% of total variance. Only a small percentage of default correlation is due to nonretained factors. These results are consistent with economic intuition and existing empirical evidence (see Vassalou and Xing, 2004), according to which during periods of recession default risk increases and clusters of bankruptcy events are observed. The factors driving default correlations are the capital market equity premium and treasury interest rates, which reflect the overall state of the economy. This is consistent with the theoretical intuition of Hull et al. (2005) when they argue that the systematic variable that drives default correlations is a capital market Wiener process. Second, the slope of the yield curve reflects investors’ expectations about the evolution of the economy. So, as long as default correlations are basically driven by systematic factors, portfolio diversification should be able to reduce default risk.

6. CONCLUSION

In this study we analyse the determinants of default correlations for a sample of the US bankrupt companies. We apply a set of structural models (Merton, 1974; Longstaff and Schwartz, 1995; Ericsson & Reneby, 1998) to estimate companies’ EDPs. Given that we observe a sharp increase in EDPs up to 2 years in advance of default event, these models provide timely and accurate estimates of companies’ default risk.

Another novel finding is the importance of idiosyncratic risk (and not total volatility) in predicting default events. This suggests that company-specific signals provide useful information to investors about the deterioration in company’s economic and financial conditions prior to bankruptcy.

We compute companies’ default correlation matrices using the copula function approach and employ factor analysis techniques to extract factors that explain default correlations. The results of prediction-oriented tests suggest that the ER model is the best model as it misclassifies only 5.5% of bankruptcy events. The results of information-related tests suggest that the M and LS models have a similar performance. Variables such as MB asset ratio and BE ratio, which other studies have found to be significant, have poor explanatory power in our regression analysis. We observe that common factors explain around 83% of the variability of default correlations. This evidence supports the belief that common factors are explained by the overall state of the economy and by the expectations of its evolution.
NOTES

1. CreditMetrics was developed by RiskMetrics Group. KMV was developed by KMV Corporation.
2. CreditRisk + was developed by Credit-Suisse Financial Products.
3. A fuller exposition is available in Frees and Valdez (1998, Nelsen (1999), and Costinot, Roncalli and Teiletche (2000).
4. See Mendes and Souza (2004) for an example of the fitting process. The authors assume that the margins of IBOVESPA and S&P500 follow a t-Student distribution and fit four copulas: the t-student, the BB1, the Gumbel, and the Gaussian copula.
5. We solve Ito’s equations

\[
\sigma_v = E_t(V_t, \sigma_v, T - t)/V_t \sigma_E N(d_1)
\]

and

\[
E_t(V_t, \sigma_v, T - t) = \hat{E}_t
\]

where \(E_t(V_t, \sigma_v, T - t)\) is the theoretical value of company’s assets, \(\sigma_E\) the volatility of equity, \(N(\cdot)\) the standard normal distribution function, and \(\hat{E}_t\) denotes the observed market value of equity.

6. This value is the weighted average, by the average leverage ratio of all S&P 500 firms, between the observed dividend yield and historical coupon rate (during the period 1973–1998). Huang and Huang (2003) also argue that the use of one payout ratio for firms with different credit ratings is not erroneous given that, probably, firms with lower credit rating may have higher debt payouts than the ones with higher credit rating but they are also likely to make less payment to shareholders.

7. We thank Kenneth French for making available this data on his web page: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french

8. We have 276 copulas because of the following reason: there are 25 bankruptcy events consisting of 23 companies, 2 of which went bankrupt twice. This leaves us with 24 events having sufficient data to form bivariate copulas. Since the copulas formed are bivariate copulas we wind up with combinations of 24 items taken two at a time which provides \(24 \times 23/2 = 276\) copulas.

9. Collin-Dufresne et al. (2001) define this variable as the difference between BBB index yield and 10-year treasury yield, which can bias the spread since these two classes of securities have different degrees of liquidity.

10. Moody’s Report (2005) presents the term structure of default rates over several period of time. Default probability of a Caa-C firm, during the period 1920–2004, at 1-year horizon, was around 15%. For the period 1983–2004, at the same horizon, was around 22%. Standard and Poor’s transition probability from CCC to default is around 19.8% (see Crouhy, Galai, & Mark, 2000).

11. This nonlinear relationship can be rewritten as a linear one \(\ln[P_{t-1}/(1-P_{t-1})] = \alpha + \beta X_{t-1}\), where the dependent variable represents the log of the odds.

12. One way to solve this problem is to use lag values in logistic regression. We did not use this procedure because it would entail loss of observations and because fixing a number of lagged values introduces bias.
13. Based on Patel and Pereira (2007), we also perform logistic regressions with model scores. We do not show these results because they are very similar to the ones reported.

14. Several copula families were also fitted including the normal and extreme values families.

15. Several studies (e.g., Collin-Dufresne et al., 2001) argue that default probabilities can be explained by nonlinear, cross term, and lagged values of regressors (such as squared and cubic slope of the yield curve or \((r_{10} \cdot r_{20}) \sigma_{\text{S&P}}\)). However, none of these terms seems to explain default correlations and that is why we restrict this analysis to the variables in Eq. (7).

ACKNOWLEDGMENT

Ricardo Pereira thanks the financial support from Fundação para a Ciência e Tecnologia.

REFERENCES


Determinants of Default Correlations


APPENDIX A. STRUCTURAL MODELS

In this section, we present a brief summary of the models by Merton (1974), Longstaff and Schwartz (1995), and Ericsson and Reneby (1998). Since our main concern here is with the empirical performance of these models, we do not discuss in detail the theoretical properties of the models. Throughout this section, we assume that uncertainty in the economy is modeled by a filtered probability space \((\Omega, \mathcal{G}, P)\), where \(\Omega\) represents the set of possible states of nature, \(\mathcal{G}_t\) is the information available to investors over time \(t\), and \(P\), the probability measure. All models assume a perfect and arbitrage-free capital market, where risky and default-free bonds and companies’ equity are traded. The risk-free numeraire (or money market account) value, at time \(t\), \(A_t\), follows the process

\[
A_t = \exp\left(\int_0^t r_s ds\right)
\]
where $r$ denotes the short-term risk-free interest rate, which can be deterministic or modeled by a stochastic process. When modeled as a stochastic process, the dynamics of $r$ is driven by a Vasicek model

$$dr_t = a(\lambda - r_t)dt + \sigma_r dW^r$$  \hspace{1cm} (A.1)

where $a$ is the short-term interest rate mean reversion speed, $\lambda$ and $\sigma_r$ are its mean reversion level and standard deviation, respectively. The variable $dW^r$ is a Wiener process. In this economy, the investors are assumed to be risk neutral, which means that the probability measure, $P$, is a martingale with respect to $A_t$. The value of a riskless discount bond that matures at $T$ is (Vasicek, 1977)

$$D(r, T) = \exp(A(T) - B(T)r)$$  \hspace{1cm} (A.2)

$$A(T) = \left(\frac{\sigma_r^2}{2a^2} - \lambda\right) T + \left(\frac{\sigma_r^2}{a^2} - \frac{a\lambda}{a^2}\right)(\exp(-aT) - 1) - \left(\frac{\sigma_r^2}{4a^2}\right)(\exp(-2aT) - 1)$$

$$B(T) = \frac{[1 - \exp(-aT)]}{a}$$

Under the risk-neutral probability space, the value of the company’s assets, $V$, follows a geometric Brownian motion ($G_t$–adapted diffusion process) given by

$$dV_t = (r_t - \delta)V_t dt + \sigma_v V_t dW^v$$  \hspace{1cm} (A.3)

where $\delta$ denotes company’s assets payout ratio and $\sigma_v$ company’s assets volatility. The variable $dW^v$ is a Wiener process under the risk-neutral probability measure. $\rho$ is the instantaneous correlation coefficient between $dW^r$ and $dW^v$.

The dynamics of company’s assets value, under the real probability space, is given by

$$dV_t = (\mu - \delta)V_t dt + \sigma_v V_t dW^{Pv}$$

where $\mu$ denotes company’s assets expected total return and $dW^{Pv}$ is a Wiener process under the real probability measure. For the dynamic process described by Eq. (A.3), and the given assumptions, the standard hedging framework leads to the following partial differential equation

$$\frac{1}{2}\sigma_v^2 V^2 \frac{\partial^2 F}{\partial V^2} + (r - \delta)V \frac{\partial F}{\partial V} - rF + \frac{\partial F}{\partial t} + P = 0$$
where \( F \) is the price of any derivative security, whose value is a function of the value of the firm, \( V \), and time, and \( P \) represents the payments received by this security. The two-factor models by Longstaff and Schwartz (1995) assume that \( F \) is a function of the value of the firm, \( V \), time and interest rates.

The standard hedging framework leads to the following partial differential equation:

\[
\frac{1}{2} \sigma_v^2 V \frac{\partial^2 F}{\partial V^2} + \rho \sigma_v \sigma_r V \frac{\partial F}{\partial V} \frac{\partial F}{\partial r} + \frac{1}{2} \sigma_r^2 \frac{\partial^2 F}{\partial r^2} + (r - \delta) V \frac{\partial F}{\partial V} + (\lambda \sigma_r - ar) \frac{\partial F}{\partial r} + \frac{\partial F}{\partial t} - rF + P = 0
\]

Given our focus on the empirical performance of structural models in predicting corporate failure, we outline only the equations relevant for the expected default probability (EDP) in each model. We refer the reader to the original papers for the full description the models.

**Merton (M) Model**

Merton (1974) model is an extension of the Black and Scholes (1973) option pricing model to value corporate securities. The company’s assets value, which corresponds to the sum of the equity and debt values, is driven by the process described by Eq. (A.3) and is assumed to be independent of company’s capital structure. Under these assumptions, equity value, \( E_t \), is defined by a call option on the assets of the firm, with maturity \( T \) and exercise price \( F \):

\[
E_t = V_t N(d_1) - e^{-r(T-t)}F N(d_2)
\]

(A.4)

where

\[
d_1 = \frac{\ln(V_t/F) + (r + 0.5\sigma_v^2)(T-t)}{\sigma_v \sqrt{T-t}}, \quad d_2 = d_1 - \sigma_v \sqrt{T-t}
\]

and \( N(\cdot) \) represents the standard normal distribution function. Debt’s value, at time \( t \), is equal to

\[
D_t = V_t - E_t
\]

If at maturity, company’s assets value, \( V_T \), is higher than the face value of its debt, \( F \), the firm does not default, bondholders receive \( F \) and shareholders
However, if $V_T < F$, the firm defaults and there is a transfer of company’s ownership from shareholders to bondholders. Firm only defaults at time $T$, and $N(-d_2)$ represents the risk-neutral probability of default.

**Longstaff and Schwartz (LS) Model**

Longstaff and Schwartz (1995) developed a two-factor model to value risky debt, extending the one-factor model of Black and Cox (1976) in two ways: (i) incorporating both default risk and interest rate risk; (ii) allowing for deviations from strict absolute priority. An important feature of this model is that firms with similar default risk can have different credit spreads if their assets have different correlations with changes in interest rates. Their assumptions are not very different from the ones used by Black-Scholes, Merton (1974), and Black and Cox (1976), except for the fact that short-term risk-free interest rate follows the dynamics described by Eq. (A.1) (and the riskless discount bond can be priced using Eq. (A.2)) and that there are bankruptcy costs, $z$. The default boundary, $K$, is constant and exogenously specified, which is consistent with the assumption of a stationary capital structure. Setting $X$ equal to the ratio $V/K$, the price of a risky discount bond that matures at $T$ is

$$D(X, r, T) = D(r, T) - zD(r, T)Q(X, r, T)$$  \hspace{1cm} (A.5)

where

$$Q(X, r, T, n) = \sum_{i=1}^{n} q_i$$  \hspace{1cm} (A.6)

$$q_1 = N(a_1)$$

$$q_i = N(a_i) - \sum_{j=1}^{i-1} q_j N(b_{ij}) \quad i = 2, 3, \ldots, n$$

$$a_i = \frac{-\ln X - M(iT/n, T)}{\sqrt{S(iT/n)}} \quad b_{ij} = \frac{M(jT/n, T) - M(iT/n, T)}{\sqrt{S(iT/n) - S(jT/n)}}$$
and where

\[ M(t, T) = \left( \frac{a\lambda - \rho \sigma_v \sigma_r}{a} - \frac{\sigma_r^2}{2a^2} \right) t + \left( \frac{\rho \sigma_v \sigma_r}{a^2} + \frac{\sigma_r^2}{2a^2} \right) \exp(-aT)(\exp(at) - 1) \]

\[ + \left( \frac{r}{a} - \frac{a\lambda}{a^2} + \frac{\sigma_r^2}{a^3} \right)(1 - \exp(-at)) - \left( \frac{\sigma_r^2}{2a^3} \right) \exp(-aT)(1 - \exp(-at)) \]

\[ S(t) = \left( \frac{\rho \sigma_v \sigma_r}{a} + \frac{\sigma_r^2}{a^2} + \frac{\sigma_r^2}{a^3} \right) t - \left( \frac{\rho \sigma_v \sigma_r}{a^2} + \frac{2\sigma_r^2}{a^3} \right) \]

\[ \times (1 - \exp(-at)) + \left( \frac{\sigma_r^2}{2a^3} \right) (1 - \exp(-2at)) \]

The term \( Q(X, r, T) \) is the limit of \( Q(X, r, T, n) \) when \( n \to \infty \) (the authors argue that the convergence between these terms is rapid and that when \( n = 200 \), the differences between the results of the terms are virtually indistinguishable).

The first term in Eq. (A.5) represents the value of a riskless bond. The second term represents a discount factor for the default of the bond. The factor can be decomposed into two components: \( \alpha D(r, T) \) is the present value of the writedown on the bond if default occurs; \( Q(X, r, T) \) is the probability, under the risk-neutral measure, that a default occurs (this probability can differ from the real one).

**Ericsson and Reneby (ER) Model**

Ericsson and Reneby (1998) demonstrate that corporate securities can be valued as portfolios of three basic claims: a down-and-out option that expires worthless if the underlying variable reaches a pre-specified lower boundary, prior to the expiration date; a down-and-out binary option that yields a unit payoff at the expiration date if the underlying asset exceeds the exercise price; and unit down-and-in option that pays off one unit the first time the underlying variable reaches a lower boundary. This formulation allows to value finite maturity coupon debt with bankruptcy costs, corporate taxes, and deviations from the absolute priority rule. The default is triggered if company’s value falls below a constant \( K \) (the reorganization barrier), at any time prior to maturity of the firm, or if, at debt’s maturity, company’s value is less than some constant \( F \), which normally is debt’s face value.
value. The time of default is denoted as \( \tau \). The price of a unit down-and-in option that matures at \( T \) and pays one monetary unit if bankruptcy happens before \( T \) and 0 otherwise, is

\[
G^K(V_\tau, t|\tau \leq T) = G^K(V_\tau|\tau \leq \infty)(1 - Q^G(t > T, V_i > K)) \tag{A.7}
\]

where

\[
G^K(V_\tau|\tau \leq \infty) = \left( \frac{V_t}{k} \right)^{-0}
\]

\[
Q^G(t > T, V_i > K) = N \left\{ d^G_T \left( \frac{K^2}{VK} \right) \right\} - \left\{ \frac{V}{K} \right\}^{-2/\sigma \mu_X^G} N \left\{ d^G_T \left( \frac{K^2}{VK} \right) \right\}
\]

\[
d^G_T(x) = \frac{\ln x}{\sigma \sqrt{t}} + \mu_X^G \sqrt{t}
\]

\[
\mu_X^B = \frac{r - \delta - 0.5\sigma^2}{\sigma} \quad \mu_X^G = \mu_X^B - \theta \sigma \quad \theta = \frac{\sqrt{(\mu_X^B)^2 + 2r + \mu_X^B}}{\sigma}
\]

Eq. (A.7) represents the EDP.

**APPENDIX B. FACTOR ANALYSIS**

A full understanding of factor analysis can be obtained by reading Sharma (1995), for example. Factor analysis uses the correlation matrix to identify the smallest number of common factors (via factor rotation) that best explain the correlation among the variables; and provide an interpretation for these common factors.

This technique assumes that the total variance of a variable can be divided into the variance explained by the common factor and the one explained by a specific factor. A factor model that contains \( m \) factors can be
represented as

\[
\begin{align*}
x_1 &= \lambda_{11} \xi_1 + \lambda_{12} \xi_2 + \cdots + \lambda_{1m} \xi_m + \epsilon_1 \\
x_2 &= \lambda_{21} \xi_1 + \lambda_{22} \xi_2 + \cdots + \lambda_{2m} \xi_m + \epsilon_2 \\
&\vdots \quad \vdots \quad \vdots \\
x_p &= \lambda_{p1} \xi_1 + \lambda_{p2} \xi_2 + \cdots + \lambda_{pm} \xi_m + \epsilon_p
\end{align*}
\]

where \(x_1, x_2, \ldots, x_p\) are variables of the \(m\) factors, \(\lambda_{pm}\) is the pattern loading of the \(p\)th variable on the \(m\)th factor, and \(\epsilon_p\), the specific factor for the \(p\)th variable. The previous construct can be represented as

\[
x = \Lambda \xi + \epsilon
\]

Eq. (B.1) is the factor analysis equation. The assumptions are that the common factors are not correlated with the specific factors, and the means and variances of variables and factors are zero and one, respectively. Variables’ correlation matrix, \(R\), is

\[
R = \Lambda \Phi \Lambda' + \Psi
\]

\(\Lambda\) is the pattern loading matrix, \(\Phi\), factors’ correlation matrix, and \(\Psi\), a diagonal matrix of the specific variances. \(R - \Psi\) gives us the variance explained by the common factors. The off-diagonals of \(R\) are the correlation among variables. Factor analysis estimate parameter matrices given the correlation matrix. The correlation between the variables and the factors is given by

\[
\Lambda = \Lambda \Phi
\]

If the \(m\) factors are (not) correlated, the factor model is referred to as an oblique (orthogonal) model. In an orthogonal model, it is assumed that \(\Phi = I\). Orthogonal rotation technique implies the identification of a matrix, \(C\), such that the new loading matrix is given by \(\Lambda^* = \Lambda C\) and \(R = \Lambda^* \Lambda^{*\prime}\).

Varimax rotation technique estimate matrix \(C\) such that each factor will be a set of different variables. This is achieved by maximizing the variance of the squared loading pattern across variables, subject to the constraint that the communality of each variable is unchanged. \(C\) is obtained maximizing
the following equation, subject to the constraint that the common variance of each variable remains the same.

\[ pV = \sum_{j=1}^{m} \sum_{i=1}^{p} \lambda_{ij}^4 - \frac{\sum_{j=1}^{m} \left( \sum_{i=1}^{p} \lambda_{ij}^2 \right)^2}{p} \]

where \( V \) is the variance explained by the common factors.
DATA MINING PROCEDURES IN GENERALIZED COX REGRESSIONS

Zhen Wei

ABSTRACT

Survival (default) data are frequently encountered in financial (especially credit risk), medical, educational, and other fields, where the “default” can be interpreted as the failure to fulfill debt payments of a specific company or the death of a patient in a medical study or the inability to pass some educational tests.

This paper introduces the basic ideas of Cox’s original proportional model for the hazard rates and extends the model within a general framework of statistical data mining procedures. By employing regularization, basis expansion, boosting, bagging, Markov chain Monte Carlo (MCMC) and many other tools, we effectively calibrate a large and flexible class of proportional hazard models.

The proposed methods have important applications in the setting of credit risk. For example, the model for the default correlation through regularization can be used to price credit basket products, and the frailty factor models can explain the contagion effects in the defaults of multiple firms in the credit market.
Survival (default) data are frequently encountered in financial (especially credit risk), medical, educational, and other fields, where the “default” can be seen as the failure to fulfill debt payments of a specific company or the death of a patient in a medical study or the inability to pass some educational tests.

Survival data usually consist of either cross-sectional data in the form of the triplet \((x_{ij}, z_{ij}, \delta_{ij})\), for \(i = 1, \ldots, m\) and \(j = 1, \ldots, J_i\), where

- \(m\) is the total number of groups and \(J_i\) is the number of subjects in the \(i\)th group.
- \(x_{ij}\) is a vector of covariates that is included to build the survival model. It can be random.
- \(z_{ij}\) is the observed failure time or censoring time, whichever comes first.
- \(\delta_{ij}\) is the default indicator for the \(j\)th subject in the \(i\)th group.

The survival data can also consist of the time series type data \((x_{ijt}, \delta_{ijt})\), for \(i = 1, \ldots, m\), \(j = 1, \ldots, J_i\), and \(t = t_{ij}, \ldots, T_{ij}\), where \(m\), \(J_i\), \(x_{ijt}\), and \(\delta_{ijt}\) are similarly defined and the covariates \(x_{ijt}\) can be stochastic processes. Our goal is to make prediction on the survival probability of each subject (entity) and the survival correlation within a pool of them using the information from the covariates. Various models are proposed, built on the past data, and their predictive powers and accuracies are assessed. One important class of models prevailing in medical and biological practices is the Cox’s (proportional) model, which also has potential applications in credit risk.

This paper introduces the basic ideas of Cox’s original proportional model for the hazard rates and extends the model within a general framework of statistical data mining procedures. Traditionally, these models are calibrated using the martingale approach, which is based on theories in counting processes (Andersen & Gill, 1982; Allen, 1971). In contrast, this paper presents a pool of various procedures that are based solely on maximum (partial) likelihoods without using the martingale properties. Moreover, the data mining procedures described in this paper can also be employed to solve other regression/classification problems that are beyond the scope of survival analysis.

The rest of the paper is divided in two parts. The first part introduces various statistical data mining procedures for (generalized) Cox regression with time-independent covariates (for cross-sectional type data). The second
part deals with time-dependent covariates. It can be seen that the tools used in Part I can also be applied to Part II and vice versa. Although the described procedures are very efficient for model calibration and many of them have very profound theoretical backgrounds, we omit most of the theoretical proofs for the purpose of emphasizing the methods themselves. For example, there is a universal “Oracle property” for many of the regularization methods that states the asymptotic efficiency of the estimation and selection procedures. Whenever possible, we present the algorithm as a recipe in several iterative steps, so that the reader can easily implement the ideas.

2. PART I: GENERALIZED COX REGRESSION WITH TIME-INDEPENDENT COVARIATES

For a brief review of the definitions in classical survival analysis, I refer the reader to Appendix A. Under the settings in Appendix A, our goal is to predict the survival probability over time of a particular entity given the current status of covariates. Furthermore, by separating the elements of the covariate $x$ into group of systematic and idiosyncratic factor components, one may further explore the correlation between the defaults of multiple names.

In this part, we only consider a single observation of default timing for each entity, which should be useful for cross-sectional data in default modeling or survival data for clinical (medical) experiments. Later we will extend the model to time series data where the covariates could also have a stochastic feature.

2.1. Generalized Cox Hazard Models

2.1.1. Proportional Hazard Model

Let us first consider the model for hazard rate:

$$h(t) = h_0(t) \exp(x^T \beta)$$

for some baseline function $h_0(t)$. To estimate the survival probability of a given subject or correlation among the defaults, it suffices to estimate the parameter $\beta$ (condition on current state of the covariates).
Given the observed data \( \{(x_i, z_i, \delta_i): i = 1, \ldots, n\} \), the likelihood function is given by (the defaults are independent given the covariates)

\[
L = \prod_{\delta_i=1} \frac{f(z_i|x_i)}{f(z_i|x_i) S(z_i|x_i)} = \prod_{\delta_i=1} \frac{h(z_i)}{h(z_i) S(z_i|x_i)}
\]

\[
= \prod_{\delta_i=1} h_0(z_i) \exp(x_i^T \beta) \prod_{i=1}^n \exp(-H_0(z_i) \exp(x_i^T \beta))
\]

where

\[
H_0(t) = \int_0^t h(u)du
\]

is the cumulative baseline hazard function.

2.1.2. Partial Likelihood Function

Usually, it is not easy to directly maximize the criterion (2). Breslow (1974) assumes that \( H_0 \) is a step function that jumps only at censored observations:

\[
H_0(t) = \sum_{\delta_j=1} h_j I(z_j \leq t)
\]

Then, the logarithm of the likelihood function (2) is

\[
\sum_{\delta_j=1} (\log(h_j) + x_j^T \beta) - \sum_{i=1}^n \left( \sum_{\delta_j=1} h_j I(z_j \leq z_i) \exp(x_j^T \beta) \right)
\]

Taking the derivative with respect to \( h_j \) generates

\[
\hat{h}_j = \left( \sum_{i=1}^n I(z_j \leq z_i) \exp(x_i^T \beta) \right)^{-1}
\]

for \( \delta_j = 1 \). Plugging \( \hat{h}_j \) into (2) generates, up to a scalar product, the so-called partial likelihood function

\[
\mathcal{L}(\beta) = \prod_{\delta_i=1} \frac{\exp(x_i^T \beta)}{\sum_{j=1}^n I(z_i \leq z_j) \exp(x_j^T \beta)}
\]
We see that it is much more easier to examine the properties of the partial likelihood function (3) than the likelihood function (2) itself.

2.1.3. Generalized Proportional Models, Parameter Regularization, and Boosting

It is natural to extend the proportional model (1) to the following form

\[ h(t) = h_0(t) \exp(\eta(x)) \]  

(4)

where \( \eta(\cdot) \) can be a generic function. The linear form \( x^T \beta \) can be seen as the first-order parameter expansion of \( \eta \). It is also easy to see that the partial likelihood function for (4) is given by

\[ L(\beta) = \prod_{\delta_i=1} \sum_{j=1}^{n} I(z_i \leq z_j) \exp(\eta(x)) \]

Since in real situations the dimensionality of the covariate is high, it is unrealistic (or less interpretable) to build a model by directly maximizing (3). For parametric model (1), we can usually use the idea of shrinkage (or regularization) to confine the parameters in a reasonable subspace. L1 regularization (LASSO) is one of the most popular methods, which by controlling the absolute sum of the parameters, can often do the job of estimation and variable selection at the same time.

\[ \ell(\beta) \]  the log-partial likelihood function for the proportional hazard model (1) can be given as

\[ \ell(\beta) = \sum_{i=1}^{n} \delta_i \left[ x_i^T \beta - \log \left( \sum_{j=1}^{n} I(z_i \leq z_j) \exp(x_j^T \beta) \right) \right] \]  

(5)

The LASSO estimate (Tibshirani, 1996, 1997) of \( \beta \) is given by

\[ \hat{\beta} = \arg \max_{\beta} \ell(\beta), \text{ subject to } \sum |\beta_j| \leq s \]  

(6)

For nonparametric estimation of \( h(t) \) by (4), we can use Friedman’s general gradient boosting machine (Friedman, 2001), with possible combination of basis expansion (spline) or kernel smoothing in the line search step.

The following sections will talk about the details of LASSO shrinkage and gradient boosting for parametric/nonparametric hazard models.
2.2. Regularized Cox Regressions

2.2.1. Least Angle Regression (LARS) for L1 Regularized Partial Likelihood

Let \( Z \equiv (Z_1, \ldots, Z_n) \) with \( Z_i = x_i^T \beta \), then (5) can be written as

\[
\ell(\beta) = \sum_{i=1}^{n} \delta_i \left[ n_i - \log \left( \sum_{j=1}^{n} I(z_j) \exp(\eta_j) \right) \right]
\]

(7)

Let \( u = (u_1, \ldots, u_n) \) with

\[
u_i = \frac{\partial \ell}{\partial \eta_i} = \delta_i - e^{\eta_i} \sum_{k=1}^{n} \frac{\delta_k I(z_k \leq z_i)}{\sum_{j=1}^{n} I(z_j \leq z_i) \exp(\eta_j)}
\]

and \( A = (a_{ij})_{n \times n} \), where

\[
a_{ii} = -\frac{\partial^2 \ell}{\partial \eta_i^2} = \frac{\partial \ell}{\partial \eta_i} = e^{\eta_i} \sum_{k=1}^{n} \frac{\delta_k I(z_k \leq z_i)}{\sum_{j=1}^{n} I(z_j \leq z_i) \exp(\eta_j)} - e^{2\eta_i} \sum_{k=1}^{n} \frac{\delta_k I(z_k \leq z_i)}{\left( \sum_{j=1}^{n} I(z_j \leq z_i) \exp(\eta_j) \right)^2}
\]

and for \( i \neq j \)

\[
a_{ij} = -\frac{\partial^2 \ell}{\partial \eta_i \partial \eta_j} = e^{\eta_i + \eta_j} \sum_{k=1}^{n} \frac{\delta_k I(z_k \leq z_i) I(z_k \leq z_j)}{\left( \sum_{j=1}^{n} I(z_j \leq z_i) \exp(\eta_i) \right)^2}
\]

The LASSO regularized partial likelihood is maximized by an iterative reweighted least square with L1 constraint procedure:

1. Fix \( s \), and initialize \( \hat{\beta} = 0 \).
2. Compute \( \eta, u, A, \) and \( z = \eta + A^{-1} u \) based on \( \hat{\beta} \), where \( A^{-1} \) is a generalized inverse of \( A \) satisfying \( AA^{-1} = A \).
3. Minimize \( (z - X\beta)^T A (z - X\beta) \), subject to \( \sum |\beta_j| \leq s \).
4. Repeat Steps 2 and 3 until \( \hat{\beta} \) does not change.

In each iteration, we need to solve an L1 regularized weighted least square problem. This task can be solved by the LARS algorithm proposed by Efron, Hastie, Johnstone, and Tibshirani (2004), which takes only a computing time of a least square fit and calculates the full LASSO path. Before applying LARS to the procedure, we should modify Step 3 a little bit so that it can fit into the LARS procedure. Let the SVD decomposition of
\( A = V D V^T \), and let \( T = D^{1/2} V^T, \tilde{z} = T z \), and \( \tilde{X} = T X \) then Step 3 is equivalent to minimizing \( (\tilde{z} - \tilde{X} \hat{\beta})^T (\tilde{z} - \tilde{X} \hat{\beta}) \) subject to \( \sum |\hat{\beta}_i| \leq s \), which can be solved by the LARS algorithm taking \( \tilde{z} \) as the response variable and the columns of \( \tilde{X} \) as predictor variables.

The LARS algorithm for least square regression model works as follows. Consider we are doing a regression, where \( y \) is the response and \( x_1, \ldots, x_p \) are the standardized predictors. The LARS algorithm works as follows:

1. Initialize \( r = y, \hat{\beta}_1 = \hat{\beta}_2 = \cdots = \hat{\beta}_p = 0 \).
2. Find predictor \( x_j \) most correlated with \( r \).
3. Increase \( \beta_j \) in the direction of sign \( \text{corr}(r, x_j) \) until some other competitor \( x_k \) has as much correlation with current residual as does \( x_j \).
4. Move \( (\hat{\beta}_j, \hat{\beta}_k) \) in the joint least squares direction for \( (x_j, x_k) \) until some other competitor \( x_{\ell} \) has much correlation with current residual.
5. Continue until all predictors have been included. Stop when \( \text{corr}(r, x_j) = 0 \) for \( \forall j \), and we get the ordinary least square solution.

It turns out that a slight modification of the above procedure can produce all the LASSO and forward stagewise regression paths.

2.2.2. \( L^d \) Regularization and Extensions

The idea of shrinkage can be extended profitably to other kinds of penalties. In a special tractable case, the \( L^2 \) regularized partial likelihood method replaces Step 3 in the last section by

\[
\text{Minimize } (z - X \beta)^T A (z - X \beta), \text{ subject to } \sum \beta_i^2 \leq s
\]

for some \( s \). This is a well-known weighted ridge regression, and the solution is given by

\[
\hat{\beta} = (X^TAX + \lambda I)^{-1}X^TAz
\]

with \( \lambda \) depending on \( s \). The optimization steps in this case can be seen as iterative reweighted ridge regressions.

If we define a (generally convex) penalty function by \( p(\cdot) \), then the Step 3 can be changed to

\[
\text{Minimize } (z - X \beta)^T A (z - X \beta), \text{ subject to } p(\beta) \leq s
\]
Minimize \((z - X\beta)^T A(z - X\beta) + \lambda \cdot p(\beta)\) (8)

In the one-dimensional case of \(p\), we see that the LASSO (L1 regularization) is a special case when \(p(x) = \sum |x_i|\), and the weighted \(L^d\) regularization corresponds to \(p(x) = \sum \omega_i |x_i|^d\). Furthermore, the definition in the penalized log-partial likelihood function in Huang and Harrington (2002) corresponds to \(p(x) = x^T \sum x\) for some positive definite matrix \(\sum\).

For a two-dimensional penalty example, let \(\lambda = (\lambda_1, \lambda_2)^T, p(x) = (\sum |x_i|, \sum x_i^2)\), then problem (8) defines a naive elastic net (Hui & Hastie, 2005). In this case, it is recommended that we use the elastic net estimate instead of the naive one.

2.2.3. Regularized Cox Regression with Basis Expansion

The concept discussed in the previous sections can also be extended by using other data mining tools. The key lies in the functional expansion of \(\eta(\cdot)\). The replacement of the linear expansion by other basis functions, can lead to other procedures, which can be used for fitting the Cox regressions.

For example, if we expand \(\eta\) by natural cubic splines with \(K\) knots, or write

\[\eta(x) = \sum_{i=1}^{K} \beta_i N_i(x)\]

where for \(0 \leq k \leq K-2\)

\[N_1(x) = 1, N_2(x) = x, N_{K+2}(x) = d_K(x) - d_{K-1}(x)\]

and

\[d_k(x) = \frac{(x - \xi_k)^3}{\xi_k - \xi_k} - \frac{(x - \xi_k)^3}{\xi_k - \xi_k}\]

then Step 3 of the iteration changes to

Minimize \((z - N\beta)^T A(z - N\beta)\), subject to \(p(\beta) \leq s\)

and the same method applies.
For another example, if we expand $\eta$ by Gaussian Kernels with

$$ \eta(x) = \sum_{i=1}^{K} \beta_i K_\eta(x, x_i) $$

where

$$ K_\eta(x_i, x_j) = e^{-z\|x_i - x_j\|^2} \quad (10) $$

for some $z > 0$ then Step 3 of the iteration changes to

$$ \text{Minimize} \ (z - K_\eta \beta)^T A (z - K_\eta \beta), \ \text{subject to} \ p(\beta) \leq s $$

In the case of Kernel expansion, the penalty function $p$ is often set to

$$ p(\beta) = \beta^T K_\eta \beta $$

In the next section, we will introduce some of the details of elastic net.

### 2.2.4. Elastic Net and Flexible Penalty

We are now considering a general setting of regression with elastic net penalty. The procedure can replace Step 3 of the iterative updating algorithm for generalized Cox regression exactly, and it shows favorable properties over LASSO and ridge regression.

Suppose we already standardize the dataset $(x_i, y_i)_{i=1}^{n}$

**Definition 1.** The naive elastic net solution $\beta_{naive}$ solves the following optimization problem

$$ \beta_{naive} = \arg\min_{\beta} L(\lambda_1, \lambda_2, \beta) $$

where

$$ L(\lambda_1, \lambda_2, \beta) = |y - X\beta|^2 + \lambda_1 \sum_{i=1}^{p} |\beta_i| + \lambda_2 \sum_{i=1}^{p} \beta_i^2 $$

Given the dataset $(x_i, y_i)_{i=1}^{n}$ and $\lambda = (\lambda_1, \lambda_2)^T$, $\beta_{naive}$ solves the LASSO problem

$$ \beta_{naive} = \frac{1}{\sqrt{1 + \lambda_2}} \arg\min_{\beta} L(\gamma, \beta) $$
where

\[ L(\gamma, \beta) = |\hat{y} - \hat{X}\beta|^2 + \gamma \sum_{i=1}^{p} |\beta_i| \]

\[ \hat{X} = (1 + \lambda_2)^{-1/2} \left( \frac{X}{\sqrt{\lambda_2 I_p}} \right) \]

\[ \hat{y} = \begin{pmatrix} y \\ 0 \end{pmatrix} \]

\[ \gamma = \frac{\lambda_1}{\sqrt{1 + \lambda_2}} \]

From our former discussion, the entire path of naive elastic net can also be solved by the LARS algorithm with the computing time of least squares.

Empirical evidence shows that the naive elastic net does not perform satisfactorily, unless it is close to LASSO or ridge regression. This introduces our rescaled elastic net estimate:

**Definition 2.** The elastic net solution $\beta^{\text{elastic}}$ is defined by

\[ \beta^{\text{elastic}} = (1 + \lambda_2) \hat{\beta} \]

It is then natural to observe that

**Theorem 1.** Given the dataset $(x_i, y_i)_{i=1}^{n}$ and $\lambda = (\lambda_1, \lambda_2)^T$, the elastic net estimate $\beta^{\text{elastic}}$ is given by

\[ \beta^{\text{elastic}} = \arg \min_{\beta} \beta^T \left( \frac{X^T X + \lambda_2 I_p}{1 + \lambda_2} \right) \beta - 2y^T X \beta + \lambda_1 \sum_{i=1}^{p} |\beta_i| \]

Comparing with the LASSO solution

\[ \beta^{\text{LASSO}} = \arg \min_{\beta} \beta^T (X^T X) \beta - 2y^T X \beta + \lambda_1 \sum_{i=1}^{p} |\beta_i| \]

we can consider the elastic net estimate as a stabilized version of LASSO. The two coincide when $\lambda_2 = 0$. The elastic net procedure is equivalent to shrinking the covariance matrix $\sum = X^T X$ toward the identity matrix $I$ by $(1 - \gamma) \sum + \gamma I$. This observation also opens the possibility of other
covariance matrix shrinkage procedures that will stabilize the estimate of LASSO or ridge regression.

2.2.5. Threshold Gradient Descent Based Forward Stagewise Selection

It has been pointed out by Hastie, Tibshirani, and Friedman (2001) and Efron et al. (2004) that the “incremental” forward stagewise strategy is closely related to the LASSO (LARS) strategy. For example, if the path of the coefficients is monotone, then the two strategies coincide exactly.

The threshold gradient descent method of Friedman and Popescu (2004) uses the above idea and provides a procedure for a more flexible way of variable selection and coefficient estimation. Specifically, under the Cox proportional model (1), the gradient of the negative log-partial likelihood with respect to parameter $\beta$ is given by

$$
g = \frac{-\partial(-\ell)}{\partial \beta} = X \frac{\partial \ell}{\partial \eta} = Xu
$$

where

$$
u_i = \frac{\partial \ell}{\partial \eta_i} = \delta_i - \phi^n \sum_{k=1}^{n} \frac{\delta_k I(z_k \leq z_i)}{\sum_{l=1}^{n} I(z_l \leq z_i) \exp(\eta_i)}
$$

and $\eta_i = x_i^T \beta$. Then if we fix a threshold value $\tau$ and a small $\Delta$ as increment size in the forward stagewise selection algorithm, the threshold gradient decent algorithm works as follows:

1. Initialize $\beta(0) = 0, s = 0$.
2. Calculate $\eta, u$, and $g$ based on current value of $\beta$.
3. Calculate $f_i(s) = I(|g_i(s)| \geq \tau \max_{0 \leq k \leq p} |g_k(s)|)$.
4. Update $\beta(s + \Delta) = \beta(s) + \Delta g(s) \cdot f(s), s = s + \Delta$.
5. Repeat Steps 2–4 for $M$ times for largest enough $M$. The optimal $\beta$ is determined by the optimal tuning parameter $s$, the selection of which will be discussed later.

If we set $\tau = 0$, then the above algorithm produces the standard gradient decent based procedure that encourages equal coefficient changes in the direction of the path. On the other hand, if $\tau = 1$, then the algorithm only changes the coefficient of the predictor variable that has the largest absolute gradient value in the direction of its path in each step. The idea of this threshold gradient decent path finding algorithm is that, by choosing a suitable threshold parameter, we can have the freedom in controlling the greediness of the algorithm in finding the optimal paths of the coefficients.
2.3. Cox Regression with Group Frailty

It is frequently argued that the Cox regression is insufficient in capturing the correlation of defaults (survival) among subjects, since under the assumptions of Cox proportional model, conditional on the covariate values (which can be stochastic themselves), the survival time of subjects are independent. However, this is unrealistic for practical purposes.

In reality, it is reasonable to assume a “group frailty” effect, i.e., a common stochastic factor within groups, which controls the in-group correlation effect. For example, it is reasonable to assume a frailty effect among patients of similar age and a similar likelihood to default among the companies with the similar credit rating or industry. Under such, conditional on the covariate values, the default timing of the subjects is correlated through group frailty (Fan & Li, 2002). To put it formally, let

$$h_{ij}(t) = h_0(t)u_i \exp(x_{ij}' \beta)$$

for the $j$th subject in the $i$th group, $i = 1, \ldots, m, j = 1, \ldots, J_i$, where $u_i$ is the frailty factor within the $i$th group. It is usually assumed that $u_i$ are i.i.d. and with mean 1 so that all the parameters in the model are estimable. Usually, we take a gamma frailty model with parameter $\alpha$ for mathematical tractability, i.e., the density for $Z_i$ is given by

$$g(u) = \frac{\alpha u^{\alpha-1}e^{-2u}}{\Gamma(\alpha)}$$

Then the likelihood condition on the frailty factor $\{u_i; i = 1, \ldots, m\}$ is given by

$$L(\beta|x, \delta, z, u) = \prod_{i=1}^{m} \prod_{j=1}^{J_i} [(h(z_{ij}))^{\delta_{ij}} S(z_{ij})] \prod_{i=1}^{m} g(u_i)$$

$$= \exp \left( \beta^T \sum_{i=1}^{m} \sum_{j=1}^{J_i} \delta_{ij}x_{ij} \right) \prod_{i=1}^{m}$$

$$\times \left[ \left( \prod_{j=1}^{J_i} (h_0(z_{ij}))^{\delta_{ij}} \right) u_i^{A_i} \exp \left( \sum_{j=1}^{J_i} H_0(z_{ij})x_{ij}' \beta \right) u_i \right] g(u_i)$$
where

\[ A_i = \sum_{j=1}^{J_i} \delta_{ij} \]

By integrating with respect to \( u_1, \ldots, u_m \), we get the likelihood function up to a multiplicative constant

\[
\mathcal{L}(\beta|x, \delta, z) \propto \exp\left( \beta^T \sum_{i=1}^{m} \sum_{j=1}^{J_i} \delta_{ij} x_{ij} \right) \prod_{i=1}^{m} \frac{\prod_{j=1}^{J_i} (h_0(z_{ij}))^{\delta_{ij}}}{\left( \sum_{j=1}^{J_i} H_0(z_{ij}) x_{ij}^T \beta + z \right)^{A_i + \alpha}}
\]

Using the previous idea, the parameter estimation and variable selection can be done via the penalized log-likelihood (up to a additive constant):

\[
\ell(\beta|x, \delta, z) = \sum_{i=1}^{m} \left\{ \sum_{j=1}^{J_i} \delta_{ij} \log h_0(z_{ij}) - (A_i + \alpha) \log \left( \sum_{j=1}^{J_i} H_0(z_{ij}) x_{ij}^T \beta + z \right) \right\}
\]

\[
+ \beta^T \sum_{i=1}^{m} \sum_{j=1}^{J_i} \delta_{ij} x_{ij} + \lambda \cdot p(\beta) \tag{11}
\]

for some penalty function \( p \).

Using the least informative step function estimate for \( H_0(\cdot) \)

\[
H_0(z) = \sum_{k=1}^{N} h_k I(z_k \leq z)
\]

where \( \{z_1, \ldots, z_N\} \) are pooled observed failure times, then differentiating (11) with respect to \( h_k \) gives

\[
h_k^{-1} = \sum_{i=1}^{m} \frac{(A_i + \alpha) \sum_{j=1}^{J_i} I(z_i \leq z_{ij}) \exp(x_{ij}^T \beta)}{\sum_{k=1}^{N} h_k \sum_{j=1}^{J_i} I(z_i \leq z_{ij}) \exp(x_{ij}^T \beta) + \alpha} \tag{12}
\]

So our estimation procedure can be done by an iterative procedure:

1. Initialize \( \beta \) and \( \{h_i\} \).
2. Update \( \{h_i\} \) by (12) using current values for \( \beta \) and \( \{h_i\} \).
3. Update \( \beta \) by maximizing (11) for fixed tuning parameter \( \lambda \) using a Newton–Raphson procedure.
4. Repeat Steps 2 and 3 until the algorithm converges.
One can also extend the above algorithm to estimate the parameter $\alpha$. The selection of the tuning parameter $\lambda$ will be discussed later. For frailty models with time-dependent or hidden covariates, one may resort to the expectation maximization (EM) algorithm with a possible combination of stochastic integration algorithms. For example, Gibbs sampling and the acceptance–rejection scheme may be employed to calculate the stochastic paths of the frailty factors.

2.4. Optimal Choice of the Parameter $s$ or $\lambda$

Generally, the optimal tuning parameter $s$ (or $\lambda$) is determined by (bootstrapped) cross validation, as in any model selection scenarios. Specifically, in our special case of Cox regression, we have at least the following two choices:

- Generalized cross validation (GCV) statistics
  
  Since the constraint in Step 3 of our $L^d$ regularization problem can be written as
  
  $$\sum \beta_j^2 |\beta_j|^{d-2} \leq s$$
  
  Hence, if we define
  
  $$W = \begin{cases} \left( \text{diag}\{ |\beta_j| \} \right)^{-1} & \text{for } d = 1 \\ \text{diag}\{ |\beta_j|^{d-2} \} & \text{for } d \geq 2 \end{cases}$$

  The effective number of parameters (degree of freedom) in our regularized fit $\hat{\beta}$ is given by
  
  $$df(s) = \text{tr}(X(X^TAX + \lambda W)^{-1}X^TA)$$

  and the generalized cross validation statistic (Wahba, 1980) is defined by

  $$\text{GCV}(s) = \frac{-\ell(\hat{\beta})}{n(1 - df(s)/n)^2}$$

  The optimal $s^*$ should minimize the criterion GCV($s$).
Cross validated partial likelihood (CVPL)

The cross validated partial likelihood is defined by

\[
CVPL(s) = -\frac{1}{n} \sum_{i=1}^{n} \left( \ell \left( \hat{\beta}_s^{(-i)} \right) - \ell^{(-i)} \left( \hat{\beta}_s^{(-i)} \right) \right)
\]

where \(\hat{\beta}_s^{(-i)}\) is the coefficient estimate without the \(i\)th subject and using tuning parameter \(s\) and \(\ell^{(-i)}\) the log-partial likelihood function without the \(i\)th subject.

Let \(\beta\) be the true value of coefficient if our model is right, then it can be seen that minimizing \(CVPL(s)\) is asymptotically equivalent to minimizing

\[
CVPL(0) + E[(\hat{\beta}_s - \beta)^T A(\hat{\beta}_s - \beta)]
\]

### 2.5. Boosting Generalized Cox Regressions

#### 2.5.1. Friedman's Gradient Boosting Machine

Suppose we are under the general setting of estimating a function \(F(\cdot)\) by the following expansion

\[
F(x) = \sum_{m=0}^{M} \beta_m h(x; a_m)
\]

where the base learner (or weak learner) \(h(x; a_m)\) can be seen as a basis function. It can be in the simple linear form \(h(x; a_m) = x\), or can be generated by more complicated machinery like spline, kernel expansion, neural net, classification and regression trees (CART), multiple adaptive regression splines (MARS), wavelets, support vector machine (SVM), etc.

For finite sample problems encountered in real world cases, we have sample \((x_i, y_i)^n_{i=1}\). Given a loss function \(L\), our goal is to minimize over the functional space the following criterion:

\[
\hat{F}(x) = \min_{F} \sum_{i=1}^{n} L(y_i, F(x_i))
\]
Friedman (2001) proposed the influential idea of a general gradient boosting machine, which trains our estimate of $F$ by the following steps:

1. Initialize $F_0(x) = \arg \min \sum_{i=1}^{n} L(y_i, p)$, or just initialize $F_0(x) = 0$.
2. For $m = 1$ to $M$ do:
   (a) Compute the “pseudo” responses $	ilde{y}_i = -\left[ \frac{\partial L(y_i, F(x_i))}{\partial F(x)} \right]_{F(x)=F_{m-1}(x)}$ for $i = 1, \ldots, N$
   (b) Calculate the least square fit $a_m = \arg \min_{a_m} \sum_{i=1}^{n} [\tilde{y}_i - \beta h(x_i; a)]^2$
   (c) Line search:
      $$\rho_m = \arg \min_{\rho} \sum_{i=1}^{n} L(y_i, F_{m-1}(x) + \rho h(x_i; a_m))$$
   (d) Update $F_m(x) = F_{m-1}(x) + \rho_m h(x; a_m)$
3. Boosted estimate: $\hat{F}(x) = F_M(x) = F_0(x) + \sum_{i=1}^{M} \rho_m h(x; a_m)$.

In particular, if $h(x; a_m)$ is a regression tree, the above steps describe the algorithm to construct a famous class of models: multiple additive regression trees (MART). See Hastie et al. (2001) for more details.

2.5.2. Boosted Cox Regression using Basis Expansion

In the setting of Cox regression by (7), our minimizing criterion (13) changes to

$$\hat{\eta}(x) = \min_{\eta} L(\eta(x))$$

where

$$L(\eta(x)) = \sum_{i=1}^{n} \delta_i \left[ \eta(x_i) - \log \left( \sum_{j=1}^{n} I(z_i \leq z_j) \exp(\eta(x_i)) \right) \right]$$

Then, in Step 2a of the gradient machine (notice that the loss criterion is not additive)

$$\tilde{y}_i = \frac{\partial L(\eta(x))}{\partial \eta(x_i)}$$

$$= \delta_i - \sum_{k=1}^{n} \frac{\delta_k I(z_k \leq z_i) \exp(\eta(x_i))}{\sum_{j=1}^{n} I(z_k \leq z_j) \exp(\eta(x_k))}$$
We have the flexibility in choosing the form of the basis $h(x; a)$ in Step 2b. In the simplest case,

$$h(x; a) = a_0 + \sum_{i=1}^{p} a_i x_i$$

then, Step 2b just fits a simple linear regression. As always, other choices of the form of $h(x; a)$ lead to other gradient boosting algorithm for the Cox regression.

For example, if $h(x; a)$ has the form

$$h(x; a) = \sum_{i=1}^{K} a_i N_i(x)$$

where $N_i$ is the basis for natural cubic spline defined in (9), then Step 2b will fit an unconstrained natural cubic spline with response $\tilde{y}$.

For another example, if $h(x; a)$ has the form

$$h(x; a) = \sum_{i=1}^{K} a_i K_2(x, x_i)$$

where $K_2$ is the Gaussian Kernel defined in (10), then Step 2b amounts to fit a kernel smoothed local regression.

Step 2c amounts to a linear proportional hazard model to the responses $(z_i, \delta_i)_{i=1}^{n}$ with predictor $h(x; a_m)$ offset $\eta_{m-1}(x)$, and regression coefficient $\rho$.

### 2.6. Bagging and Subsample Aggregating

Bagging (or bootstrap aggregating) and subsample aggregating (also called subagging) are model-averaging methods which are designed to stabilize the fitting and predicting results. It turns out that they also increase the accuracy of parameter estimation or prediction as well.

The bootstrap was introduced by Efron (1979) and extended by many others. The original idea of bootstrap is to sample from the data using the same size with replacement and use the bootstrap sample to do estimation and inferences. For example, we can use the bootstrap procedure to estimate the standard deviation of any statistic. This is particularly useful for small datasets. For large dataset, we do not need to reuse the original data so many times and one can use the so-called subsampling procedure, which is
to sample a subset of the data each time without replacement and use the resulting subsamples to do estimation and inferences.

Now, we consider fitting a model to the dataset \( \{x_i, y_i\}_{i=1}^n \), and make prediction \( \hat{f}(x) \) at future inputs \( x \). Bagging makes the prediction by averaging over a collection of bootstrap samples, and subsample aggregating makes the prediction by averaging over a collection of subsamples. Similarly, Bagging (subsample aggregating) estimates the parameter by average over a collection of estimated parameters from bootstrap samples (subsamples). This works for all the data mining procedures mentioned before.

Specifically, for each bootstrap sample \( \{x_{ib}, y_{ib}\}_{i=1}^n \) for \( b = 1, \ldots, B \), we fit our model and give the prediction \( \hat{f}^{*b}(x) \), the bagging estimation is

\[
\hat{f}_{\text{bag}}(x) = \frac{1}{B} \sum_{b=1}^B \hat{f}^{*b}(x)
\]

Similar procedure works for subsample aggregating. Asymptotical results for bagging and subsample aggregating can be found in standard references. For example, see Peter Bühlmann (2003), which also propose a more robust version of bagging.

3. PART II: GENERALIZED COX REGRESSION WITH TIME-DEPENDENT AND HIDDEN COVARIATES

3.1. Time-Varying Covariates

When our dataset has a time series feature, instead of only one survival time or censoring time for each subject, we have observations over time of the survival of a subject. For example, we may have the data consisting the time series of multiple firms with default indicator for each observation. In this case, the classical Cox proportional model (1) should be extended.

Now assume we have \( m \) entities, each with a time series observation. Suppose the observation time for the \( i \)th entity is \([t_i, T_i]\) and in time \( t \) we have the triplet \((x_{it}, \delta_{it}, t) : i = 1, \ldots, m\), where \( \delta_{it} \) is the default (survival) indicator for the \( i \)th entity at time \( t \). Again, we model the hazard rate \( h_i \) for the \( i \)th entity by the proportional model

\[
h_i(t) = h_{0i}(t) \exp(x_{it}^T \beta)
\]
Let us first suppose the covariates are deterministic. The complete likelihood for the data is given by

\[
L(\beta|x, z, \delta) = \prod_{i=1}^{m} \left( e^{-\sum_{t=i}^{T_i} h_i(t) \Delta t} \prod_{t=i}^{T_i} [\delta_{it} h_i(t) + (1 - \delta_{it})] \right)
\]

and the log-likelihood function is given by

\[
\ell(\beta|x, z, \delta) = \sum_{i=1}^{m} \sum_{t=i}^{T_i} (-h_i(t) \Delta t + \log(h_i(t)) \delta_{it})
\]

\[
= \sum_{i=1}^{m} \sum_{t=i}^{T_i} (-h_{0i}(t) \exp(x_i^T \beta) \Delta t + \log(h_{0i}(t) \exp(x_i^T \beta)) \delta_{it})
\]

(14)

where \( \Delta t = t_{i+1} - t_i \) is the tenor for the ith observation.

If we model \( h_{0i}(t) \) by the least informative approach (piecewise constant), i.e., \( h_{0i}(0) = 0 \),

\[ h_{0i}(t) = \lambda_{li} \quad \text{for} \quad z_{l-1} < t \leq z_l \]

or

\[ h_{0i}(t) = \sum_{l=1}^{N} \lambda_{li} I(z_{l-1} < t \leq z_l) \]

(15)

where \( \{z_1, \ldots, z_N\} \) are pooled observed failure times and \( z_0 = 0 \) for positive constants \( \lambda_{li}, l = 1, \ldots, N_i \) and \( i = 1, \ldots, m \). Plugging in this estimate and taking the derivatives of (14) with respect to \( \lambda_{li} \) generates:

\[ \dot{\lambda}_{li} = \frac{\sum_{t=i}^{T_i} I(z_{l-1} < t \leq z_l) \exp(x_i^T \beta) \Delta t}{\sum_{t=i}^{T_i} I(z_{l-1} < t \leq z_l) \exp(x_i^T \beta) \Delta t} \]

So basically the problem can be solved by the following iterative scheme:

1. Initialize the values for \( \beta \).
2. For current values of \( \beta \), calculate the current values for \( \dot{\lambda}_{li} \) for \( l=1, \ldots, N \), and \( i=1, \ldots, m \). Then, we have the current values for \( h_{0i}(t) \) by (15).
3. Plug the current values for $h_0(t)$ in (14) and solve for the following penalized (partial) log-likelihood problem:

$$\min_{\beta} \ell(\beta|x,z,\delta) + \lambda \cdot p(\beta)$$

for some penalty function $p$, using Newton–Raphson’s procedure.

4. Repeat Steps 2 and 3 until the algorithm converges.

Another treatment of $h_0(z)$ is just simply set $h_0(z) = 1$ for all $i$. Then the penalized log-likelihood reduces the following optimization problem,

$$\min_{\beta} \sum_{i=1}^{m} \sum_{t_{i}} (-\exp(x_{it}^T\beta)\Delta t + x_{it}^T\beta \delta_{ht}) + \lambda \cdot p(\beta)$$

which is much simpler to solve.

3.2. Stochastic Covariate Processes

The model in the previous section simply assumes that the covariates are deterministic. However, we may encounter problems in survival (failure, default) analysis where the covariates themselves have the stochastic nature.

For example, to use macroeconomic variables like interest rates, indices, GDP growth, etc. or firm-specific variables like return, debt, asset, etc. to model the defaults or default correlation of multiple firms, we may as well take in consideration that the covariates are also stochastically varying over time. For another, to predict the failure time in a medical experiment, the temperature as an external factor and the level of a certain chemical in the human body as an internal factor may both have some stochastic nature that is not captured by deterministic processes.

Now we suppose that the covariate $X$ has a parametric form of stochastic process and its likelihood is given by $\mathcal{L}(\gamma|X)$. Under the doubly stochastic assumption (Duffie, Saita, & Wang, 2007), conditional on the paths of the covariate process, the default (survival) timing of the entities are independent, and the full likelihood function is given by

$$\mathcal{L}(\beta, \gamma|x,z,\delta) = \mathcal{L}(\gamma|x)\mathcal{L}(\beta|x,z,\delta)$$

So, the maximum likelihood estimator $(\hat{\beta}, \hat{\gamma})$ can be obtained by maximizing $\mathcal{L}(\gamma|x)$ and $\mathcal{L}(\beta|x,z,\delta)$ respectively.
For mathematical convenience, we usually model the covariate process by simple time series models, for instance, autoregressive integrated moving average (ARIMA) or its vector version (VARIMA). Here, I briefly review the vector autoregressive model with possible cointegration effect.

Suppose \( x_t \) is a \( d \times 1 \) vector for each \( t \). A multivariate extension of AR\( (p) \) model for \( x \) (with 0 mean) is given by

\[
x_t = A_1 x_{t-1} + A_2 x_{t-2} + \ldots + A_p x_{t-p} + \epsilon_t
\]

We say that \( x \sim \text{VAR}(p) \).

If the components of \( x \) are \( I(1) \) and cointegrated, then it has the Granger’s representation (VAR with cointegration)

\[
\Delta x_t = \alpha \beta^T x_{t-1} + B_1 \Delta x_{t-1} + \ldots + B_{p-1} \Delta x_{t-p+1} + \epsilon_t
\]

where \( \beta \) is the cointegration vector for \( x \). The system is also called error correction form.

Engle and Granger (1987) suggest the two-stage estimation method for

\[
\Delta x_t = \alpha \beta^T x_{t-1} + B_1 \Delta x_{t-1} + \ldots + B_{p-1} \Delta x_{t-p+1} + \epsilon_t
\]

(1) Estimate \( \beta \) by least square regression.
(2) Estimate \( \alpha, B_j, j = 1, \ldots, p-1 \) by maximum likelihood.

It is proved that if we estimate the parameters in this way, then

(1) \( \hat{\beta} \) is super-consistent: it converges to \( \beta \) at the rapid rate \( T^{-1} \), where \( T \) is the sample size.
(2) \( \hat{\alpha}, \hat{\beta}, j = 1, \ldots, p-1 \) are consistent and asymptotically normal.

Another tractable type of model is affine process, which is discussed in Section 4 in Zhen (2006), and the estimation procedures are discussed in Section 6.

### 3.3. Frailty Factor for Modeling Dependence

The assumption that conditional on the paths of the covariate processes, the default (survival) timing of the entities is independent may be violated in reality. For example, previous models generally cannot capture the “contagion effect” of the defaults of multiple firms, which is essential to understand the risks in the credit market. One remedy for this is to assume some subject-specific or time-dependent frailty factors that may have the effect of adding an additional source of certainty in the default and default correlation among subjects.
Consider the following model

\[ h_t(t) = h_0(t)Y_tS_t \exp(x_t^T \beta) \]  

(16)

where \( Y_t \) is a time-dependent frailty factor and \( S_t \) a subject-specific frailty factor. Under the gamma model, suppose \( S_t \) is i.i.d. with density

\[ g(s) = \frac{z^2s^{z-1}\exp(-zs)}{\Gamma(z)} \]  

(17)

and \( Y_t \) is a positive process. For example, \( Y_t \) can be a gamma process, geometric Brownian motion, or exponential of a variance gamma process. One can refer to Appendix B for a review of gamma and variance gamma processes, see also Madan and Seneta (1990).

A popular scheme of calibrating frailty related models is through Markov chain Monte Carlo expectation maximization (MCMC EM) algorithm, which provides a general scheme for optimization involving stochastic (or hidden) factors. The following subsections give a procedure for the MCMC EM optimization problem and we will narrow our focus on our Cox regression settings.

3.3.1. The Expectation Maximization Algorithm

The expectation maximization (EM or Baum–Welch) algorithm is originally proposed for maximizing likelihoods in cases with latent (unobserved or missing) data. The latent data (frailty factor in the setting of Cox regression) can be introduced by model construction or by data augmentation in order to simplify the maximizing problem. Naturally, the EM algorithm can be used for minimizing any loss criterion with latent factor.

Suppose we are going to maximize a general function \( \ell(\theta; X) \) of parameter \( \theta \) and observed data \( X \). For example, \( l \) can be the (penalized) log-likelihood function or the negative value of any loss function. The augmented function \( \ell_0(\theta; X, Z) \) also depends on the latent or missing data \( Z \) and usually we have \( \ell_0(\theta; X) = E(\ell_0(\theta; X, Z)|X, \theta) \). However, this relationship need not hold in general. The EM algorithm works as follows:

1. Initialize our guess for the parameters \( \theta^{(0)} \).
2. Expectation step: At the \( j \)th step, compute the expectation

\[ Q(\theta, \theta^{(j)}) = E(\ell_0(\theta; X, Z)|X, \theta^{(j)}) \]

as a function of \( \theta \).
3. Maximization step: Solve the maximization problem:

$$\theta^{(j+1)} = \arg \max_{\theta} Q(\theta, \theta^{(j)})$$  \hspace{1cm} (18)

4. Repeat Steps 2 and 3 until the algorithm converges.

At least two complications arise in the above EM algorithm. The first is that it is hard to compute the expectation in Step 2 in general because the distribution of $Z$ condition on $X$ may not have explicit solution. Even the conditional density can be computed explicitly, the expectation involves (high-dimensional) numerical integration, which is usually not stable and reliable. One remedy for this is to use the Markov chain Monte Carlo methods, specifically, the Gibbs sampler and the Metropolis (Metropolis–Hasting) algorithm for sampling the posterior distributions and use the sample average

$$\frac{1}{M} \sum_{n=N+1}^{N+M} \ell_0(\theta; X, Z^{(n)})$$

to compute the expectation, where $N$ is the “burn-in” period in our sample generating procedure. We will introduce MCMC shortly.

The second problem is the maximization in Step 3. The Newton–Raphson algorithm does not necessarily lead to nice solutions. Other optimization schemes like simulated annealing, the genetic algorithm may be employed to find the global maximum too. For computational reasons, it is advisable just to apply one-step Gradient Descent to Step 3 and hope the iterative procedure of Step 2 and Step 3 will lead to a suboptimal solution. Our focus then will be the Markov chain Monte Carlo calibration of the expectation in Step 2.

3.3.2. The Markov Chain Monte Carlo Methods

A particularly useful and the most simple Markov chain simulation algorithm is called the Gibbs sampler or the alternating conditional sampling. Following the notions of previous section, suppose we want to sample the distribution of $Z = (Z_1, \ldots, Z_k)$ condition on $X$, then the Gibbs sampler works as follows:

1. Initialize $Z_k^{(0)}$, $k = 1, \ldots, K$.
2. At the $t$th step, sample $Z_k^{(t)}$ from the conditional distribution

$$Z_k^{(t)} | Z_1^{(t)}, \ldots, Z_{k-1}^{(t)}, Z_{k+1}^{(t-1)}, \ldots, Z_K^{(t-1)}, X$$

for $k = 1, \ldots, K$. 

Data Mining Procedures in Generalized Cox Regressions
3. Repeat Step 2, until the joint conditional distribution of $Z^{(t)}|X = Z_1^{(t)}, \ldots, Z_k^{(t)}|X$ does not change or simply repeat a designated number of steps.

The Gibbs sampler works pretty well if we can easily generate the conditional sample in Step 2. However, it may not be easy to generate such a sample directly, even if we have the explicit formula for the density. A more flexible sampling method is called the Metropolis algorithm, which can be seen as an adaptation of a random walk that uses an acceptance/rejection rule to converge to the target distribution. It only requires we know the likelihood ratio of the target density.

Suppose we want to sample from a (conditional) density function $L(\theta; Z, X)$ for $Z$. The Metropolis algorithm works as follows:

1. Initialize $Z_k^{(0)}, k = 1, \ldots, K$.
2. At the $t$th step, sample a proposal $Z^*$ from a jumping distribution (or proposal distribution) $J_t(Z^*|Z^{(t-1)}, X)$. The proposal distribution $J_t$ should be symmetric, or $J_t(Z_1|Z_2, X) = J_t(Z_2|Z_1, X)$.
3. Calculate the likelihood ratio

$$r = \frac{L(\theta; Z^*, X)}{L(\theta; Z^{(t-1)}, X)}$$

where $L$ is the (conditional) density function for $Z$.
4. Draw $U$ uniformly from $[0, 1]$, and set

$$Z^{(t)} = \begin{cases} Z^* & \text{if } U < \min(r, 1) \\ Z^{(t-1)} & \text{otherwise} \end{cases}$$

5. Repeat Steps 2–4, until the conditional distribution $Z^{(t)}|X = Z_1^{(t)}, \ldots, Z_k^{(t)}|X$ does not change or simply repeat a designated number of steps.

Usually, we set $J_t(Z^*|Z^{(t-1)}, X) = N(Z^{(t-1)}, \sigma I_K)$, where $I_K$ is a $K$-dimensional identity matrix. The Metropolis–Hastings algorithm slightly generalizes the above algorithm to asymmetric proposal distribution functions. If $J_t(Z_1|Z_2, X) = J_t(Z_2|Z_1, X)$ does not hold for any $Z_1, Z_2$, then in Step 3, the formula for $r$ changes to

$$r = \frac{L(\theta; Z^*, X)/J_t(Z^*|Z^{(t-1)}, X)}{L(\theta; Z^{(t-1)}, X)/J_t(Z^{(t-1)}|Z^*, X)}$$
It can be proved that in this way, the sampling distribution will converge to the stationary distribution of a Markov chain, with the stationary distribution the same as our target distribution.

The Steps 2–4 of above algorithm can also consist of a series of $K$ iterations, where we perform acceptance–rejection updating procedure for each of the $Z_j^{(t)}$ for $j=1, \ldots, K$ in each step. If we define the proposal distribution by

$$J_{j,t}^{\text{Gibbs}}(Z^*|Z^{(t-1)}, X) = \begin{cases} \mathcal{L}(\theta; Z^*_j, Z^{(t-1)}_{-j}, X) & \text{if } Z^*_j = Z^{(t-1)}_{-j} \\ 0 & \text{otherwise} \end{cases}$$

Then, it can be seen that the Metropolis–Hastings algorithm produces a Gibbs sampler.

### 3.3.3. Calibrate the Frailty Model

Our basic scheme is to use the MCMC version of the EM algorithm to estimate the model parameters (and possibly to select the covariates by regularization). As mentioned before, we are most interested in Step 2 of the EM algorithm. Since we can separate the estimation of the covariate processes and the hazard rate model, our likelihood function will only consist of the parameters in the hazard formulation. Under the frailty model (16), the augmented likelihood function is given by

$$\mathcal{L}(\beta, \theta; x, Y_t, S, \delta) = \prod_{i=t}^{m} \left( e^{-\sum_{t_{i-1}}^{t_i} h_i(t) \Delta t} \prod_{t_{i-1}}^{T} [\delta_i h_i(t) + (1 - \delta_i)] \right)$$

where

$$h_i(t) = Y_t S_i \exp(x_i^T \beta)$$

and $\theta$ is the parameter for the frailty factors $Y_t$, $S_t$, where $S_t$ has the gamma model (17). We set $h_0(z) = 1$ for simplicity. Under a Markov model for $Y_t = g(W_t)$ (geometrical Brownian motion, gamma process, exponential of a variance gamma process) where $W$ is a Markov and $g$ a deterministic function. Then the conditional density of $W_t$ given $W_{(-t)} = (W_1, \ldots, W_{t-1}, W_{t+1}, \ldots, W_T)$ is

$$f(W_t|\beta, \theta, x, W_{(-t)}, S, \delta) \propto \mathcal{L}(\beta, \theta; x, Y_t, S, \delta) f(W_t|W_{t-1}, \theta) f(W_t|W_{t+1}, \theta)$$
The sampling procedure for fixed parameters ($\beta, \theta$) works as follows:

1. Initialize $Y_t = 1$ and $S_t = 1, 0 \leq t \leq T$ and $1 \leq i \leq m$.
2. Given the current values of $Y_t, S_t$ in step $j$, use the Metropolis–Hastings algorithm to draw $Y_{t,j+1}$ by the following scheme, where we can use a Gaussian proposal distribution in all cases:

   (a) If $Y_t = e^{b W_t}$ is a geometric Brownian motion, where $W_t$ is a standard Brownian motion. Then, the (conditional) likelihood ratio in Step 3 of the Metropolis algorithm is given by

   $$r = \frac{f(W_t^* | \beta, \theta, x, W_{t-1}, S, \delta)}{f(W_t | \beta, \theta, x, W_{t-1}, S, \delta)} = \frac{(W_t^*-W_{t-1})/\sqrt{2 \pi}}{g_{1/2}(1/2 \beta)} \cdot \frac{f(W_t^* | W_{t-1}, \theta) f(W_t^* | W_{t+1}, \theta)}{f(W_t | W_{t-1}, \theta) f(W_t | W_{t+1}, \theta)} \quad (19)$$

   where $Y_t^* = e^{b W_t}$, and

   $$f(W_t | W_{t-1}, \theta) = \frac{1}{\sqrt{2 \pi}} e^{-(W_t-W_{t-1})^2/2}$$

   $$f(W_t | W_{t+1}, \theta) = \frac{1}{\sqrt{2 \pi}} e^{-(W_t-W_{t+1})^2/2}$$

   (b) If $Y_t$ is a gamma process $\{G(t, u, v) : t \geq 0\}$, then $Y_t = W_t$,

   $$f(W_t | W_{t-1}, \theta) = g_{u^2/v, y/v} (W_t - W_{t-1})$$

   $$f(W_t | W_{t+1}, \theta) = g_{u^2/v, y/v} (W_{t+1} - W_t)$$

   where $g_{\alpha, \beta}(x)$ is a gamma density

   $$g_{\alpha, \beta}(x) = \frac{1}{\beta^2 \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}$$

   and the (conditional) likelihood ratio in Step 3 of the Metropolis algorithm is also given in (19).

   (c) If $Y_t$ is an exponential of a variance gamma process, then $Y_t = e^{b W_t}$, where $W_t$ is a variance gamma process $\{\text{VG}(t, \theta, \sigma, v) : t \geq 0\}$. We have

   $$f(W_t | W_{t-1}, \theta) = f_{\text{VG}(1, \theta, \sigma, v)} (W_t - W_{t-1})$$

   $$f(W_t | W_{t+1}, \theta) = f_{\text{VG}(1, \theta, \sigma, v)} (W_{t+1} - W_t)$$
where \( f_{VG(t,\theta,\sigma,v)} \) is the density function of a variance gamma process with parameters \((\theta, \sigma, v)\) at time \( t \). The formula for \( f_{VG(t,\theta,\sigma,v)} \) is given in Appendix B, and since it is in the form of an integral, numerical approximation procedures should be employed. The (conditional) likelihood ratio in Step 3 of the Metropolis algorithm is also given in (19).

3. Given \( Y^{(j+1)}_i \) and \( S^{(j)}_i \), we proceed to draw the conditional distributions of \( S^{(j+1)}_i \) for \( i = 1, \ldots, m \).

Under the i.i.d. gamma model, since \( S^{(j+1)}_i \) are conditional independent, and each with density (assume \( \Delta t = 1 \))

\[
\begin{align*}
&f(s| Y^{(j+1)}, X, \delta, \beta, \theta) \propto s^{a-1} \exp(-as) \exp(\sum_{t=t_i}^{T_i} Y^{(j+1)}_t (x^T_i \beta) s) \prod_{t=t_i}^{T_i} \left[ \delta_{it} Y^{(j+1)}_t (x^T_i \beta) s + (1 - \delta_{it}) \right] \\
&\text{so, we can draw } S^{(j+1)}_i \text{ from the gamma distribution } \Gamma(A_i, B_i) \text{ where}
\end{align*}
\]

\[
\begin{align*}
A_i &= \alpha + \sum_{t=t_i}^{T_i} Y^{(j+1)}_t \exp(x^T_i \beta) \\
B_i &= \alpha + \sum_{t=t_i}^{T_i} I(\delta_{it} = 1)
\end{align*}
\]

Particularly, if we ignore recovery from default, and consider entity \( i \) defaults only when \( \delta_{iT_i} = 1 \), then

\[
B_i = \alpha + \delta_{iT_i}
\]

Our maximization step (18) may involve drawing the conditional samples for each set of parameter \( \theta \) for a discretized grid in the parameter space. For computational considerations, we can make very raw grid of the parameters in the initial steps and refine the grid in later steps for better results.

An extension of the model (16) is given by

\[
\begin{align*}
h_{ij}(t) &= h_{0ij}(t) Y_i S_i O_j \exp(x^T_{ij} \beta) \\
\text{where } Y_i \text{ is a time-dependent frailty factor, } S_i \text{ a subject-specific frailty factor, } \\
\text{and } O_j \text{ a group-specific factor for } i = 1, \ldots, m, j = 1, \ldots, J_i \text{ and } t = t_i, \ldots, T_i.
\end{align*}
\]

For example, it is reasonable to assume a group frailty factor among
companies in the same field, e.g., information technology, energy, material, retail, etc. Calibration for the model (20) is straightforward by our proposed MCMC EM method, where Metropolis–Hastings algorithm is used to draw samples from conditional distribution for $Y_t$. Under gamma models for $S_t$ and $O_j$, their conditional distributions in the iterative steps are gamma distributions, making the sampling procedure easy to implement.

4. CONCLUDING REMARKS

This paper introduces various statistical data mining procedures within the context of generalized Cox regressions. It is noteworthy that many iterative procedures described above can be embedded into others so that the practitioners really have a diversified pool of tools to build statistical models for survival probabilities and correlations. On the other hand, the practitioners should also be cautious when employing these data mining procedures. For example, it really makes no sense to bag a boosting procedure or vice versa in the hope of getting better results. Although both of them have the advantage of reducing prediction variance, the combination of the two will not procedure more accurate results. On the contrary the computation is too prohibitive (especially for large datasets) to have any practical value.

Having constructed a parametric model for the hazard rate for the $i$th subject, for example, in the time-dependent covariate case:

$$h_i(t) = \hat{h}_0(t) \exp(x^T_{1i} \hat{\beta}_1 + x^T_{2i} \hat{\beta}_2)$$

where $x_1$ is the systematic covariate and $x_2$ the idiosyncratic covariate, then the joint survival function of the ($i, j$)th subjects are specified by

$$P(\tau_i > T_1, \tau_j > T_2|\mathcal{F}_t) = E(P(\tau_i > T_1, \tau_j > T_2|\mathcal{F}_t, x_{1i}, x_{2i}, x_{1j}, x_{2j})|\mathcal{F}_t)$$

$$= E_i \left( e^{-\int^T_{1i} h_i(s) ds} - \int^T_{1i} h_i(s) ds \right)$$

which can be calculated by Markov chain Monte Carlo methods. The information filtration $\{\mathcal{F}_t : t \geq 0\}$ often rejects the econometrician or market investor’s accumulation of knowledge in the setting of financial modeling. We ignore its construction just for the simplicity of illustration. The formula (21) also holds for frailty models, except that the conditional expectation
will be calculated with respect to both the covariate processes and the frailty factors.

Other applications of these procedures can be found in Computational Biology, Genetics, Bioinformatics, Clinical trial, Actuarial Science, Education and Product Reliability tests.

NOTES

1. \( l(1) \) refers to nonstationary series with stationary first difference.
2. The superscript \((j+1)\) means we are in the \((j+1)\)th step of our iteration procedure. We ignore this notation below when there is no confusion.
3. For a formal definition of independent censoring, refer to Andersen, Borgan, Gill, and Keiding (1993).

ACKNOWLEDGMENTS

The author is grateful for the comments from Prof. T. L. Lai, Department of Statistics, Stanford University and the editors of Advances in Econometrics. The author is thankful to Prof. Thomas Fomby in Southern Methodist University for his insightful suggestions and persistent encouragement.

REFERENCES


### APPENDIX A. COUNTING AND INTENSITY PROCESSES

Suppose a probability space \((\Omega, \mathcal{F}, P)\) with information filtration \((\mathcal{F}_t)_{t \geq 0}\) satisfying the usual conditions:

- **(increasing)** \(\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}\) for all \(s < t\)
- **(right continuous)** \(\mathcal{F}_s = \bigcap_{r > s} \mathcal{F}_t\) for all \(s\)
- **(complete)** \(A \subset B \in \mathcal{F}, P(B) = 0\) implies \(A \in \mathcal{F}_0\)
Definition A.1 Counting process. A $k$-dimensional counting process $N = (N_1, \ldots, N_k)$ is a vector of $k$ càdlàg processes, all zero at time zero, with piecewise constant and nondecreasing sample paths, having jumps of size 1 only and no two components jumping at the same time.

For a classical survival analysis, we have a sample of $n$ observations where $z_i = \min\{T_i, C_i\}$ the minimum of the survival time $T_i$ and censoring time $C_i$ for the $i$th observation. $\delta_i = I_{\{T_i \leq C_i\}}$ is the corresponding censoring indicator.

For a complete description of the picture, we need to further define:

- $R(t) = \{i : Z_i \geq t\}$ as the risk set just before time $t$.
- $Y(t) = \#R(t)$ as the number at risk just before time $t$.
- $S(t) = P(T_i > t)$ as the survival function for the default times.

Under the differentiability assumption of $S(t)$, we have the density function of the survival time as $f(t) = \frac{1}{S(t)}$ and the hazard rate function $h(t) = \frac{f(t)}{S(t)}$. It can be verified that

$$S(t) = \exp\left(-\int_0^t h(s)ds\right) = \exp(-H(t))$$

(22)

where $H(t) = \int_0^t h(s) ds$ is the cumulative hazard function.

Under the assumption of independent censoring, which means roughly that the survival experience at any time $t$ is independent of $\mathcal{F}_{t-}$, we have

$$P(Z_i \in [t, t+dt)\delta_i = 1|\mathcal{F}_{t-}) = 1_{\{Z_i \geq t\}}h(t) dt$$

(23)

Summing the above formula over $i$, we get

$$E(\#\{i : Z_i \in [t, t+dt), \delta_i = 1\}|\mathcal{F}_{t-}) = Y(t)h(t)dt$$

(24)

On the other hand, if we define a process $N = (N(t))_{t \geq 0}$ counting the observed failures, where

$$N(t) = \#\{i : Z_i \leq t, \delta_i = 1\}$$

If we denote

$$dN(t) = N((t+dt)-) - N(t-)$$
as the increment of observed failure over the small interval \([t, t + dt]\), then we can rewrite (24) as

\[
E(dN(t)|\mathcal{F}_{t-}) = \lambda(t) \, dt
\]  
(25)

where \(\lambda(t) = Y(t)h(t)\) is called the intensity process of the survival times.

**Proposition A.1.** Under above assumptions, \(M(t) = N(t) - \Lambda(t)\) is a martingale \((t \geq 0)\), where

\[
\Lambda(t) = \int_0^t \lambda(s) \, ds
\]

is called the cumulative intensity process for the default times.

Since by definition \(dN(t)\) can only be 0 or 1, (25) is equivalent to

\[
P(dN(t) = 1|\mathcal{F}_{t-}) = \dot{\lambda}(t) \, dt
\]

Hence, if we consider \(N(t)\) to the total number of observed defaults in a pool of names, then informally

\[
P(\text{no default in } [t, s] | \mathcal{F}_{t-}) = \prod_{s \leq u \leq t} P(dN(u) = 0|\mathcal{F}_{u-})
\]

\[
= \prod_{s \leq u \leq t} (1 - \dot{\lambda}(u) \, du)
\]

\[
= e^{-\int_t^s \dot{\lambda}(u) \, du}
\]

In the case of constant hazard rate \(h(t) \equiv h\), the intensity process \(\lambda\) is a piecewise constant and decreasing process.

Under the settings of credit risk, we can think that the censoring is caused due to other types of exit except default of the entity of interest. So we can define the survival function for censoring as

\[
S^C(t) = P(C_i > t)
\]

and the corresponding hazard rate function as \(\beta(t)\). Then, similar to the previous arguments, we have

\[
E(\#i : Z_i \in [t, t + dt), \delta_i = 0)|\mathcal{F}_{t-}) = \omega(t) \, dt
\]  
(26)
where \( \omega(t) = Y(t)\beta(t) \) is the intensity process of the censoring times and

\[
\Omega(t) = \int_0^t \omega(s) \, ds
\]

is the cumulative intensity process for the censoring times.

**APPENDIX B. GAMMA AND VARIANCE GAMMA PROCESSES**

Before we go into the definition and properties of the gamma and variance gamma processes, we first introduce a much more larger class of processes: Lévy process (Bertoin, 1996) and some familiar examples.

**Definition B.1 Lévy process.** A Lévy process is any continuous-time stochastic process that starts at 0, admits càdlàg modification, and has “stationary independent increments.”

**Example B.1.** If \( X_t \) is a Lévy process and the increment \( X_t - X_s \) has a Gaussian distribution with mean 0 and variance \( (t-s) \) for \( t \geq s \), then \( X_t \) is called a standard Brownian motion or Wiener process, often denoted by \( X_t = B_t \) or \( W_t \).

**Example B.2.** If \( X_t \) is a Lévy process and the increment \( X_t - X_s \) has a Poisson distribution with parameter \( \lambda \) for \( t \geq s \), or

\[
P(X_t - X_s = k) = \frac{e^{-\lambda(t-s)}(\lambda(t-s))^k}{k!} \quad \text{for } k = 0, 1, \ldots
\]

then \( X_t \) is called a Poisson process with intensity parameter (or rate) \( \lambda \).

The independent increment property of Lévy processes can account for the following.

**Corollary B.1.** If \( X_t \) is a Lévy process and \( \phi(\theta) = E(e^{i\theta X_1}) \) the characteristic function of \( X_1 \), then the characteristic function of \( X_{t+s} - X_s \) is \( (\phi(\theta))^t \) for \( t, s \geq 0 \). Particularly, if \( s = 0 \), then \( X_t \) has characteristic function \( (\phi(\theta))^t \).

**Theorem B.1.** Lévy–Khintchine representation: If \( X_t \) is a Lévy process, then its characteristic function satisfies the Lévy–Khintchine
representation:

\[ E(e^{i\theta X_t}) = \exp \left( i\gamma \theta t - \frac{1}{2} \sigma^2 \theta^2 t + t \int_{\mathbb{R}\setminus \{0\}} (e^{i\theta x} - 1 - i\theta x 1_{|x| < 1}) \nu(dx) \right) \]

where \( \gamma \in \mathbb{R}, \sigma \geq 0 \) and \( \nu \) is a measure defined on \( \mathbb{R}\setminus \{0\} \) called the Lévy measure satisfying

\[ \int_{\mathbb{R}\setminus \{0\}} (x^2 \wedge 1) \nu(dx) < \infty \]

Thus, a Lévy process can be seen as comprising of three components: a drift, a Brownian motion, and a jump component.

From above corollary and theorem, we see that the Lévy–Khintchine representation is equivalent to

\[ \psi(\theta) = \log(\phi(\theta)) = i\gamma \theta - \frac{1}{2} \sigma^2 \theta^2 + \int_{\mathbb{R}\setminus \{0\}} (e^{i\theta x - 1 - i\theta x 1_{|x| < 1}}) \nu(dx) \]

and \( \psi(\theta) \) is called the characteristic component of \( X_t \).

Before I introduce the gamma process, we first give a definition to the compound Poisson process:

**Definition B.2 Compound Poisson process.** A compound Poisson process \( Y_t \) with rate \( \lambda > 0 \) and jump size distribution \( G \) is a continuous stochastic process given by

\[ Y_t = \sum_{i=1}^{N(t)} D_i \]

where \( N(t) \) is a Poisson process with rate \( \lambda \), and \( D_i \) is i.i.d. with distribution \( G \), which is also independent of \( N(t) \).

It is easy to show that a Compound Poisson process is also a Lévy process by direct calculating its characteristic function. It turns out that we can construct a series of compound Poisson processes converging to a limit which is also a Lévy process. The limit process has independent increments with gamma distribution, thus has the name gamma process. See, for example, David and Howard (1993) for the construction.
Let $\Gamma(\alpha, \beta)$ denote the gamma distribution with density

$$f(x) = \frac{1}{B(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta}, \quad x > 0$$

**Definition B.3.** A Lévy process $X_t$ is a gamma process with mean parameter $u > 0$ and variance parameter $v > 0$, if the increment $X_t - X_s$ has the gamma distribution $\Gamma(u^2(t-s)/v, v/u)$ for $t > s$.

It can be shown by direct calculations that the characteristic function of a gamma process $X_t$ is given by

$$\phi(\theta) = E(e^{i\theta X_t}) = \left(\frac{1}{1 - i\theta(v/u)}\right)^{(u^2\theta)/v}$$

and the Lévy measure for $X_t$ is given by

$$v(dx) = \frac{u^2 \exp(-u/v)x}{v x} dx$$

The gamma process is always non-negative and increasing, which may restrict the practical applications of such processes. It turns out that by taking the difference of two independent gamma processes with some specific parameters, we will get another Lévy process, which behaves somewhat like Brownian motion, but has more preferable properties over Brownian motion. Interestingly, this kind of process also has a construction closely related to Brownian motion:

**Definition B.4 Variance gamma process or VG process.** A VG process is obtained by evaluating Brownian motion with drift at a random time given by a gamma process. Specifically, let $b(t; \theta, \sigma) = \theta + \sigma W_t$, where $W_t$ is a standard Brownian motion. The VG process $X_t$ with parameter $(\theta, \sigma, v)$ is given by $X_t = b(Y_t; \theta, \sigma)$ where $Y_t$ is a gamma process with mean parameter 1 and variance parameter $v$. The process can also be seen as being generated by the independent increments:

$$X_{t+s} - X_t = \theta(Y_{t+s} - Y_t) + \sigma(W(Y_{t+s}) - W(Y_t))$$

By Markov property, we see $X_{t+s} - X_t \sim b(Y_s, \theta, \sigma)$ for $t, s \geq 0$. 
If $X_t$ is a variance gamma process with parameter $(\theta, \sigma, \nu)$, then $X_t$ has density function

$$f(x) = \int_0^\infty \frac{1}{\sigma \sqrt{2\pi s}} e^{-((x-\theta)^2/2\sigma^2)} \frac{s^{t/v-1}e^{-s/v}}{v^{t/v}\Gamma(t/v)} ds$$

and calculation shows it has the characteristic function

$$\phi(\theta) = E(e^{i\theta X_t}) = \left( \frac{1}{1 - i\theta \nu u + (\sigma^2 v/2)\theta^2} \right)^{t/v}$$

The variance gamma process $X_t$ with parameter $(\theta, \sigma, \nu)$ can be expressed as the difference of two independent gamma processes (Madan, Carr, & Chang, 1998): $X_t = Y_{1t} - Y_{2t}$ where $Y_{1t}$, $Y_{2t}$ are gamma processes with parameter $(u_1, v_1)$ and $(u_2, v_2)$ respectively and

$$u_1 = \frac{1}{2} \sqrt{\theta^2 + 2\sigma^2 \nu} + \frac{\theta}{2}$$

$$u_1 = \frac{1}{2} \sqrt{\theta^2 + 2\sigma^2 \nu} - \frac{\theta}{2}$$

$$v_1 = u_1^2 \nu$$

$$v_2 = u_2^2 \nu v N.$$ 

It can also be shown that the Lévy measure for $X_t$ is given by

$$v(dx) = \begin{cases} 
\frac{u_1^2}{v_1 x} e^{-u_1 x} \, dx & \text{for } x > 0 \\
-\frac{u_2^2}{v_2 x} e^{-u_2 x} \, dx & \text{for } x > 0 \\
\frac{e^{\theta x/\sigma^2} - \sqrt{2/\nu} e^{\sigma^2/2}}{v|x|} \, dx & \text{for } x > 0
\end{cases}$$

which is symmetric if only $\theta = 0$. 

---

ZHEN WEI

194
The credit migration process contains important information about the dynamics of a firm’s credit quality, therefore, it has a significant impact on its relevant credit derivatives. We present a jump diffusion approach to model the credit rating transitions which leads to a partial integro-differential equation (PIDE) formulation, with defaults and rating changes characterized by barrier crossings. Efficient and reliable numerical solutions are developed for the variable coefficient equation that result in good agreement with historical and market data, across all credit ratings. A simple adjustment in the credit index drift converts the model to be used in the risk-neutral setting, which makes it a valuable tool in credit derivative pricing.
1. INTRODUCTION

Credit derivatives have become one of the major sectors in financial trading and investment, and the credit rating of a firm by one of the major rating agencies, such as Moody’s or Standard and Poor’s, can be the most crucial indicator for the credit quality of that firm. Modeling credit rating transitions naturally is recognized as a major topic in credit risk research. The Markov chain approach is an obvious choice, and the early work by Jarrow, Lando, and Turnbull (1997) highlighted and clarified many crucial issues in this area. However, the underlying process still remains open to different interpretations. There is a strong incentive to link the process to the balance sheets of the firm, which suggests that structural models, originally pioneered by Merton (1974), are natural to serve as strong candidates for modeling the default and rating transitions. The appealing aspects of structural models, such as the first passage of the default boundary, and the crossing of a barrier at any time, make these approaches quite promising.

It has been widely recognized that Brownian motion alone is inadequate to explain all the rating transitions, for the obvious reason that transitions across several ratings simultaneously occur at non-negligible rates, but they cannot be modeled by a continuous process. The immediate remedy is to combine a jump process with the conventional Wiener process. The early work of Zhou (2001) in defaultable bonds suggests that jumps may hold the key to these credit modeling issues. In contrast, the stochastic volatility approach by Fouque, Sircar, and Sølna (2006) provides another way to avoid some of the handicaps of the Brownian diffusion approach. However, the complication of estimating hidden volatility information in a stochastic volatility model brings in yet more unknown factors, especially in the rating transition problem where the transition process is quite different from the stock or the interest rate processes by nature.

A great deal of effort has been focused on the Lévy process recently, after the jump process was first introduced by Merton (1976), simply because of the jumps allowed in the process, and other desirable properties, such as the Markov property and the resulting partial integro-differential equation (PIDE). Some useful works have been done in deriving analytical solutions, as well as numerical methods to solve the equations (see, e.g., Carr & Hirsa, 2003; Cont & Tankov, 2003). The works of Albanese, Campolieti, Chen, and Zavidonov (2003), and Albanese and Chen (2006) extend the distance-to-default framework (Hull & White, 2000; Avellaneda & Zhu, 2001) by introducing jumps to the credit index in a region separated by multibarriers. Crucial to the success of that approach, some technical conditions, such as that
the model is integrable in terms of the gamma function, must be met, and
general time-dependent coefficients cannot be conveniently incorporated.
Nevertheless, the results are quite encouraging in that consistently good
agreements with historical and market data are achieved.

Here we take a different approach based on similar ideas, more from a
differential equation point of view, which allows general variable coefficients in
the equation and takes advantage of the vast resources in numerical methods.
In this generalization, we no longer require the analytical tractability, instead
we develop efficient numerical algorithms that have the flexibility to
accommodate any state and time dependence in parameters, and exhibit more
transparency in connecting the model coefficients to data. The current model
follows the distance-to-default approach used by Avellaneda and Zhu (2001),
and the PIDE extension in a previous work of Zhu (2006), by utilizing the
information contained in the survival probability density function. The need to
model different rating regions by different parameters poses no difficulty for
the PIDE formulation and the finite difference method used to obtain
approximate solutions. For each set of parameters, a collection of solutions
corresponding to all initial ratings is generated and compared with the
historical transition frequency matrix. The calibration of the model is
performed by obtaining the best fit through the use of an optimization
package BFGS developed by Zhu, Byrd, and Noedel (1997). The stability
of the PIDE solutions ensures that the fitting procedure is well behaved.
We have tested with two similar rating transition frequency matrices from
Standard and Poor’s and Moody’s. Good agreements are found in both cases
and it is obvious that the jump diffusion model generates superior fits
compared to the Brownian motion model. Another simplification in this model
is the straightforward procedure to change from the real-world measure to the
risk-neutral measure. Because of the complicated issues of recovery rates and
tax rates, it is not expected for the model to be able to fit the market data.
Instead, it is hoped that sensible choices of the default boundary can provide
insight into the credit spread fraction of the yield spread as observed on the
market. The close connection between the default boundary found in
Avellaneda and Zhu (2001) and that found in the current model makes it
possible to calibrate the risk-neutral parameters based on other market sources.

The paper is organized as follows. In Section 2, we briefly introduce the
process and the resulting partial PIDE. In Section 3, we establish the rating
migration model through the survival density function. Then in Section 4,
we present some numerical results for the real-world calibration and an
application in a risk-neutral market in Section 5. We also offer some
concluding remarks in Section 6.
We will use the distance-to-default of a firm to determine its credit rating. Naturally a larger distance-to-default implies a higher credit quality. To model the process for the distance-to-default, it is important to include jumps in the index process, in addition to the Brownian motion. This is particularly crucial in rating migration modeling, as transitions with more than two ratings involved are quite significant and they account for over 10% of all the transitions, according to the study of Carty (1997). We consider the following discontinuous Lévy process:

\[ dX_t = a(X_t, t)dt + \sigma(X_t, t)dW_t + dq_t, \quad X_0 > 0 \]

Here \( W_t \) is the standard Brownian motion, \( \sigma \) is the volatility associated with the Brownian motion, \( a \) is a deterministic drift term, and \( q_t \) is a Poisson process with intensity \( \lambda \). Once a jump occurs, the probability measure of the jump amplitude (from \( x \) to \( y \)) is given by a known distribution

\[ G(x, dy) = P[x \rightarrow (y, y + dy)] \quad (1) \]

We assume that \( W_t, q_t \), and the jump amplitude are mutually independent.

According to the definition of the distance-to-default (Hull & White, 2000; Avellaneda & Zhu, 2001), a default of the firm is triggered when the distance-to-default \( X_t \) crosses the line \( x = 0 \) for the first time, that is, the default time is the first passage time

\[ \tau = \inf\{t \geq 0 : X_t \leq 0\} \quad (2) \]

Following the notations in Hull and White (2000) and Avellaneda and Zhu (2001), the cumulative default probability by time \( t \) is the probability that the default has occurred by \( t \), or more precisely,

\[ P(t) = P[\tau < t] \quad (3) \]

One of the goals of this model is to be able to compute this cumulative default probability, its default probability density defined through

\[ P'(t)dt = P[t < \tau \leq t + dt] \quad (4) \]

and match them with the information observed on the market.

To specify and study the credit state of the firm based on the information contained in the distance-to-default, our approach is to analyze the survival
probability density function $u$, which is defined by

$$u(x, t)\, dx = P[x < X_t \leq x + dx, \tau \geq t] \quad (5)$$

In order to obtain the equation for $u(x, t)$, it is necessary to derive the infinitesimal generator of the process, with the consideration of the killing when a path crosses the default barrier $x = 0$. Since the generator for the Gaussian diffusion with a known drift, and with the killing boundary condition at $x = 0$ is well known (see, e.g., Lamperti, 1977), we focus on the part associated with the jump diffusion with the same killing boundary condition. In the following, we briefly describe the derivation of the generator associated with the Poisson process $q_t$ with emphasis on the boundary condition.

Given a Poisson process with intensity $\lambda$, if we define $Z_t$ to be the number of jump occurrences in $[0, t]$, the probability of $n$ occurrences in this time interval is

$$P_n(t) = P[Z_t = n] = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$$

For small $t > 0$, we have the approximation

$$P_1(t) = \lambda t + o(t), \quad P_0(t) = 1 - \lambda t + o(t)$$

and

$$P_n(t) = o(t), \quad n \geq 2$$

Now we consider a smooth function $f(x)$ defined for $x \geq 0$, with the boundary value $f(0) = 0$. To overcome the overshooting problem discussed by Kou and Wang (Kou, 2002; Kou & Wang, 2003), we extend $f$ so that $f(x) = 0$ for $x < 0$, and also allow $\lambda$ to be state and time dependent. Assuming small $t$, the conditional expectation of $f(q_t)$ with $q_0 = x$ is given by

$$E^x[f(q_t)] = \sum_{n=0}^{\infty} P_n(t)E^x[f(q_t)|Z_t = n]$$

$$= P_0(t)f(x) + P_1(t)E^x[f(q_t)|Z_t = 1] + o(t)$$

$$= (1 - \lambda t)f(x) + \lambda t \int_0^\infty f(y)G(x, dy) + o(t)$$

Notice that in the last integral, we used the fact that if a jump results in the path crossing the default boundary, the corresponding value of $f$ is zero. As a consequence, the integral is only over the half space $x > 0$. 

Jump Diffusion in Credit Barrier Modeling

199
To obtain the infinitesimal generator, we evaluate
\[
\lim_{t \to 0^+} \frac{E^t[f(q_t)] - f(x)}{t} = \lambda \left( \int_0^\infty f(y)G(x, dy) - f(x) \right)
\]  
(6)

This is the infinitesimal generator associated with Poisson jumps, with the killing boundary condition at \( x = 0 \). For our combined process (2), the infinitesimal generator is given by
\[
Af(x) = \frac{1}{2}\sigma^2 f_{xx} + af_x + \lambda \left( \int_0^\infty f(y)G(x, dy) - f(x) \right)
\]  
(7)

with the boundary condition \( f|_{x=0} = 0 \). This generator is different from the operator used in Kou and Wang (2003) because of the extension of \( f \) and the boundary condition.

The evolution equation for the survival probability density function \( u \) can be obtained by the Fokker–Planck equation associated with this generator, via the adjoint operator. As a first step in building a class of models, we assume that the distribution function \( G \) has a simple form
\[
G(x, dy) = g(y - x)dy
\]  
(8)

for some density function \( g \).

The Fokker–Planck equation for the survival density function therefore, is
\[
\frac{\partial u}{\partial t} = \frac{1}{2}(\sigma^2 u)_{xx} - (au)_x + \int_0^\infty \lambda(y, t)u(y, t)g(x - y)dy - \lambda u, \quad x > 0
\]  
(9)

with initial condition
\[
u(x, 0) = \delta(x - X_0)
\]  
(10)

where \( X_0 > 0 \) is the initial distance-to-default, and the boundary conditions
\[
u|_{x=0} = 0, \quad \lim_{x \to \infty} u = \lim_{x \to \infty} u_x = 0, \quad \text{for } t > 0
\]  
(11)

For the jump amplitude distribution \( g(x) \), two natural suggestions arise due to their obvious convenience. One is the normal distribution
\[
g(x) = \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left(-\frac{(x - \mu_j)^2}{2\sigma_j^2}\right)
\]  
(12)

where the mean \( \mu_j \) and the variance \( \sigma_j^2 \) can be specified. The other choice, which has found many applications in recent years (Kou, 2002;
Kou & Wang, 2003), is the two-sided exponential distribution

\[
g(x) = \begin{cases} 
\frac{x}{\beta_+} e^{-\frac{x}{\beta_+}} & x > 0 \\
\frac{x}{\beta_+} + \frac{1-x}{\beta_-} & x = 0 \\
\frac{1-x}{\beta_-} e^{\frac{x}{\beta_-}} & x < 0 
\end{cases}
\] (13)

with parameters

\[0 < \alpha < 1, \quad \beta_+, \beta_- > 0\]

We use a normal distribution in this work.

3. CREDIT RATING MIGRATION MODEL

To use the distance-to-default process to determine the credit rating of an entity, we let \( R(t) \) to denote the rating of the entity at time \( t \), which is assigned by a rating agency such as Moody’s, based on the firm’s credit condition using the rating agency’s rating method. This rating will have a significant price impact on debts issued by the entity, therefore the transition dynamics of the rating is a central issue in credit modeling. Extensive historical data are available (in transition frequencies) and comprehensive studies are well documented (such as Carty, 1997). The first mathematical model for the transitions is based on Markov chains (Jarrow et al., 1997) and it is focused on the transition probability matrix \( Q(t, T) = (q_{i,j}(t, T)) \) that describes the transition of rating from time \( t \) to a later time \( T \), where

\[q_{i,j}(t, T) = P[R(T) = j, |R(t) = i], \quad t < T\]

In the time-homogeneous case where \( Q \) has the form \( Q(t, T) = Q(T-t) \), one of the approaches (Lando & Skødeberg, 2002) to analyze the transition matrix is to find the generating matrix \( \Lambda \) such that

\[Q(t) = [q_{i,j}(t)] = e^{\Lambda t}\] (14)

A common theme in the works along this direction concerns with modeling the generator \( \Lambda \) by some stochastic process.

The distance-to-default methodology provides us with a natural way to determine the rating at time \( t \). Assuming that there are \( K \) non-default
ratings, we first divide the half space \([0, \infty)\) into \(K\) subregions for the distance-to-default, separated by \(x = b_j, j = 1, \ldots, K - 1\), and the default boundary \(x = b_0 = 0\). We introduce the rating system in which the rating is determined to be \(j\) at time \(t\) if the distance-to-default \(X_t\) satisfies \(b_{j-1} \leq X_t < b_j\), for some \(1 \leq j \leq K - 1\), or \(K\) if \(X_t \geq b_{K-1}\). If \(X_t \leq 0\), then the firm has entered the default state, which is an absorbing state that implies no recovery in the future. An illustration of the credit barriers is given in Fig. 1.

A similar approach is taken in Albanese et al. (2003) and Albanese and Chen (2006) with a probabilistic treatment. However, due to the restriction of analytic tools, the model parameters in the probabilistic approach are limited to constants. The general PIDE setting described above and the availability of extensive numerical tools suggest that these restrictions could be removed in the current approach, and that will lead to a model flexible enough to incorporate all the key factors found in the economy and the market, and also connect with the Markov chain generator in Eq. (14).

For a given set of drift, volatility, and jump intensity, the solution \(u(x, t)\) to Eq. (9) for \(t > 0\) provides useful information that can be interpreted according to the locations of the barriers. To be more specific, let \(u^j\) denote a solution that represents the survival probability density for a firm which has initial rating \(j\), with the initial condition

\[
u^j(x, 0) = \delta(x - X^j_0)
\]

Here the initial index value \(X^j_0\) satisfies

\[
b_{j-1} < X^j_0 \leq b_j, \text{ if } j = 1, \ldots, K - 1
\]

or

\[
X^j_0 > b_{K-1}, \text{ if } j = K
\]

![Fig. 1. An Illustration of the Credit Barrier Model.](image-url)
To be more general, we only need
\[ u^i(x, 0) \begin{cases} \geq 0, & b_{j-1} < x < b_j, \\ 0, & \text{elsewhere} \end{cases} \quad j = 1, \ldots, K - 1 \] (18)

and
\[ u^K(x, 0) \begin{cases} \geq 0, & x > b_{K-1} \\ 0, & \text{elsewhere} \end{cases} \] (19)
satisfying
\[ \int_{b_{j-1}}^{b_j} u^i(x, 0) \, dx = 1, \quad j = 1, \ldots, K - 1 \]

and
\[ \int_{b_{K-1}}^{\infty} u^K(x, 0) \, dx = 1 \]

With such solutions, the following integrals
\[ q_{j,k} = \int_{b_{k-1}}^{b_k} u^j(x, T) \, dx, \quad k = 1, \ldots, K - 1 \] (20)

and
\[ q_{j,K} = \int_{b_{K-1}}^{\infty} u^j(x, T) \, dx \] (21)
give the transition probabilities from rating \( j \) to other non-default ratings over the time period \([0, T]\). The default probability for firms with initial rating \( j \) can also be recovered from the solutions
\[ P_j(t) = 1 - \sum_{k=1}^{K} q_{j,k} = 1 - \int_{0}^{\infty} u^j(x, T) \, dx \] (22)

It is expected that this continuous time model should cover the situations described by the original discrete Markov chain model. Furthermore, there is an added advantage that defaults and rating changes can occur at any time, represented simply by barrier crossings. The inclusion of Poisson jumps avoids the limitation in a Brownian diffusion model that multibarrier crossings are excluded. The transparency of the continuous model is
obvious: the parameters $a$, $\sigma$, and $\lambda$ can be naturally interpreted and required to match economic and market conditions. If properly calibrated, the volatility structure can be viewed as a media property of the material to support the economic activities, whereas the drift and intensity parameters can be used to describe market adjustments and firm's shock conditions. It is, therefore, possible to model multiname credit dynamics in this setting and impose a correlation structure among different names. This media property approach provides a platform to explore many such issues, especially when efficient Monte Carlo simulations are developed.

The feasibility of the PIDE calibration lies in the effectiveness of the numerical methods to solve the equation, and the accuracy of the solutions. We develop a stable finite difference method that is second order in space but first order in time. The discretization takes the following form: the half space $[0, \infty)$ is truncated to $[0, L]$, where $L$ is large enough so the finite interval covers sufficient information contained in the original problem. A small time step $\Delta t$ is also introduced, so the time derivative can be approximated by a finite difference. For approximations to the spatial derivatives, the standard centered difference is used for $u_{xx}$, and an upwind scheme is introduced for the drift term, an implicit treatment for the diffusion part is introduced for the consideration of numerical stability. The integral is approximated by an explicit trapezoidal rule, involving solutions at the previous time step only. The resulting linear system is similar to that of the Crank–Nicolson approximation to solve the heat equation. The numerical scheme has proved to be robust and accurate for our applications.

4. CALIBRATION TO HISTORICAL RATING TRANSITION MATRICES

The calibration of the model takes two sets of parameters: the volatility and jump structure, and the drift information. As the first step, we proceed to fit the historical data. The volatility structure, when viewed as a media property, can be carried over from the real-world measure to the risk-neutral measure in pricing applications. This can be justified by using iso-volatility arguments. For studies of the historical data, extensive analysis is performed by Carty (1997) and we use the conclusions there as our benchmark test criteria. The goal is to choose those parameters so that the transition matrix generated from the solutions of Eq. (9) matches the historical transition matrix as closely as possible. In calculations contained
in this work, we assume $K = 7$, $b_k = k$, for $k = 0, 1, \ldots, 6$, and a truncated state–space interval $[0, 20]$, which is subdivided into 200 equal-length subintervals for finite difference approximations. For the time discretization we use $\Delta t = 0.01$, which satisfies the numerical stability condition.

In the quest to minimize the difference between the model output and the historical data, we choose the following objective function

$$f(a, \sigma, \lambda) = \|Q_d - Q_{a,\sigma,\lambda}\|_{F,W}^2$$

(23)

where $Q_d$ is the historical transition probability matrix, $Q_{a,\sigma,\lambda}$ is the transition probability matrix from the model, and the matrix norm is the weighted Frobenius norm with weights $W$. Here $a$, $\sigma$, and $\lambda$ represent the sets of drift, volatility, and jump parameters. For the historical data sets listed below, there will be a total of 21 parameters to be determined. They are the drift values $a_k$, volatility values $\sigma_k$, and jump intensity values $\lambda_k$, for $k = 1, \ldots, 7$. The variable coefficients used in Eq. (9) are constructed based on these values in a piecewise linear form, with proper treatment at the two ends. For the distribution of jump sizes, we choose the standard normal distribution to limit the number of free parameters in the optimization procedure. The starting positions of the processes for different ratings are another factor in this model, and we choose to fix them at the centers of the subregions, with an initial smoothed delta function to meet practical requirements in a numerical approximation. It should be pointed out that should there be a need to improve the accuracy, one can easily allow these extra parameters to vary so a better fit may be generated. The optimization package we choose to use is L-BFGS-B (version 2.1) (Zhu et al., 1997) in which linear bound constraints can be imposed.

As the first example, we use the transition frequency matrix from Standard and Poor’s 1981–1991 data (Table 1) to test our model and study the difference between the Brownian diffusion model and the jump diffusion model.

In Tables 2 and 3, we demonstrate the advantage of the jump diffusion model by showing a marked improvement in fitting the transition probabilities, with emphasis on the off-diagonal entries corresponding to migrations crossing one or two neighboring ratings. Here the individual errors for the fit are included in parentheses below the main entries.

Next we study the transition matrix from Moody’s 1980–2000 data (Duffie & Singleton, 1999) in Table 4, in contrast to the data set from Carty’s (1997) study. The choice is made on the basis that the transition frequency matrix from (Carty, 1997) covers a long period (1920–1990) and
many world and economic events played important roles in the data. Since we have the goal to link with the credit derivative market, it seems more appropriate to use data that are more relevant to our current time.

The transition matrix listed below is for average transition frequencies over a 1-year period, and the transition probabilities are adjusted to reflect withdrawn ratings. According to the study of Carty (1997), ratings are

### Table 1. Historical Average 1-Year Rating Transition Frequencies, Standard and Poor’s 1981–1991.

<table>
<thead>
<tr>
<th>Rating From</th>
<th>AAA</th>
<th>AA</th>
<th>A</th>
<th>BBB</th>
<th>BB</th>
<th>B</th>
<th>CCC</th>
<th>Default</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAA</td>
<td>89.10</td>
<td>9.63</td>
<td>0.78</td>
<td>0.19</td>
<td>0.30</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>AA</td>
<td>0.86</td>
<td>90.10</td>
<td>7.47</td>
<td>0.99</td>
<td>0.29</td>
<td>0.29</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>A</td>
<td>0.09</td>
<td>2.91</td>
<td>88.94</td>
<td>6.49</td>
<td>1.01</td>
<td>0.45</td>
<td>0.00</td>
<td>0.09</td>
</tr>
<tr>
<td>BBB</td>
<td>0.06</td>
<td>0.43</td>
<td>6.56</td>
<td>84.27</td>
<td>6.44</td>
<td>1.60</td>
<td>0.18</td>
<td>0.45</td>
</tr>
<tr>
<td>BB</td>
<td>0.04</td>
<td>0.22</td>
<td>0.79</td>
<td>7.19</td>
<td>77.64</td>
<td>10.43</td>
<td>1.27</td>
<td>2.41</td>
</tr>
<tr>
<td>B</td>
<td>0.00</td>
<td>0.19</td>
<td>0.31</td>
<td>0.66</td>
<td>5.17</td>
<td>82.46</td>
<td>4.35</td>
<td>6.85</td>
</tr>
<tr>
<td>CCC</td>
<td>0.00</td>
<td>0.00</td>
<td>1.16</td>
<td>1.16</td>
<td>2.03</td>
<td>7.54</td>
<td>64.93</td>
<td>23.19</td>
</tr>
</tbody>
</table>

### Table 2. Fitting the Standard and Poor’s Data by the Brownian Diffusion Model.

<table>
<thead>
<tr>
<th>Rating From</th>
<th>AAA</th>
<th>AA</th>
<th>A</th>
<th>BBB</th>
<th>BB</th>
<th>B</th>
<th>CCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAA</td>
<td>89.87</td>
<td>10.44</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>AA</td>
<td>(0.77)</td>
<td>(0.81)</td>
<td>(−0.78)</td>
<td>(−0.19)</td>
<td>(−0.30)</td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>A</td>
<td>1.65</td>
<td>90.79</td>
<td>8.13</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>BBB</td>
<td>(0.79)</td>
<td>(0.69)</td>
<td>(0.66)</td>
<td>(−0.99)</td>
<td>(−0.29)</td>
<td>(−0.29)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>BB</td>
<td>0.00</td>
<td>3.32</td>
<td>89.70</td>
<td>7.26</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>B</td>
<td>(−0.09)</td>
<td>(0.41)</td>
<td>(0.76)</td>
<td>(0.77)</td>
<td>(−1.01)</td>
<td>(−0.45)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>CCC</td>
<td>0.00</td>
<td>0.00</td>
<td>7.96</td>
<td>85.17</td>
<td>7.37</td>
<td>0.01</td>
<td>0.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rating From</th>
<th>AAA</th>
<th>AA</th>
<th>A</th>
<th>BBB</th>
<th>BB</th>
<th>B</th>
<th>CCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAA</td>
<td>(−0.06)</td>
<td>(−0.43)</td>
<td>(1.40)</td>
<td>(0.90)</td>
<td>(0.93)</td>
<td>(−1.59)</td>
<td>(−0.18)</td>
</tr>
<tr>
<td>AA</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>7.11</td>
<td>79.14</td>
<td>11.36</td>
<td>0.00</td>
</tr>
<tr>
<td>A</td>
<td>(−0.04)</td>
<td>(−0.22)</td>
<td>(−0.79)</td>
<td>(−0.08)</td>
<td>(1.50)</td>
<td>(0.93)</td>
<td>(−1.27)</td>
</tr>
<tr>
<td>BBB</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>9.22</td>
<td>84.56</td>
<td>8.24</td>
<td>0.00</td>
</tr>
<tr>
<td>BB</td>
<td>(0.00)</td>
<td>(−0.19)</td>
<td>(−0.31)</td>
<td>(−0.66)</td>
<td>(4.05)</td>
<td>(2.10)</td>
<td>(3.89)</td>
</tr>
<tr>
<td>B</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>6.90</td>
<td>65.47</td>
<td>0.00</td>
</tr>
<tr>
<td>CCC</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(−1.16)</td>
<td>(−1.16)</td>
<td>(−2.03)</td>
<td>(−0.64)</td>
<td>(0.54)</td>
</tr>
</tbody>
</table>
withdrawn for various reasons and the majority cases are due to the fact that an issue had either matured or had been called, which does not necessarily imply a deterioration in its credit condition. Similar transition frequencies are available for 3- and 5-year periods, but substantial noises are contained in these data sets. Instead, for periods beyond 1 year, we choose to fit the default probabilities only. In Table 5, we average the default

### Table 3. Fitting the Standard and Poor’s Data by the Jump Diffusion Model.

<table>
<thead>
<tr>
<th>Rating From</th>
<th>AAA</th>
<th>AA</th>
<th>A</th>
<th>BBB</th>
<th>BB</th>
<th>B</th>
<th>CCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAA</td>
<td>89.28</td>
<td>9.80</td>
<td>1.05</td>
<td>0.13</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>(0.18)</td>
<td>(0.17)</td>
<td>(0.27)</td>
<td>(−0.06)</td>
<td>(−0.29)</td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>AA</td>
<td>1.25</td>
<td>90.30</td>
<td>7.68</td>
<td>0.39</td>
<td>0.06</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>(0.39)</td>
<td>(0.20)</td>
<td>(0.21)</td>
<td>(−0.60)</td>
<td>(−0.23)</td>
<td>(−0.28)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>A</td>
<td>0.71</td>
<td>2.93</td>
<td>89.06</td>
<td>6.63</td>
<td>1.01</td>
<td>0.17</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>(0.62)</td>
<td>(0.02)</td>
<td>(0.12)</td>
<td>(0.14)</td>
<td>(0.00)</td>
<td>(−0.28)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>BBB</td>
<td>0.16</td>
<td>1.30</td>
<td>6.47</td>
<td>84.31</td>
<td>6.42</td>
<td>1.56</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>(0.10)</td>
<td>(0.87)</td>
<td>(−0.09)</td>
<td>(0.04)</td>
<td>(−0.02)</td>
<td>(−0.04)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>BB</td>
<td>0.02</td>
<td>0.21</td>
<td>1.75</td>
<td>7.12</td>
<td>77.79</td>
<td>10.62</td>
<td>1.82</td>
</tr>
<tr>
<td></td>
<td>(−0.02)</td>
<td>(−0.01)</td>
<td>(0.96)</td>
<td>(−0.07)</td>
<td>(0.15)</td>
<td>(0.19)</td>
<td>(0.55)</td>
</tr>
<tr>
<td>B</td>
<td>0.00</td>
<td>0.01</td>
<td>0.11</td>
<td>0.70</td>
<td>5.59</td>
<td>82.75</td>
<td>4.82</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(−0.18)</td>
<td>(−0.20)</td>
<td>(0.04)</td>
<td>(0.42)</td>
<td>(0.29)</td>
<td>(0.47)</td>
</tr>
<tr>
<td>CCC</td>
<td>0.00</td>
<td>0.00</td>
<td>0.03</td>
<td>0.26</td>
<td>2.00</td>
<td>7.57</td>
<td>64.92</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(−1.13)</td>
<td>(−0.90)</td>
<td>(−0.03)</td>
<td>(0.03)</td>
<td>(−0.01)</td>
</tr>
</tbody>
</table>

### Table 4. Historical Average 1-Year Rating Transition Frequencies, Moody’s 1980–2000, Normalized for Withdrawn Ratings.

<table>
<thead>
<tr>
<th>Rating From</th>
<th>Aaa</th>
<th>Aa</th>
<th>A</th>
<th>Baa</th>
<th>Ba</th>
<th>B</th>
<th>Caa-C</th>
<th>Default</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aaa</td>
<td>89.14</td>
<td>9.78</td>
<td>1.06</td>
<td>0.00</td>
<td>0.03</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Aa</td>
<td>1.14</td>
<td>89.13</td>
<td>9.25</td>
<td>0.32</td>
<td>0.11</td>
<td>0.01</td>
<td>0.00</td>
<td>0.04</td>
</tr>
<tr>
<td>A</td>
<td>0.06</td>
<td>2.97</td>
<td>90.28</td>
<td>5.81</td>
<td>0.69</td>
<td>0.18</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>Baa</td>
<td>0.06</td>
<td>0.36</td>
<td>7.01</td>
<td>85.47</td>
<td>5.82</td>
<td>1.02</td>
<td>0.08</td>
<td>0.18</td>
</tr>
<tr>
<td>Ba</td>
<td>0.03</td>
<td>0.07</td>
<td>0.59</td>
<td>5.96</td>
<td>82.41</td>
<td>8.93</td>
<td>0.58</td>
<td>1.43</td>
</tr>
<tr>
<td>B</td>
<td>0.01</td>
<td>0.04</td>
<td>0.22</td>
<td>0.61</td>
<td>6.43</td>
<td>82.44</td>
<td>3.29</td>
<td>6.96</td>
</tr>
<tr>
<td>Caa-C</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.95</td>
<td>2.85</td>
<td>6.15</td>
<td>62.36</td>
<td>27.69</td>
</tr>
</tbody>
</table>
probabilities for alphanumeric ratings contained in Albanese and Chen (2006) to obtain a collection of historical default probabilities for seven different ratings. Together with the transition frequency matrix above, they form the main data set we set out to fit by choosing parameters to minimize the differences between the data and the model outputs. In Table 6, the transition matrix generated from the model for the Moody’s data and the fitting errors are listed. We find the similar behavior in errors as that in Table 3 for Standard and Poor’s data. The similarity supports our claim that the developed algorithm is quite robust.

Table 5. Averaged Default Probabilities Estimated from Moody’s Data.

<table>
<thead>
<tr>
<th>Rating</th>
<th>1 Year (%)</th>
<th>3 Years (%)</th>
<th>5 Years (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aaa</td>
<td>0.00</td>
<td>0.00</td>
<td>0.39</td>
</tr>
<tr>
<td>Aa</td>
<td>0.04</td>
<td>0.07</td>
<td>0.54</td>
</tr>
<tr>
<td>A</td>
<td>0.00</td>
<td>0.38</td>
<td>0.83</td>
</tr>
<tr>
<td>Baa</td>
<td>0.18</td>
<td>1.47</td>
<td>2.43</td>
</tr>
<tr>
<td>Ba</td>
<td>1.43</td>
<td>8.71</td>
<td>16.22</td>
</tr>
<tr>
<td>B</td>
<td>6.96</td>
<td>26.93</td>
<td>38.80</td>
</tr>
<tr>
<td>Caa-C</td>
<td>27.69</td>
<td>50.14</td>
<td>60.12</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>Rating From</th>
<th>Rating To</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Aaa</td>
</tr>
<tr>
<td>Aaa</td>
<td>89.21</td>
</tr>
<tr>
<td></td>
<td>(0.07)</td>
</tr>
<tr>
<td>Aa</td>
<td>1.20</td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
</tr>
<tr>
<td>A</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>(0.43)</td>
</tr>
<tr>
<td>Baa</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
</tr>
<tr>
<td>Ba</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>(−0.02)</td>
</tr>
<tr>
<td>B</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>(−0.01)</td>
</tr>
<tr>
<td>Caa-C</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
</tr>
</tbody>
</table>
One advantage of this PIDE model is the transparency of the model parameters that comes across in the calibration procedure. It is, therefore, important to study the parameter structures resulted from the optimization. The drift term $a$ is intended to show the general trend in each of the regions, reflecting different behaviors for different ratings. It is observed by Carty (1997) that in general most firms tend to be downgraded. However, those lower grades, if they managed to survive in their early stages, often have a tendency to slightly move up later, unless they are very close to default, such as Caa-C ratings.

In Fig. 2, we plot the parameter structures for the periods of the first year, the second and the third years, and the fourth and the fifth years. Here we impose a fixed structure for the drift, volatility, and jump intensity in each of the time periods. In the first year, the maximum drift (moving toward the higher ratings) appears in the state–space interval $1 \leq x \leq 2$, which indicates that issues with initial B ratings actually tend to move up. The minimum (moving toward default) occurs in the lowest rating Caa-C, which comes as

![Fig. 2. Parameter Structures Obtained from Optimization, by Matching the Transition Frequency Matrices and Default Probabilities.](image-url)
no surprise. Also as expected, for high quality ratings, the general tendency is to move downward, although at smaller and smaller magnitudes as we move to higher ratings. These properties are quite clearly verified in Fig. 2. As we move to later time periods, the situation changes quite drastically, with the most significant change in the drift for the lowest rating: it is now quite large and tends to push the rating toward the higher rating regions. This suggests that once a low rating company survives the first year, it has a strong possibility to be upgraded. The volatility and jump intensity structures have similar behaviors in that the significant contributions are mostly in the lower rating regions: volatility tends to be high in the next-to-lowest rating region, except that it becomes almost monotone decreasing in the time period of fourth and fifth years, where the maximum clearly occurs in the lowest rating region. For the jump intensity, it is obvious that jumps are most significantly needed in the lowest rating region. From all the discussions above, there is strong evidence to suggest that the interpretation from the parameter structures is consistent with economic observations and intuitions.

In Fig. 3, we plot the comparisons of default probabilities between the Moody’s data and the model output for time periods of 1 year, 3 years, and
5 years. Good agreements are found in most of these comparisons, except in the first-year period where the optimization is based on fitting the transition matrix, which compromises the accuracy in default probability fits. Also the fit for high-grade ratings is not as good. This is due to the fact that we use uniform weights in the optimization procedure and the small magnitude entries suffer larger relative errors.

5. CHANGE TO THE RISK-NEUTRAL MEASURE

One of the major applications of a credit model is to price credit derivatives, which requires to bring the model to the risk-neutral world. To carry out the change of measure, a transformation for the state variable is required. Unlike the CIR process used in Albanese et al. (2003) and Albanese and Chen (2006), the jump process we use allows a simple linear transformation, which will retain the form of the PIDE with only a change in the drift term. To see this, we change the default boundary from $x = 0$ to $x = b(t)$ for some $b(t) \geq 0$. Eq. (9) is consequently modified in the boundary location and the integral domain. If we introduce a new process

$$Y_t = X_t - b(t)$$

(24)

the form of the PIDE in the state variable $y$ will be the same as the one for the state variable $x$ in the case of default boundary $x = 0$. The boundary and integral domain are the same as those in the original equation, except that the drift $a(x, t)$ is replaced by $a(y + b(t), t) - b'(t)$. This simple change of variables leaves the volatility invariant, which is appropriate if we do not include the consideration of the market price for jump risk (Albanese et al., 2003).

The choice of the correct default boundary, however, involves many factors that have not been adequately addressed in mathematical models, such as the recovery rates and the tax adjustments for different ratings. In this work, we ignore the issue of finding the correct default boundary, instead we demonstrate with a simple curve the flexibility and transparency achieved by the current model. Here we assume a piecewise cubic default boundary for $t \in [0, 5]$ that has a continuous derivative. The boundary is also monotone increasing in time, which implies that the drift adjustment always enhances the tendency to move toward a lower rating. This corresponds to a risk-neutral premium which ensures that the implied default rate in the risk-neutral world is always greater than or equal to the default probability in the real world. For the purpose of comparison, we consider the market yield spreads used in (Albanese et al., 2003), the real-world default rate data in
Table 5, and the model generated credit yields, based on our fitting for the transition frequency matrix in Table 4 and default rates in Table 5, adjusted to the default boundary mentioned above. To convert between the yield spread and the default probabilities, we assume zero-coupon bonds, adopt the same recovery rates as in (Albanese et al., 2003), and use the relation

\[ e^{-s(T)}T = 1 - P(T)(1 - Re) \]  

(25)

where \( s(T) \) is the yield spread with maturity \( T \), \( P(T) \) the default probability before time \( T \), and \( Re \) the expected recovery rate which depends on the rating. In Fig. 4, we compare the market yield spread, the credit spread from our model, and the spread implied from the actual default probabilities for ratings A, Ba, and B, under the same assumption of recovery rates. The default boundary used for the example is also plotted in the figure. We consistently notice the observation by Huang and Huang (2003) that credit spread is a small part of the yield spread in higher ratings and is a larger fraction in lower ratings. In the result for the 1-year period, the credit spread from the model is
lower than the implied spread from actual default probabilities. This is because of the poor fit in default probabilities, resulting from the objective function that leads to fitting the transition frequency matrix, rather than just the default probabilities in this time period. There are conflicting considerations in fitting individual matrix entries versus fitting the sums of matrix rows (default probabilities). This suggests that it is necessary to be flexible with the objective function in the fit if we use the model only to study default probabilities and spreads.

6. CONCLUSION

We introduce a jump diffusion model to study the credit rating migration problem, and also the default probability versus the credit spread problem. Instead of focusing on analytically tractable processes, we develop a PIDE formulation with general variable coefficients and show that efficient numerical solutions are practical. The lack of dependence on analytic tractability makes the model quite robust to fit a wide variety of data. The model and the efficient numerical methods are straightforward to implement, and good agreements with several historical data sets have been achieved. It is shown that the introduction of Poisson jumps significantly improves the fitting of the transition frequency matrices. Since there are still many issues concerning the historical rating transition data, it is important to have a model that is stable with respect to the changes in collected data. The extensive knowledge about the PIDE and the existing numerical tools suggest that this approach can be used in many practical applications, and the simple and flexible process adopted should lead to efficient Monte Carlo simulations in credit derivative pricing and management.

ACKNOWLEDGMENTS

The author thanks Tongshu Ma and Zhifeng Zhang for helpful discussions, and the referee for useful comments.

REFERENCES


BOND MARKETS WITH STOCHASTIC VOLATILITY

Rafael DeSantiago, Jean-Pierre Fouque and Knut Solna

ABSTRACT

We analyze stochastic volatility effects in the context of the bond market. The short rate model is of Vasicek type and the focus of our analysis is the effect of multiple scale variations in the volatility of this model. Using a combined singular-regular perturbation approach we can identify a parsimonious representation of multiscale stochastic volatility effects. The results are illustrated with numerical simulations. We also present a framework for model calibration and look at the connection to defaultable bonds.

1. INTRODUCTION

In this paper we illustrate the role of stochastic volatility in the case of interest rate markets. Our main example is the pricing of zero-coupon bonds when the interest rate is defined in terms of a Vasicek model, as well as the pricing of options on bonds. We use the singular perturbation framework set forth in Fouque, Papanicolaou, and Sircar (2000) and extend the results in Cotton, Fouque, Papanicolaou, and Sircar (2004) to the case where the volatility is driven by a slow process in addition to the fast process considered there.
The fact that zero-coupon bonds are parameterized by two time indices (the time at which the contract begins, and the maturity date) means that arbitrage restrictions across different maturities have to be taken into account. Also, note that options on bonds can be written on an infinite number of bonds indexed by their different maturity date and each bond cannot be treated independently, for bonds of different maturities are correlated. The most important difference between the classic Black–Scholes scenario and interest rate markets is the fact that the short rate is not the price of a traded asset.

The main building block to price many other financial instruments in interest rate markets is the zero-coupon bond. In Section 2, we show two different approaches to price zero-coupon bonds. In the following section, we review briefly a class of models that have desirable properties in terms of modeling interest rate markets, namely, the affine models for the short rate. We focus then on one such model, the Vasicek model. In Section 4, we introduce stochastic volatility in the Vasicek framework by letting the volatility be driven by two stochastic processes that vary on two different time scales. In Section 5, we compute an asymptotic approximation to the bond price. This gives a parsimonious representation that is useful for calibration purposes, as presented in Section 7. Finally, in Section 8, we discuss the connection to the case with a defaultable bond.

2. PRICING BONDS

We define a zero-coupon bond with maturity $T$ as a contract, subscribed at the present time $t$ that guarantees the holder one dollar to be paid at time $T$ (with $t < T$). We begin by assuming that, under the subjective measure $\mathcal{P}$ the short rate follows the dynamics

$$
\text{d}r_t = \kappa(t, r_t)\text{d}t + \beta(t, r_t)\text{d}W_t
$$

(1)

where $W_t$ is a standard $\mathcal{P}$-Brownian motion, and we assume that $\kappa$ and $\beta$ are continuous with respect to $t$, and such that they satisfy the usual conditions for a strong solution. The money market account, $B_t$, is defined by

$$
\text{d}B_t = r_tB_t\text{d}t
$$

(2)

No-arbitrage pricing consists in pricing a contingent claim (the *derivative*) in terms of some underlying asset. In the Black–Scholes setting one typically has two processes: one that represents the price of the risky asset (usually a stock), and another one that represents the money market account.
In our case we also have two processes, \( r_t \) and \( B_t \), given by Eqs. (1) and (2), and it would seem natural to price the zero-coupon bond as a “derivative of the short rate”: that is, in order to compute the no-arbitrage price of the bond we would like to find a replicating strategy, based on the money account and some underlying asset, that gives $1 at time \( T \). The problem is that Eq. (1) does not represent the price of a traded asset. The only asset whose price is given exogenously is the money account, so we do not have interesting ways of forming replicating strategies (or even self-financing strategies).

We can get a better understanding of what the problem is if we try to price a zero-coupon bond. This can be done in two different ways. One is to form strategies with bonds of different maturities. But note that this is quite different than what is done in the Black–Scholes case, where a typical replicating strategy consists of holdings of the money account and of the underlying risky asset; now our portfolio would contain holdings of two different contingent claims, that is, of two derivatives (two bonds of different maturities), and maybe of the money account. If we consider the price process of one of the bonds as given, then we could price the other bond relative to the given benchmark bond.

The other way is to find an appropriate martingale measure that allows us to compute the price according to the general theory of derivative pricing. A zero-coupon bond can be considered as a contingent claim with payoff equal to one. Hence, the bond price, \( P(t, T) \), is given by

\[
P(t, T) = E^Q \left[ e^{-\int_t^T r_s \, ds} \cdot 1 \mid \mathcal{F}_t \right] = E^Q \left[ e^{-\int_t^T r_s \, ds} \mid \mathcal{F}_t \right]
\]

(3)

where \( \mathcal{F}_t \) is the filtration associated with the Brownian motion \( W_t \). Note, however, that we now need to know the equivalent martingale measure \( Q \).

In the Black–Scholes case, the equivalent martingale measure was found in the following way: if we assume that the price of the underlying asset, \( S_t \), is given by

\[
dS_t = \mu_t S_t dt + \sigma_t S_t dW_t
\]

(4)

where and \( \mu_t \) and \( \sigma_t \) satisfy some appropriate conditions, then the new measure \( Q \) was defined by the Radon–Nikodym derivative

\[
\frac{dQ}{d\mathcal{P}} := e^{-\int_0^t \theta_s dW_s - \frac{1}{2} \int_0^t \theta_s^2 ds}
\]

(5)

with \( \theta_t = (\mu_t - r_t)/\sigma_t \). Note that \( \theta_t \), and therefore \( Q \), are uniquely determined by the given \( \mathcal{P} \)-dynamics of the risky asset (\( \mu_t \) and \( \sigma_t \)).
In our case the short rate dynamics is not enough to uniquely determine the equivalent martingale measure.

2.1. The Term Structure Equation

In this section we recall standard arguments for obtaining the partial differential equation giving the price of a zero-coupon bond.

**Assumption 1.** We assume that the bond price depends on the short rate,

\[ P(t, T) = P(t, r_t, T) \]

We also assume, from now on, that \( P(t, x, T) \) has continuous partial derivatives up to second order with respect to the first two variables, and up to first order with respect to \( T \).

As we do not have a risky (underlying) asset, we form a portfolio with bonds of two different maturities. In particular, we let our portfolio contain \( \theta_1 \) bonds with maturity \( T_1 \) and \( \theta_2 \) bonds with maturity \( T_2 \). When no risk of confusion with the time indices exists we will use the notation \( P_1 = P(t, T_1) \) and \( P_2 = P(t, T_2) \) for the corresponding prices of the bonds. Applying Itô’s formula, we get

\[
dP_1 = \frac{\partial P_1}{\partial t} dt + \frac{\partial P_1}{\partial x} (\kappa dt + \theta dW_t) + \frac{1}{2} \sigma^2 \frac{\partial^2 P_1}{\partial x^2} dt
\]

and the analogous equation for \( P_2 \). If we let

\[
\mu_1 = \frac{1}{P_1} \left( \frac{\partial P_1}{\partial t} + \kappa \frac{\partial P_1}{\partial x} + \frac{\theta^2}{2} \frac{\partial^2 P_1}{\partial x^2} \right) \quad \text{and} \quad \sigma_1 = \frac{\theta}{P_1} \frac{\partial P_1}{\partial x}
\]

we can then write the price dynamics of the \( T_1 \) bond as

\[
dP_1 = \mu_1 P_1 dt + \sigma_1 P_1 dW_t
\]

The analogous equation holds for the \( T_2 \) bond. If we impose the condition that our strategy has to be self-financing, and we choose our portfolio in such a way that we eliminate the random part of the portfolio dynamics, then, by absence of arbitrage, this portfolio must have a rate of return equal to the short rate. This leads to a relationship between \( r_t \) and the drift and volatility of bonds of each maturity \( T \) (the details can be found in Bjork, 1998,
Chapter 16). If we let \( \mu_T \) and \( \sigma_T \) have the analogous meaning as Eq. (6), but for an arbitrary maturity time \( T \), then we obtain the following result.

**Proposition 1.** If the bond market is arbitrage-free, there exists a process \( \lambda \), such that

\[
\frac{\mu_T(t) - r_t}{\sigma_T(t)} = \lambda_t
\]  

(8)

holds for all \( t \), and for every maturity time \( T \). The process \( \lambda \) is known as the market price of risk.

Proposition 1 may also be expressed like this: if the bond market is free of arbitrage, bonds of all maturities must have the same market price of risk. The dynamics of the bond price, \( P_T = P(t, T) \), is now given by

\[
dP_T = \mu_T P_T dt + \sigma_T P_T dW_t
\]

If we substitute the expressions for \( \mu_T \) and \( \sigma_T \) (which are given by Eq. (6) with the subscript \( T \)) in the last equation, we obtain

\[
\frac{\partial P_T}{\partial t} + \kappa \frac{\partial P_T}{\partial x} + \frac{1}{2} \sigma^2 \frac{\partial^2 P_T}{\partial x^2} - r_T P_T = \Theta \lambda \frac{\partial P_T}{\partial x}
\]

As \( \lambda \) is independent of maturity we do not need to keep track of \( T \), so we let \( P = P_T \). We now rearrange the last equation to obtain the following result.

**Theorem 1.** If the bond market is arbitrage-free, the price of a bond of maturity \( T \) is given by the boundary value problem

\[
\begin{cases}
\frac{\partial P}{\partial t} + (\kappa - \lambda \Theta) \frac{\partial P}{\partial x} + \frac{1}{2} \sigma^2 \frac{\partial^2 P}{\partial x^2} - xP = 0 \\
P(T, x, T) = 1
\end{cases}
\]  

(9)

This partial differential equation is referred to as the term structure equation.

Note that now \( \lambda \) is not determined within the model: it is determined by the market.

### 2.2. Probabilistic Representation of the Bond Price

We have seen earlier that the bond price could be computed as a conditional expectation, but we did not know with respect to which measure. Proposition 1 shows how to construct the measure.
Let \( P^*(t, T) := e^{-\int_0^T r_s \ud s} P(t, T) \). An equivalent martingale measure \( Q \) is said to be an equivalent martingale measure for the bond market if \( Q \) is equivalent to \( P \) and the process \( P^*(t, T) \) is a martingale under \( Q \), for all maturities \( T \). Thus, we let \( \bar{T} \) denote the largest maturity and identify the measure \( Q \) by the Radon–Nikodym derivative

\[
\frac{\ud Q}{\ud P} := e^{-\int_0^{\bar{T}} \lambda_s \ud W_s - \frac{1}{2} \int_0^{\bar{T}} \lambda_s^2 \ud s}
\]

where \( \lambda \) is given by Eq. (8). Then, restricted to \( \mathcal{F}_{\bar{T}} \), we have

\[
\frac{\ud Q}{\ud P} := e^{-\int_0^{\bar{T}} \lambda_s \ud W_s - \frac{1}{2} \int_0^{\bar{T}} \lambda_s^2 \ud s}
\]

Proposition 1 guarantees the existence of the quantity \( \lambda \), as determined by the market. Besides the condition imposed by Proposition 1, we assume that \( \lambda \) is such that the process defined by

\[
Z_t := e^{-\int_0^t \lambda_s \ud W_s - \frac{1}{2} \int_0^t \lambda_s^2 \ud s}
\]

is a \( P \)-martingale. In practice, \( \lambda \) is frequently taken to be a constant.

If \( \lambda \) is such that Eq. (11) is a martingale, by Girsanov’s theorem it follows that \( \bar{W}_t = W_t + \int_0^t \lambda_s \ud s \) is a standard \( Q \)-Brownian motion. Under the new measure \( Q \), the dynamics of the short rate is given by

\[
\ud r_t = (\kappa - \lambda \theta) \ud t + \theta \ud \bar{W}_t
\]

### 3. AFFINE MODELS

#### 3.1. General Case

In the literature, there are many different models for the dynamics of the short rate. The most popular ones are the so-called affine models, due to their pleasing properties from analytical and computational points of view. These models are characterized by the assumption that the short rate is an affine function of a vector of unobserved state variables \( v_t = (v_1(t), \ldots, v_N(t)) \). Specifically, it is assumed that

\[
r_t = \delta_0 + \sum_{i=1}^N \delta_i v_i(t)
\]
where the vector $v_t$ follows an “affine diffusion,”

$$d v_t = \alpha(v_t) dt + \beta(v_t) dW_t$$

With the proper choice of $\alpha$ and $\beta$, corresponding to the affine family, one obtains that the price at time $t$ of a zero-coupon bond with maturity $T$ can be written as

$$P(t, T) = e^{A(T-t) - B(T-t)'v_t}$$

where $T$ means transpose, and $A$ and $B$ are obtained as solutions of a set of ordinary differential equations (see Duffie & Kan, 1996; Dai & Singleton, 2000). The important point is that the yield curve of these models is affine in the state variables, where we define the yield curve in Eq. (24) below.

3.2. The Vasicek Model

Among the affine class, one of the most popular ways of modeling the short rate is the Vasicek model, in which the short rate is considered to be a Gaussian process that satisfies the following stochastic differential equation:

$$d r_t = a(r_{\infty} - r_t) dt + \sigma dW_t$$

where $a$, $r_{\infty}$, and $\sigma$ are constants, and $W_t$ is a standard Brownian motion. In this case, the only state variable is the “unobserved” short rate. The drawback of this model is that, due to the Gaussian nature of the $r_t$ process, there is a positive probability that the short rate is negative, which is unreasonable from a financial point of view. Despite this drawback, the Vasicek model is frequently used because it allows explicit computations and many results can be obtained in closed form, making it easier to highlight the important points of further analysis.

Our first goal is to price zero-coupon bonds. Following the analysis of the previous section, we let $Q$ be defined by

$$\frac{dQ}{dP} := e^{-\lambda W_T - \frac{\lambda^2}{2} T}$$

where we assume the market price of risk, $\lambda$, to be a constant. Under this measure, the dynamics of the short rate is given by the Ornstein–Uhlenbeck process

$$d r_t = a(r^* - r_t) dt + \sigma d\tilde{W}_t$$

(14)
where \( r^* = r_\infty - (\lambda \sigma / a) \), and \( \tilde{W}_t = W_t + \lambda t \) is a standard \( \mathbb{Q} \)-Brownian motion.

Let \( B_V(t, x; T, \sigma, r^*) \) denote the Vasicek price at time \( t \) of a zero-coupon bond with maturity \( T \) when \( r_t = x \). By Theorem 1, this no-arbitrage price is determined by the term structure equation:

\[
\begin{align*}
\mathcal{L}_V(\sigma, r^*)B_V &= 0 \\
B_V(T, x; T, \sigma, r^*) &= 1
\end{align*}
\]  

(15)

where the Vasicek operator for the parameters \( \sigma \) and \( r^* \) is given by

\[
\mathcal{L}_V(\sigma, r^*):= \frac{\partial}{\partial t} + \frac{1}{2} \sigma^2 \frac{\partial^2}{\partial x^2} + a(r^* - x) \frac{\partial}{\partial x} - x
\]  

(16)

Note that this operator depends on \( \lambda \) through \( r^* = r_\infty - (\lambda \sigma / a) \). Let \( \tau = T - t \) be the time to maturity. Trying a solution of the form

\[
B_V(t, x; T, \sigma, r^*) = A(\tau)e^{-B(\tau)x}
\]  

(17)

we get that \( A \) and \( B \) satisfy the following ordinary differential equations

\[
\dot{A} = \frac{1}{2} \sigma^2 B^2 - a r^* B
\]  

(18)

\[
\dot{B} = 1 - a B
\]  

(19)

with initial conditions \( A(0) = 1 \), and \( B(0) = 0 \), and where the dot means differentiation with respect to \( \tau \). Solving these differential equations we get that

\[
B(\tau) = \frac{1}{a} (1 - e^{-a \tau})
\]  

(20)

and

\[
A(\tau) = e^{-R_\infty \tau + R_\infty (1/a)(1-e^{-a \tau}) - (\sigma^2/4a^3)(1-e^{-a \tau})^2}
\]  

(21)

where

\[
R_\infty = r^* - \frac{\sigma^2}{2a^2}
\]  

(22)

Hence, the Vasicek bond price is given by

\[
B_V(t, r_t; T, \sigma, r^*) = e^{-\{R_\infty \tau + (r_t - R_\infty)B + (\sigma^2/4a^3)B^2\}}
\]  

(23)

For fixed \( t \), the graph of \( P(t, T) \) as a function of \( T \) is called the bond price curve at time \( t \).
3.3. The Yield Curve

If at time $t$ we buy a zero-coupon bond with maturity $T$, the continuously compounded return on this investment, which we denote $R(t, T)$, is obtained from

$$ P(t, T) e^{(T-t)R} = 1 $$

This quantity $R$, which gives us the “internal rate of return” of the bond, is called the yield and it plays an important role in interest rate markets. At time $t$ one would be indifferent to buy the $T$-bond or to invest the amount $P(t, T)$ during the period $[t, T]$ at the rate $R(t, T)$. If we keep $t$ fixed and we let maturity vary, we obtain useful information about the interest rate market, namely, we get an idea of what the market thinks about the future evolution of interest rates.

The continuously compounded zero-coupon yield $R(t, T)$ is given by

$$ R(t; T) = \frac{\log P(t, T)}{T - t} $$  \hspace{1cm} (24)  

For fixed $t$, the graph of $R(t, T)$ is called the yield curve, or the term structure at $t$.

The yield for the Vasicek model is given by

$$ R_V(t, T) = R_{\infty} + (r_t - R_{\infty}) \frac{B}{T - t} + \frac{\sigma^2}{4a} \frac{B^2}{T - t} $$ \hspace{1cm} (25)  

which is an affine function of the short rate $r_t$. In Fig. 1, we show the Vasicek bond price (as a function of maturity) and the corresponding yield curve for some specific values of the parameters. We can see how the mean reverting property of the Ornstein–Uhlenbeck process brings the yield back to its long-term value, $R_{\infty}$.

4. THE VASICEK MODEL WITH STOCHASTIC VOLATILITY

In the case of the Black–Scholes model, typical historical data of the standard deviation of returns indicate that the volatility is not constant. The distributions of returns are not normal (they show fat tails), moreover, one can observe a smile effect in the implied volatility and similarly for the bond market (Andersen & Lund, 1997; Brennen, Harjes, & Kroner, 1996).
Fig. 2 shows two different paths of a non-constant volatility. In the top figure, the volatility is low (under 14%) for the first 17 years, and then it is high for the rest of the time. In the bottom figure, the volatility is high for several months, and then low for a similar period. Then high again, and so forth. The second path exhibits volatility clustering, the tendency of volatility to come in rapid bursts. This burstiness is closely related to mean reversion. We want to incorporate this type of volatility variations schematically in our modeling. The analysis of such an extended model will inform us about the significance of volatility heterogeneity and the shortcoming of the constant parameter model. As we show below we can do so in a robust way: Essentially, the presence of such volatility time scales is what is important, not their detailed modeling.

We now introduce a stochastic volatility model as follows. Let

$$\sigma_t := f(Y_t, Z_t)$$

where $f$ is a smooth bounded positive function, bounded away from zero, and $Y_t$ and $Z_t$ are two diffusion processes that vary, respectively, on a fast time
scale and on a slow time scale. Under the subjective probability measure $\mathcal{P}$, the short rate follows the stochastic differential equation

$$
\mathrm{d} r_t = a(r_\infty - r_t)\mathrm{d}t + \sigma_i \mathrm{d}W^0_t
$$

where $a$ and $r_\infty$ are constants, and $W^0_t$ is a standard Brownian motion.

4.1. The Fast Scale Volatility Factor

We choose the first factor driving the volatility, $Y_t$, to be a fast mean reverting Ornstein–Uhlenbeck process

$$
\mathrm{d} Y_t = \alpha(m - Y_t)\mathrm{d}t + \beta \mathrm{d}W^1_t
$$

where $\alpha$, $m$, and $\beta$ are constants, and $W^1_t$ is a standard Brownian motion whose covariation with $W^0_t$ is given by

$$
\mathrm{d}[W^0_t, W^1_t] = \rho_1 \mathrm{d}t
$$

We assume $\rho_1$ to be constant and $|\rho_1|<1$. The process $\{Y_t\}_{t \geq 0}$ is an ergodic process whose invariant distribution is $\mathcal{N}(m, \nu^2)$, with $\nu^2 = \beta^2/2\alpha$. 
Under the invariant distribution, the covariance is given by

\[ E[(Y_t - m)(Y_s - m)] = v^2 e^{-2|t-s|} \]

which shows that the exponential rate of decorrelation of \( \{Y_t\} \) is \( \alpha \). Hence, \( 1/\alpha \) can be thought of as the typical correlation time. The parameter \( v^2 \) controls the size of the fluctuations in the volatility associated with variations in \( Y_t \). We assume that \( v^2 \) is constant and consider a regime of fast mean reversion or \( \alpha \) large (i.e., \( \beta = O(\sqrt{\alpha}) \)). Increasing \( \alpha \) and keeping \( v \) fixed changes the degree of burstiness of the volatility without affecting the magnitude of the fluctuations (see Fouque et al., 2000, for more details).

Define \( \varepsilon = 1/\alpha \). Then, the small parameter \( \varepsilon \) can be interpreted as the mean reversion time of the volatility associated with fluctuations in \( Y_t \). The asymptotic analysis that we introduce in Section 5.2 is then for the case \( \varepsilon \downarrow 0 \), with \( v^2 \) fixed, with \( \beta = v\sqrt{2/\varepsilon} \).

### 4.2. The Slow Scale Volatility Factor

We choose the second factor, \( Z_t \), to follow the stochastic differential equation

\[ dZ_t = \delta c(Z_t)dt + \sqrt{\delta}g(Z_t)dW^2_t \]  \tag{28}  

where \( \delta \) is a small parameter, \( W^2_t \) is a standard Brownian motion, and we assume that the functions \( c(\cdot) \) and \( g(\cdot) \) are smooth and at most linearly growing at infinity. As the parameter \( \delta \) is assumed to be small, this makes \( Z_t \) vary on a slow scale: namely, \( Z_t \) varies on the \( O(1/\delta) \) scale. Note that we now have three relevant time scales:

1. The \( O(1) \) scale, which is the time-to-maturity scale (\( T \)).
2. The slow scale or \( O(1/\delta) \) scale, which is the characteristic time scale of the process \( Z_t \) (\( T < (1/\delta) \)).
3. The fast scale or \( O(\varepsilon) \) scale, which is the mean reversion time of the process \( Y_t \) (\( \varepsilon < T \)).

### 4.3. The Model Under the Risk-Neutral Measure

The introduction of two new sources of randomness gives rise to a family of equivalent martingale measures that will be parameterized by the market price of risk, \( \lambda \), and two market prices of volatility risk, which we denote by \( \gamma \) and \( \xi \), associated respectively with \( Y_t \) and \( Z_t \). All these market prices
are not determined within the model, but are fixed exogenously by the market.

We now assume that these market prices of risk do not depend on the short rate: that is, they have the form \( \lambda(Y_t, Z_t), \gamma(Y_t, Z_t), \) and \( \zeta(Y_t, Z_t) \). We also assume that \( \lambda, \gamma, \) and \( \zeta \) are smooth bounded functions.

We define the new equivalent martingale measure, \( Q \), by

\[
\frac{dQ}{dP} = e^{-\int_0^T H_t dB_t - \frac{1}{2} \int_0^T \|H_t\|^2 dt}
\]

where \( H_t = (\lambda(Y_t, Z_t), \gamma(Y_t, Z_t), \zeta(Y_t, Z_t)) \) and \( B_t = (B^0_t, B^1_t, B^2_t) \). By Girsanov’s Theorem, \( B^*_t = B_t + \int_0^T H_s ds \) is a \( Q \)-Brownian motion, with \( B^*_t = (B^0*_t, B^1*_t, B^2*_t) \). We now define \( (W^0*_t, W^1*_t, W^2*_t) \) with the following correlation structure:

\[
\begin{pmatrix} W^0*_t \\ W^1*_t \\ W^2*_t \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ \rho_1 & \sqrt{1 - \rho^2_1} & 0 \\ \rho_2 & \tilde{\rho}_{12} & \sqrt{1 - \rho^2_2 - \tilde{\rho}^2_{12}} \end{pmatrix} \begin{pmatrix} B^0*_t \\ B^1*_t \\ B^2*_t \end{pmatrix}
\]

(29)

where we assume \( |\rho_1| < 1 \) and \( \rho^2_2 + \tilde{\rho}^2_{12} < 1 \). Note that with this structure \( d[W^1_t, W^2_t] = \rho_{12} dt \), with \( \rho_{12} := \rho_1 \rho_2 + \tilde{\rho}_{12} \sqrt{1 - \rho^2_1} \). Under the risk-neutral measure, our model is therefore:

\[
dr_t = (a(r_\infty - r_t) - \lambda_t f_t) dt + f_t dW^0*_t
\]

(30)

\[
dY_t = \left( \frac{1}{e}(m - Y_t) - \frac{1}{\sqrt{e}} \Lambda_t \right) dt + \frac{1}{\sqrt{e}} dW^1*_t
\]

(31)

\[
dZ_t = \left( \delta c(Z_t) - \sqrt{\delta g(Z_t)} \right) \Gamma_t dt + \sqrt{\delta g(Z_t)} dW^2*_t
\]

(32)

where we have used \( \lambda_t \) for \( \lambda(Y_t, Z_t) \), and the analogous convention for \( \gamma_t, \zeta_t, f_t, \Lambda_t, \) and \( \Gamma_t \), together with:

\[
\Lambda(y, z) := \rho_1 \lambda(y, z) + \gamma(y, z) \sqrt{1 - \rho^2_1}
\]

(33)

\[
\Gamma(y, z) := \rho_2 \lambda(y, z) + \tilde{\rho}_{12} \gamma(y, z) + \zeta(y, z) \sqrt{1 - \rho^2_2 - \tilde{\rho}^2_{12}}
\]

(34)
5. THE BOND PRICE WITH STOCHASTIC VOLATILITY

In this section we present the asymptotic analysis for the bond price, and we find an approximation to the price that accounts for stochastic volatility. In order to emphasize the dependence of the approximation on the small parameters $\varepsilon$ and $\delta$, we will denote the no-arbitrage Vasicek price of a zero-coupon bond with maturity $T$ by $P^{\varepsilon,\delta}(t, x, y, z; T)$. Using the probabilistic representation, this price is given by

$$P^{\varepsilon,\delta}(t, x, y, z; T) = E^Q \left[ e^{-\int_t^T r_s ds} | r_t = x, Y_t = y, Z_t = z \right]$$ (35)

Because in the Vasicek model $r_t$ can take any negative value with positive probability, the expectation in Eq. (35) is not trivially finite. One can prove that the expectation is indeed finite by showing that $r_t$ and $\int_0^t r_s ds$ have exponential moments (see Cotton et al., 2004; De Santiago, 2008, Section 6.1).

5.1. The Bond Price Expansion

An application of Feynman–Kac’s result to Eq. (35) shows that $P^{\varepsilon,\delta}$ is a solution of the following problem

$$\begin{align*}
\mathcal{L}^{\varepsilon,\delta} P^{\varepsilon,\delta} &= 0 \\
P^{\varepsilon,\delta}(T, x, y, z; T) &= 1
\end{align*}$$ (36)

where the operator $\mathcal{L}^{\varepsilon,\delta}$ is given by

$$\mathcal{L}^{\varepsilon,\delta} = \frac{1}{\varepsilon} L_0 + \frac{1}{\sqrt{\varepsilon}} L_1 + L_2 + \sqrt{\delta} \mathcal{M}_1 + \delta \mathcal{M}_2 + \sqrt{\varepsilon} \mathcal{M}_3$$ (37)

with

$$L_0 = (m - y) \frac{\partial}{\partial y} + v^2 \frac{\partial^2}{\partial y^2}$$ (38)

$$L_1 = v \sqrt{2} \left( \rho_{1f}(y, z) \frac{\partial^2}{\partial x \partial y} - \Lambda(y, z) \frac{\partial}{\partial y} \right)$$ (39)
\[ \mathcal{L}_2 = \p{t} + \frac{1}{2} f^2(y, z) \p[2]{x} + (a r_\infty - x) - \lambda(y, z) f(y, z) \frac{\p{x}}{} - x \quad (40) \]

\[ \mathcal{M}_1 = \rho_2 f(y, z) g(z) \p[2]{x} - g(z) \Gamma(y, z) \frac{\p{}}{z} \quad (41) \]

\[ \mathcal{M}_2 = c(z) \frac{\p{}}{z} + \frac{1}{2} g^2(z) \p[2]{z} \quad (42) \]

\[ \mathcal{M}_3 = \sqrt{2} \rho_{12} g(z) \p[2]{y} \quad (43) \]

If we fix \( y \) and \( z \), and we let \( \sigma = f(y, z) \) and \( r^* = r_\infty - (\lambda(y, z) f(y, z)/a) \), we can write Eq. (40) as

\[ \mathcal{L}_2 = \p{t} + \frac{1}{2} \sigma^2 \p[2]{x} + a(r^* - x) \frac{\p{x}}{} - x \]

which is the Vasicek operator defined in Eq. (16). That is, \( \mathcal{L}_2 \equiv \mathcal{L}_V(\sigma, r^*) \).

The small parameter \( \varepsilon \) gives rise to a singular perturbation problem. In the limit when \( \varepsilon \) goes to zero, the leading problem becomes the Poisson equation associated with the operator \( \mathcal{L}_0 \) rather than the Vasicek problem. The terms associated only with the small parameter \( \delta \) give rise to a regular perturbation problem about the Vasicek operator \( \mathcal{L}_2 \). In the following sections, we carry out the combined regular and singular perturbation expansion.

In order to carry out the asymptotic analysis, we begin writing \( P^{\varepsilon, \delta} \) in powers of \( \sqrt{\delta} \):

\[ P^{\varepsilon, \delta} = P_0^{\varepsilon} + \sqrt{\delta} P_1^{\varepsilon} + \delta P_2^{\varepsilon} + \cdots \quad (44) \]

Substituting Eq. (44) in the PDE (36), and considering the \( O(1) \) terms (with respect to \( \delta \)) and the \( O(\sqrt{\delta}) \) terms, we define the problems that will determine \( P_0^{\varepsilon} \) and \( P_1^{\varepsilon} \).

**Definition 1.** The leading order term \( P_0^{\varepsilon} \) is defined as the unique solution to

\[ \left( \frac{1}{\varepsilon} \mathcal{L}_0 + \frac{1}{\sqrt{\varepsilon}} \mathcal{L}_1 + \mathcal{L}_2 \right) P_0^{\varepsilon} = 0 \quad (45) \]

\[ P_0^{\varepsilon}(T, x, y, z; T) = 1 \quad (46) \]
Definition 2. The term $P^e_1$ is defined as the unique solution to the problem
\[
\left( \frac{1}{\varepsilon} L_0 + \frac{1}{\sqrt{\varepsilon}} L_1 + L_2 \right) P^e_1 = -\left( M_1 + \frac{1}{\sqrt{\varepsilon}} M_3 \right) P^e_0
\]
\[ P^e_1(T, x, y, z; T) = 0 \] (47)
\[ P^e_1(T, x, y, z; T) = 0 \] (48)

In the following sections, we expand $P^e_0$ and $P^e_1$ in powers of $\sqrt{\varepsilon}$ to obtain the approximation to the price, $P^{e, \delta}$.

5.2. The Fast Scale Correction

First, we expand $P^e_0$ as
\[
P^e_0 = P_0 + \sqrt{\varepsilon} P_{1,0} + \varepsilon P_{2,0} + \varepsilon^{3/2} P_{3,0} + \cdots
\] (49)

In order to find explicit expressions for $P_0$ and $P_{1,0}$, we insert Eq. (49) into Eq. (45). From the $\mathcal{O}(1/\varepsilon)$ and $\mathcal{O}(1/\sqrt{\varepsilon})$ terms we obtain that $P_0$ and $P_{1,0}$ do not depend on $y$: hence, we can write $P_0 = P_0(t, x, z)$, and $P_{1,0} = P_{1,0}(t, x, z)$. The $\mathcal{O}(1)$ terms give that $P_0$ is determined by the problem:
\[
\begin{cases}
\langle L_V \rangle P_0 & = 0 \\
P_0(T, x, z) & = 1
\end{cases}
\] (50)

where the bracket notation $\langle \cdot \rangle$ means integration with respect to the invariant distribution of the $Y$ process (i.e., integration with respect to a normal $\mathcal{N}(m, v^2)$). That is, $\langle L_V \rangle$ is the Vasicek operator $L_V(\tilde{\sigma}, \tilde{r}^*)$ with parameters
\[
\tilde{\sigma}(z) := \sqrt{\langle f^2(\cdot, z) \rangle}, \quad \tilde{r}^*(z) := r_\infty - \frac{\langle \lambda(\cdot, z)f(\cdot, z) \rangle}{a}
\] (51)

Hence, we have that
\[
P_0(t, x, z) = B_V(t, x; T, \tilde{\sigma}(z), \tilde{r}^*(z))
\] (52)

where $B_V$ was defined in Eq. (17). This is the constant volatility Vasicek price evaluated at effective parameters $\tilde{\sigma}(z)$ and $\tilde{r}^*(z)$, which result from averaging with respect to the fast variable, and from “freezing” the slow factor at its current level $z$. The explicit form of $P_0$ is given by
\[
P_0(t, x, z) = A(\tau, z)e^{-B(\tau)x}
\] (53)
where $\tau = T - t$ is the time to maturity, and the expressions for $A(\tau, z)$ and $B(\tau)$ are now given by:

$$\begin{align*}
A(\tau, z) &= e^{-R_\infty(z)\tau + R_\infty(z)(1/a)(1 - e^{-at})} \frac{\hat{\sigma}^2(z)}{2a^2} (1 - e^{-at})^2 \\
B(\tau) &= \frac{1}{a} (1 - e^{-at})
\end{align*}$$

(54)

(55)

with

$$R_\infty(z) = \tilde{r}^*(z) - \frac{\hat{\sigma}^2(z)}{2a^2}$$

(56)

Define now $\tilde{P}_{1,0} = \sqrt{\varepsilon} P_{1,0}$, and the operator

$$\lambda := \sqrt{\varepsilon}(L_1 L_0^{-1}(L_2 - (L_2)))$$

(57)

The $O(\sqrt{\varepsilon})$ terms give that $\tilde{P}_{1,0}$ is determined by the problem:

$$\begin{align*}
\langle L_2 \rangle \tilde{P}_{1,0} &= \lambda P_0 \\
\tilde{P}_{1,0}(T, x, z) &= 0
\end{align*}$$

(58)

In order to obtain an expression for the operator $\lambda$ we introduce $\phi(y, z)$ and $\psi(y, z)$, solutions of the following Poisson equations with respect to $y$:

$$L_0 \phi(y, z) = f^2(y, z) - \hat{\sigma}^2(z)$$

(59)

$$L_0 \psi(y, z) = \dot{\lambda}(y, z) f(y, z) - \langle \dot{\lambda}(\cdot, z) f(\cdot, z) \rangle$$

(60)

Both $\phi$ and $\psi$ are defined up to an additive function that does not depend on $y$, and that will not affect $\lambda$ since the operator $L_1$ (which is included in $\lambda$) takes derivatives with respect to $y$. We then have

$$L_0^{-1}(L_2 - (L_2)) = \frac{1}{2} \phi(y, z) \frac{\partial^2}{\partial x^2} - \psi(y, z) \frac{\partial}{\partial x}$$

and therefore

$$L_1 L_0^{-1}(L_2 - (L_2))$$

$$= \sqrt{2} \left\{ \rho_1 f(y, z) \frac{\partial^2}{\partial x \partial y} \left( \frac{1}{2} \phi \frac{\partial^2}{\partial x^2} - \psi \frac{\partial}{\partial x} \right) - \Lambda \frac{\partial}{\partial y} \left( \frac{1}{2} \phi \frac{\partial^2}{\partial x^2} - \psi \frac{\partial}{\partial x} \right) \right\}$$
Expanding, and using the fact that $P_0(t, x, z)$ does not depend on $y$, we have that

\[
\mathcal{A} P_0 = \frac{v \sqrt{\varepsilon}}{\sqrt{2}} \rho_1 \langle f(\cdot, z)\phi_y(\cdot, z) \rangle \frac{\partial^3 P_0}{\partial x^3} + \frac{v \sqrt{\varepsilon}}{\sqrt{2}} \langle (\Lambda(\cdot, z)\phi_y(\cdot, z)) + 2 \rho_1 \langle f(\cdot, z)\psi_y(\cdot, z) \rangle \rangle \frac{\partial^2 P_0}{\partial x^2} + \frac{v \sqrt{2\varepsilon}}{2} \langle (\Lambda(\cdot, z)\psi_y(\cdot, z)) \rangle \frac{\partial P_0}{\partial x} \]

where $\phi_y$ represents the partial derivative of $\phi$ with respect to $y$, and the analogous convention holds for $\psi_y$. If we now let

\[
V_1^c(z) := \frac{v \sqrt{2\varepsilon}}{2} (\Lambda\psi_y)
\]

\[
V_2^c(z) := -\frac{v \sqrt{\varepsilon}}{\sqrt{2}} \langle (\Lambda\phi_y) + 2 \rho_1 \langle f\psi_y \rangle \rangle
\]

\[
V_3^c(z) := \frac{v \sqrt{\varepsilon}}{\sqrt{2}} \rho_1 \langle f\phi_y \rangle
\]

then we can write

\[
\mathcal{A} = V_1^c \frac{\partial}{\partial x} + V_2^c \frac{\partial^2}{\partial x^2} + V_3^c \frac{\partial^3}{\partial x^3}
\]

The problem for $\tilde{P}_{1,0}$ then becomes

\[
\begin{align*}
\langle \mathcal{L}_2 \rangle \tilde{P}_{1,0} &= V_1^c \frac{\partial P_0}{\partial x} + V_2^c \frac{\partial^2 P_0}{\partial x^2} + V_3^c \frac{\partial^3 P_0}{\partial x^3} \\
\tilde{P}_{1,0}(T, x, z) &= 0
\end{align*}
\]

The problem for $\tilde{P}_{1,0}$ can easily be solved explicitly, to give a representation in the form:

\[
\tilde{P}_{1,0} = D^c(\tau, z)A(\tau, z)e^{-B(\tau)x}
\]

(see Cotton et al., 2004; De Santiago, 2008). However, before we give the explicit form we shall in Section 6 carry out a group parameter reduction that will simplify the representation of the solution.
5.3. The Slow Scale Correction

Let us now expand $P_1^e$ (the second term on the expansion (44)) in terms of the small parameter $\varepsilon$,

$$P_1^e = P_{0,1} + \sqrt{\varepsilon} P_{1,1} + \varepsilon P_{2,1} + \varepsilon^{3/2} P_{3,1} + \cdots \quad (68)$$

and substitute this expression in the terminal value problem that defines $P_1^e$ (Eq. (47)). The leading order term gives that $P_{0,1}$ does not depend on $y$, that is, $P_{0,1} = P_{0,1}(t, x, z)$.

Define $\tilde{P}_{0,1} = \sqrt{\delta} P_{0,1}$ and $\tilde{\mathcal{M}}_1 = \sqrt{\delta} \mathcal{M}_1$. From the $O(1)$ terms we obtain that the problem that determines $\tilde{P}_{0,1}$ is

$$\begin{cases}
(L_2) \tilde{P}_{0,1} &= - (\tilde{\mathcal{M}}_1) P_0 \\
\tilde{P}_{0,1}(T, x, z) &= 0 
\end{cases} \quad (69)$$

where we recall that $P_0 = A(\tau, z)e^{-\mu(\tau)x}$. Let

$$V_{0}^{\delta}(z) = -\sqrt{\delta} g(z)(\Gamma(\cdot, z)) \quad (70)$$

$$V_{1}^{\delta}(z) = \sqrt{\delta} \rho_2 g(z)(f(\cdot, z)) \quad (71)$$

and recall that $\mathcal{M}_1$ is given in Eq. (41). Then we can write $\langle \tilde{\mathcal{M}}_1 \rangle$ as

$$\langle \tilde{\mathcal{M}}_1 \rangle = \frac{\partial}{\partial z} V_{0}^{\delta} + \frac{\partial^2}{\partial x \partial z} V_{1}^{\delta}$$

Again, the problem for $\tilde{P}_{0,1}$ can easily be solved explicitly, to give a representation in the form:

$$\tilde{P}_{0,1} = D(\tau, z) A(\tau, z)e^{-\mu(\tau)x} \quad (72)$$

(see De Santiago, 2008). However, in Section 6 we carry out a group parameter reduction that will simplify the representation of this solution. Before we go into the details of the group parameter transformation, we present in the next section a numerical illustration of some typical corrections to the bond and yield that derives from our multiscale model.
5.4. The Bond Price Approximation

We define the bond price approximation as

\[ \tilde{P}^{e,\delta} := P_0 + \sqrt{\varepsilon}P_{1,0} + \sqrt{\delta}P_{0,1} = P_0 + \tilde{P}_{1,0} + \tilde{P}_{0,1} \]

From Eqs. (67), (72), and (53) it follows that we can write the approximation explicitly as

\[ \tilde{P}^{e,\delta} = (1 + D^e + D^\delta)A e^{-Bx} \]  \hspace{1cm} (73)

where \( D^e \) and \( D^\delta \) are \( O(\sqrt{\varepsilon}) \) and \( O(\sqrt{\delta}) \), respectively, and \( A \) and \( B \) are defined in Eqs. (54) and (55). The corresponding approximation to the yield curve, \( \tilde{R}^{e,\delta} = -\frac{1}{\tau} \log \tilde{P}^{e,\delta} \), is given by

\[ \tilde{R}^{e,\delta} = R_0 - \frac{1}{\tau} \log(1 + D^e + D^\delta) \approx R_0 - \frac{1}{\tau} (D^e + D^\delta) \]  \hspace{1cm} (74)

where \( R_0 = -\frac{1}{\tau} \log P_0 \) is the yield corresponding to the constant volatility price \( P_0 \).

The influence of the corrections \( D^e \) and \( D^\delta \) will affect the shape of the yield curve, so we expect a richer variety of shapes.

In the following figures we use the values of the parameters \( a = 1, \tilde{\sigma} = 0.1, \) and \( \tilde{\rho} = 0.1 \) and the initial value \( x = 0.07 \) and show results corresponding to the approximation derived above. If we assume that only the fast scale process \( Y_t \) has influence on \( \sigma_t \) (i.e., we assume \( \delta = 0 \)), we obtain the graphics in Fig. 3, where the prices and yields are computed for \( \varepsilon = 0.01 \). The yield curve is increasing for very short maturities, it becomes decreasing for medium-range maturities (from 1.5 to 9 years) and then it becomes slowly increasing again.

In Fig. 4, we assume that only the slow scale process affects the volatility (i.e., we take \( \varepsilon = 0 \)). The yield curve is increasing up to 6 years and it decreases for longer maturities. Fig. 5 shows the case when the volatility is driven by both processes, \( Y_t \) and \( Z_t \). As one would expect, it seems that the slow scale seems to have a greater impact on the long range, while the fast scale seems to affect the medium-maturity yields. In Fig. 6 we can see that the larger the value of \( \delta \), the more pronounced is the influence of the slow scale volatility.
6. GROUP PARAMETER REDUCTION

The leading order bond price, \( P_0 \), depends on the parameters \( a, \tilde{\sigma}(z) \), and \( \tilde{r}^* (z) \), which are those that define the operator \( \mathcal{L}_V (\tilde{\sigma}, \tilde{r}^*) \) (see Eq. (51)). The first-order corrections, \( \tilde{P}_{1,0} \) and \( \tilde{P}_{0,1} \) depend in particular also on the \( z \)-dependent group parameters

\[
V_1^e, V_2^e, V_3^e, V_0^\delta, V_1^\delta
\]  

If we define the price correction

\[
P_{c}^{e,\delta} := \sqrt{e}P_{1,0} + \sqrt{\delta}P_{0,1}
\]  

then the problem characterizing \( P_{c}^{e,\delta} \) can be written as

\[
\begin{cases}
\mathcal{L}_V (\tilde{\sigma}, \tilde{r}^*) P_{c}^{e,\delta} = -\mathcal{H}^{e,\delta} P_0 \\
P_{c}^{e,\delta} (T, x, z; T) = 0
\end{cases}
\]  

Fig. 3. Bond Price and Yield Curves with Constant Volatility (Thin Line), and Fast-Scale Stochastic Volatility (Dashed Line). For This Example, \( \varepsilon = 0.001 \).
where in the source term we have used the notation

\[ \mathcal{H}^{\varepsilon, \delta} = \sum_{k=1}^{3} V^k_k \frac{\partial^k}{\partial x^k} + \left( V^0_0 \frac{\partial}{\partial z} + V^1_1 \frac{\partial^2}{\partial z \partial x} \right) \]  

(78)

Note that the price approximation does not depend on \( y \), and that \( z \) is a fixed parameter (obtained by “freezing” the slow factor at its current level). In this section, we discuss how to effectively reduce the number of degrees of freedom.

If we now define

\[ m_1 := \frac{\partial R_\infty(z)}{\partial z} \]

\[ m_2 := \frac{\partial (\sigma^2(z)/4a)}{\partial z} \]
by making use of Eqs. (53) and (54), we can write the source operator in terms of the “Greeks” as

$$-\mathcal{H}_V^{\varepsilon, \delta} P_0 = -\left\{ U_1 \frac{\partial}{\partial \chi} + U_2 \frac{\partial^2}{\partial \chi^2} + U_3 \frac{\partial^3}{\partial \chi^3} - \tau m_1 \left( V_0^\delta + V_1^\delta \frac{\partial}{\partial \chi} \right) \right\} P_0 \quad (79)$$

with

$$U_1 = V_1^\varepsilon - m_1 V_0^\delta, \quad U_2 = V_2^\varepsilon - m_2 V_0^\delta - m_1 V_1^\delta, \quad U_3 = V_3^\varepsilon - m_2 V_1^\delta$$

If we now define

$$\sigma^*(z) = \sqrt{\sigma^2(z) + 2U_2(z)} \quad (80)$$

$$r^{**}(z) = r^*(z) + \frac{U_1(z)}{a} \quad (81)$$

*Fig. 5.* Bond Price and Yield Curves with Constant Volatility (Thin Line), Fast-Scale Stochastic Volatility (Dashed Line) and Two-Factor Stochastic Volatility (Dotted Line). For This Example, $\varepsilon = 0.001, \delta = 0.01$. 
we have that
\[ \mathcal{L}_V(\sigma^*, r^{**}) = \mathcal{L}_V(\bar{\sigma}, \bar{r}^*) + U_2 \frac{\partial^2}{\partial x^2} + U_1 \frac{\partial}{\partial x} \]
where we have dropped the \( z \) dependence. Therefore, using Eqs. (50), (77),
and the fact that \( \bar{P}^{\varepsilon, \delta} = P_0 + P_{c}^{\varepsilon, \delta} \), we get that
\[ \mathcal{L}_V(\sigma^*, r^{**}) \bar{P}^{\varepsilon, \delta} = U_1 \frac{\partial P_{c}^{\varepsilon, \delta}}{\partial x} + U_2 \frac{\partial^2 P_{c}^{\varepsilon, \delta}}{\partial x^2} - U_3 \frac{\partial^3 P_0}{\partial x^3} + \tau m_1 \left( V^\delta_0 + V^\delta_1 \frac{\partial}{\partial x} \right) P_0 \]
where the source is very similar to Eq. (79), except for the first two terms. Note that the source terms
\[ U_1 \frac{\partial P_{c}^{\varepsilon, \delta}}{\partial x} \quad \text{and} \quad U_2 \frac{\partial^2 P_{c}^{\varepsilon, \delta}}{\partial x^2} \]
are of order \( O(\varepsilon + \delta) \), since \( V^\varepsilon_1, V^\varepsilon_2 \), and \( \bar{P}_{1,0} \) are all of order \( O(\sqrt{\varepsilon}) \), and \( V^\delta_0, V^\delta_1 \), and \( \bar{P}_{1,0} \) are all of order \( O(\sqrt{\delta}) \). Thus, the first two terms of the source are negligible compared to the other source terms, and therefore the corrected price, \( P_0 + P_{c}^{\varepsilon, \delta} \), has the same order of accuracy as \( P_0^* + P_{c}^* \), where
$P^*_0$ satisfies
\[
\begin{align*}
\mathcal{L}_V(\sigma^*, r^{**})P^*_0 &= 0 \\
P^*_0(T, x; T) &= 1
\end{align*}
\]
and $P^*_c$ satisfies
\[
\begin{align*}
\mathcal{L}_V(\sigma^*, r^{**})P^*_c &= -\mathcal{H}_V^* P^*_0 \\
P^*_c(T, x; T) &= 0
\end{align*}
\]
with the new source operator
\[
\mathcal{H}_V^* = U_3 \frac{\partial^3}{\partial x^3} - \tau \left( W_0 + W_1 \frac{\partial}{\partial x} \right)
\]
where we defined
\[
W_0 = m_1 V_0^\delta, \quad W_1 = m_1 V_0^\delta
\] (82)

Note that the set of parameters that are to be calibrated is thus reduced to, first, the $O(1)$ parameters
\[
a, \quad r^{**}, \quad \sigma^*
\] (83)
and, in addition, the small parameters
\[
U_3, \quad W_0, \quad W_1
\] (84)
that are $O(\sqrt{\varepsilon} + \sqrt{\delta})$.

### 6.1. Yield Correction

From Eqs. (67) and (72) we know that the first-order correction to the bond price can be written as $P^{c*\delta} = (D^e + D^d)P_0$. Because $P^*_c$ has the same order of accuracy as $P^{c*\delta}$, we may write
\[
P^*_c = D^* P^*_0
\]
so that the corrected yield takes the form
\[
\tilde{R}^{c*\delta} = R^*_0 - \frac{1}{\tau} \log(1 + D^*) \approx R^*_0 - \frac{1}{\tau} D^*
\]
Here $R^*_0 = -(1/\tau) \log P^*_0$ is the yield corresponding to the constant volatility bond price $P^*_0$ evaluated at the effective short rate and volatility $(r^{**}, \sigma^*)$. 

Bond Markets with Stochastic Volatility

239
and

\[ D^*(\tau) = U_3g_1(\tau) + W_0g_2(\tau) + W_1g_3(\tau) \]  

(85)

with

\[ g_1(\tau) = \frac{B(\tau) - \tau}{a^3} + \frac{B^2(\tau)}{2a^2} + \frac{B^3(\tau)}{3a} \]

\[ g_2(\tau) = -\frac{\tau^2}{2} \]

\[ g_3(\tau) = \frac{\tau}{a^2} + \frac{\tau^2}{2a} - B(\tau) \left( \frac{\tau}{a} + \frac{1}{a^2} \right) \]

Note that in this parameterization the structure of the yield curve correction is not affected by the fast factor as long as the slow is present, only the interpretation of the parameters does. Observe also that only the temporal aspect of the yield curve is affected by the stochastic volatility modulation. The spatial part is still determined by the modulation \( xB(\tau)/\tau \) in the expression for the leading yield surface

\[ R_0(t, \tau; x, z) = R_\infty^* + (x - R_\infty^*) \frac{B(\tau)}{\tau} + \frac{(\sigma^*)^2 B^2(\tau)}{4a} \frac{\tau}{\tau} \]

where we defined

\[ R_\infty^* = r^{**}(z) - \frac{(\sigma^*)^2(z)}{2a^2} \]

Observe that in order to calibrate the parameters \( c_i, i = 1, 2, 3 \), we would need to regress the observed yield corrections relative to the constant volatility model against the yield curve terms factors \( g_i, i = 1, 2, 3 \). We comment on the calibration in more detail in the next section.

7. CALIBRATION OF THE MODEL

In this section we discuss one way of calibrating the model to market data. Assume for various \( t \) and \( T \) we can observe the yield \( R(t, T; x) \), the market yield at time \( t \) of a bond with maturity \( T \), and current short rate level \( x \) and time to maturity \( \tau = T - t \). To emphasize that it is an observed market yield we will write \( R^{obs}(t, T) \). We then seek to estimate

\[ a, \quad r^{**}, \quad \sigma^*, \quad U_3, \quad W_0, \quad W_1 \]
so that for the set of observed yields we have

\[ R_{\text{obs}}(t, T; x) \approx R_0^*(\tau, a, r^{**}, \sigma^*; x) - \frac{U_3 g_1(\tau) + W_0 g_2(\tau) + W_1 g_3(\tau)}{\tau} \]

where \( R_0^* \) corresponds to the constant parameter Vasicek yield as given in Eq. (25), but evaluated at the corrected parameters. This can, for instance, be accomplished by first fitting \( R_0^* \) to the data by estimating the \( O(1) \) parameters \( a, r^{**}, \) and \( \sigma^* \) via a least squares procedure to get a priori estimates of their leading values. The multiscale correction will affect the "wings" of the yield term structure relatively strongest, corresponding to small and large maturities. Thus, in the second step we include the correction terms and estimate also the small correction parameters in addition to an updated estimate of the parameters \( r^{**}, \sigma^* \). Note that when exploiting the a priori parameter estimates this second step actually becomes, in view of the form of Eq. (25), a linear least squares problem relative to a set of term structure factors defined in terms of the a priori estimates.

We conclude that the asymptotic framework provides a robust approach to parameterize the yield term structure; moreover, the parameters of the fitted term structure can be used in the pricing of related, potentially less liquid, contracts.

8. CONNECTION TO DEFAULT ABLE BONDS

We consider a particular obligor corresponding to an underlying name. The event that the obligor defaults is modeled in terms of the first arrival of a Poisson process with stochastic intensity or hazard rate \( \lambda^{(1)} \). Conditioned on the path of the hazard rate the probability that the obligor has survived till time \( T \) is thus \( \exp(-\int_0^T \lambda^{(1)} ds) \). The probability of survival till time \( T \) is then under the doubly stochastic framework

\[ E^Q\left[ e^{-\int_0^T \lambda^{(1)} ds} \right] \]

with the expectation taken with respect to the risk-neutral pricing measure so that this corresponds to the bond price expression (3). In the Vasicek setup, we model the intensity so that

\[ \lambda^{(1)}_t = X^{(1)}_t \]
is an Ornstein-Uhlenbeck process:

\[
dX^{(1)}_t = a_1 \left( X^*_1 - X^{(1)}_t \right) dt + \sigma_1 dW^{(1)}_t
\]  

(87)

Thus, we can re-interpret all our results on multiscale stochastic volatility bond prices as results on the survival probability in the context of a defaultable bond in the case where the underlying hazard rate process is modeled in terms of a multiscale stochastic volatility Vasicek process (see also Fouque, Sircar, & Solna, 2006; Papageorgiou & Sircar, 2008a). In the multiname case correlations in-between names is essential and there is an important “gearing” effect of stochastic volatility impact in terms of the size of the name portfolio regarding its joint survival probability (see Cotton, Fouque, Sircar, & Solna, 2008; Papageorgiou & Sircar, 2008b).

REFERENCES

TWO-DIMENSIONAL MARKOVIAN MODEL FOR DYNAMICS OF AGGREGATE CREDIT LOSS

Andrei V. Lopatin and Timur Misirpashaev

ABSTRACT

We propose a new model for the dynamics of the aggregate credit portfolio loss. The model is Markovian in two dimensions with the state variables being the total accumulated loss $L_t$ and the stochastic default intensity $\lambda_t$. The dynamics of the default intensity are governed by the equation $d\lambda_t = \kappa(\rho(L_t, t) - \lambda_t)dt + \sigma\sqrt{\lambda_t}dW_t$. The function $\rho$ depends both on time $t$ and accumulated loss $L_t$, providing sufficient freedom to calibrate the model to a generic distribution of loss. We develop a computationally efficient method for model calibration to the market of synthetic single tranche collateralized debt obligations (CDOs). The method is based on the Markovian projection technique which reduces the full model to a one-step Markov chain having the same marginal distributions of loss. We show that once the intensity function of the effective Markov chain consistent with the loss distribution implied by the tranches is found, the function $\rho$ can be recovered with a very moderate computational effort. Because our model is Markovian and has low dimensionality, it offers a convenient framework for the pricing of dynamic credit instruments, such as options on indices and tranches, by backward induction. We calibrate the model to a set of recent market data.
quotes on CDX index tranches and apply it to the pricing of tranche options.

1. INTRODUCTION

Synthetic collateralized debt obligations (CDOs) are derivatives of the aggregate loss sustained by the seller of the protection on a portfolio of credit default swaps. The majority of standard CDO tranches can be statically replicated by a set of long and short positions in more elementary instruments called stop-loss options. The payoff from a stop-loss option with maturity $t$ and strike $X$ is equal to $\max(L_t - X, 0)$, where $L_t$ is the loss accumulated in the underlying portfolio by the maturity time $t$.

This replication is not directly useful for hedging purposes because stand-alone stop-loss options are not currently traded. However, it is extremely useful in model-based valuation because of the simplicity of stop-loss options, which only depend on the distribution of loss at a single time horizon. An important consequence is that the value of an instrument replicated by a portfolio of stop-loss options only depends on the one-dimensional marginal distributions of the loss process and is insensitive to the dynamics of the temporal loss evolution. Therefore, it is possible to construct viable valuation models for synthetic CDO tranches by focusing solely on producing correct distributions of loss on a grid of relevant time horizons and ignoring the implied dynamics (or even leaving the dynamics undefined). Such models are often referred to as static. Most static models in active use today belong to the framework of factor models, reviewed by Andersen and Sidenius (2005b).

There are two main practical reasons to go beyond the static models. First, there are instruments that do not admit a replication by stop-loss options. These include forward-starting CDO tranches, options on tranches, leveraged super-senior tranches, and other innovative structured products such as constant proportion portfolio insurance (CPPI) and constant proportion debt obligations (CPDO). The second reason is the ambiguity of dynamic hedging and the difficulty of managing the risk of forward exposures on the basis of static models, even for positions in standard tranches.

While the potential of the new generation of factor models to build adequate dynamics of portfolio loss starting from loss distributions of individual obligors is certainly far from exhausted (see, e.g., Andersen, 2006;
There is growing appreciation of the benefits of the direct modeling of the loss process $L_t$. The general framework of aggregate-loss-based approaches to basket credit derivatives was put forward by Giesecke and Goldberg (2005), Schönbucher (2005), and Sidenius, Piterbarg, and Andersen (2008). Examples of specific models can be found in the works by Bennani (2005), Brigo, Pallavicini, and Torresetti (2007), Errais, Giesecke, and Goldberg (2006), Ding, Giesecke, and Tomecek (2006), and Longstaff and Rajan (2008).

Both Schönbucher (2005) and Sidenius et al. (2008) aimed to build a credit portfolio counterpart of the Heath-Jarrow-Morton (HJM) framework of the interest rate modeling. In the HJM framework, the problem of fitting the initial term structure of interest rates is non-existent because the discount curve serves as an initial condition and not as a calibration constraint. In the calibration of credit portfolio models, the role of the discount curve is played by the surface of the loss distribution, $\pi(L,t) = P[L_t \leq L]$. In the spirit of the HJM framework, Schönbucher (2005) and Sidenius et al. (2008) eliminated the problem of the calibration to the distribution of loss by making it an initial condition. This, however, was achieved at a price of losing the ability to simulate the loss process without introducing a large number of additional stochastic degrees of freedom, which led to severe computational problems and accentuated the need for a more specific approach.

While the HJM framework indeed provides ultimate flexibility in fitting the market, many of the short-rate models developed before HJM are also capable of fitting the entire discount curve. The flexibility of the calibration was achieved due to the presence of a free function of time in the drift term of the model-defining stochastic differential equation (SDE). The models developed within this scheme had a tremendous impact on the field and are still highly popular among practitioners. In view of this success, it appears reasonable to try an adaptation of the framework of short rates to the problem of credit portfolio loss.

As was pointed out by Schönbucher (2005), models based on an explicit, short-rate-like modeling of the loss intensity generally run into a problem with the calibration to the distribution of loss. Indeed, fitting an entire two-dimensional surface of loss can require a large number of free calibration parameters and is likely to be computationally burdensome. It might be argued that the information about the loss distribution coming from the standard tranches is too sparse to define a complete surface of loss. Brigo et al. (2007), Errais et al. (2006), Ding et al. (2006), and Longstaff and Rajan (2008) reported successful calibration to the tranche market. Their models are
formulated in an open-ended way, so that it might in principle be possible to calibrate increasingly many tranches by adding new terms to the defining equations. However, the problem of finding a model based on an explicit equation for the loss intensity, and amenable to a reasonably fast calibration to a generic distribution of loss, has remained unresolved.\(^2\)

In this work, we introduce a new two-dimensional intensity-based Markovian model of the aggregate credit loss. This model can be easily calibrated to a generic distribution of portfolio loss without sacrificing tractability and robustness. The calibration procedure consists of two steps. On the first step, we find the intensity of an auxiliary one-step Markov chain model consistent with the CDO tranches. Because the intensity of the Markov chain is a deterministic function of accumulated loss and time, it can be called local intensity, to distinguish it from the stochastic intensity of the full model. On the second step, the full two-dimensional model is calibrated to match the local intensity. The idea of exploring the link between the local intensity and the stochastic intensity is borrowed from the method of Markovian projection used by Dupire (1994) and Piterbarg (2007) to relate the local volatility and the stochastic volatility. For the purpose of credit applications, we extended the original formulation of the Markovian projection given by Gyöngy (1986) from diffusions to jump processes.

A model calibrated to the market quotes on CDO tranches can be used to price more complicated dynamic instruments. In this paper, we consider an application to the tranche option which, as we show, can be evaluated easily using the backward induction technique. In a numerical example, we calibrated our model to a set of recent quotes for the tranches on Dow Jones CDX.NA.IG.7 and calculated the values of the tranche option at different strikes.

The rest of the paper is organized as follows. In Section 2, we define our model and give a general discussion of its properties. Section 3 is devoted to model calibration. In Section 3.1, we assume (unrealistically) that a full surface of loss distribution is available. This would be the case if we had arbitrage-free quotes for stop-loss options at all strikes and maturities. We show how to build an efficient numerical procedure for the calibration of the two-dimensional Markovian model once we know the local intensity of the auxiliary one-step Markov chain model. We show, furthermore, that finding the local intensity from a complete loss distribution is straightforward. In practice, only a handful of quotes for CDO tranches with a sparse set of attachment points and maturities are available, so that it is not possible to restore the full distribution of loss
without interpolation assumptions. We address this issue in Section 3.2. Instead of interpolating the loss, we choose to do a parametric fit for the coefficients in a specific analytical form for the local intensity. Numerical results for the calibration are given in Section 3.3.

Applications to dynamic pricing problems are discussed in Section 4. We begin by describing the general backward induction setup in Section 4.1, which is followed by a discussion of numerical results for tranche options in Section 4.2. We conclude in Section 5.

The appendix consists of three parts completing the main text. Appendix A describes the cash flows of the single tranche CDO and explains its replication by a portfolio of stop-loss options. Appendix B contains a digression into the method of Markovian projection for stochastic volatility modeling and our extension of Gyöngy’s lemma for jump processes. Appendix C gives the details of the discretization scheme used in the numerical implementation.

2. THE MODEL

We work in the top-down framework and model the loss as an intensity-based process (see, e.g., Duffie, 2005, for general properties and definitions, and Giesecke & Goldberg, 2005, for a conceptual description of the top-down framework). Other possibilities for introducing a compact formulaic description of the loss process were tried. For example, Bennani (2005) postulates an SDE on the loss process without introducing an intensity. However, this sacrifices the discrete nature of default arrivals, which is generally considered an important feature to keep.

The minimal number of independent variables necessary to describe a state of an intensity-based process is two, the accumulated loss \( L_t \) and the intensity \( \lambda_t \). We do not introduce any additional variables and postulate the following dynamics for the intensity,

\[
d\lambda_t = \kappa(\rho(L_t, t) - \lambda_t)dt + \sigma \sqrt{\lambda_t}dW_t
\]  

(1)

where \( W_t \) is the standard Brownian motion. We follow the work of Errais et al. (2006) in allowing for a back action of the loss onto the intensity process (thus going beyond the class of Cox processes which exclude such an action). We found, however, that restricting the model to the affine framework limits its ability to achieve a stable calibration to the market. The
function $\rho(L_t, t)$, in general, is not linear in the accumulated loss $L_t$, and therefore our model is not affine. This function serves to provide sufficient freedom to calibrate to a generic distribution of loss $\pi(L_t, t)$. In contrast to the affine setup of Errais et al. (2006), our model has no transform-based analytical solution for the stop-loss option. We will show, nevertheless, that an efficient numerical calibration to an arbitrary distribution of loss is possible.

Throughout this paper, we assume that the value of loss-given-default (LGD) is equal to the same amount $h$ for all assets in the basket. Many authors, including Brigo et al. (2007) and Errais et al. (2006), point out the importance of a non-trivial LGD distribution for the market-matching ability of their models. We believe that our model can describe the market data even in its simplest form, with a deterministic LGD, because sufficient flexibility is already built-in via the function $\rho(L_t, t)$. Our model can be generalized to include stochastic LGD at the cost of introducing the third dimension, which in our numerical experiments reduced the performance without a significant functional improvement.

Note that the calibration to the loss distribution will be achieved only by adjusting the function $\rho(L_t, t)$ in the drift term, but not the multiplier $\kappa$ or the volatility $\sigma$ in the diffusion term of Eq. (1). The volatility term is kept available to tune the dynamics of the model. Given the potential for the growth in the variety and liquidity of dynamics-sensitive credit instruments, we can envisage a scenario where the volatility will be calibrated to simpler instruments (e.g., European tranche options) and then used to value more complex ones (e.g., Bermudan tranche options). If necessary, a constant volatility can be generalized to a term structure. This is similar to the calibration strategy for the classic short-rate models of interest rates, including Hull–White (HW), Black–Karasinski (BK), and Cox–Ingersoll–Ross (CIR). For these models, the term structure of volatilities is fitted to the options in a cycle of solver iterations, with the free function of time in the drift term being fitted to the discount curve inside each iteration.

We chose CIR-like dynamics (1) for the intensity only as a starting point. Similar models are possible based on single-factor or multi-factor BK-like equations. It is also possible to introduce jump terms in the intensity. The procedures described in this paper will remain applicable provided the model has a free function of loss and time in the drift term and does not lead to negative values of intensity. In the case of CIR-like dynamics (1) the flow of intensities through the boundary $\lambda = 0$ to the negative values is avoided as long as $\rho(L_t, t) > 0$. 
3. CALIBRATION

For a given function $\rho(L, t)$, the model defined by Eq. (1) can be easily treated numerically. Depending on the financial instrument, it can be solved either by a numerical integration of the backward Kolmogorov equation (as discussed in detail in Section 4.1) or by a forward Monte Carlo simulation. However, a direct iterative calibration would certainly be too time-consuming. In this section, we develop a computationally efficient procedure for the calibration to CDO tranches, which avoids massive iterative fitting.

The goal of the calibration is to find the function $\rho(L, t)$ consistent with the information about the loss distribution available from the market data. As mentioned in the previous section, the reversion strength $\kappa$ and the volatility $\sigma$ are not subject to calibration at this stage. We want to keep them unconstrained by the static information and potentially available for the calibration to dynamically sensitive instruments, such as options on tranches.

The starting point of our calibration procedure is the forward Kolmogorov equation for the joint density $p(\lambda, L, t)$ of intensity $\lambda$ and loss $L$, following from Eq. (1),

$$
\frac{\partial p(\lambda, L, t)}{\partial t} = \left( -\kappa \frac{\partial}{\partial \lambda} (\rho(L, t) - \lambda) + \frac{1}{2} \frac{\partial^2}{\partial \lambda^2} \sigma^2 \lambda \right) p(\lambda, L, t)
+ \lambda (1_{L \geq h} p(\lambda, L - h, t) - p(\lambda, L, t))
$$

(2)

Here $\lambda$ is restricted to be non-negative, and the loss $L$ takes the values $L = 0, h, 2h, \ldots, N_{\text{max}} h$, where $N_{\text{max}}$ is the number of assets in the basket. The absence of the probability flow through the boundary $\lambda = 0$ to the negative intensities is ensured by the boundary condition

$$
\left( -\kappa (\rho(L, t) - \lambda) + \frac{1}{2} \frac{\partial}{\partial \lambda} \sigma^2 \lambda \right) p(\lambda, L, t) = 0, \quad \lambda \to 0
$$

(3)

We also need an initial condition to Eq. (2), which obviously has the following form

$$
p(\lambda, L, 0) = p_0(\lambda) \cdot 1_{L = 0}
$$

(4)

The function $p_0(\lambda)$ is not fully fixed by the market because we cannot observe the distribution of default intensity. The choice of this function will be discussed in Section 3.2.
The probability density of loss is obtained by integrating the joint density over $\lambda$,

$$P(L, t) = \int_{0}^{\infty} p(\lambda, L, t)d\lambda$$  \hspace{1cm} (5)

### 3.1. Calibration to Local Intensity

In this section, we assume that the calibration target is the entire set of one-dimensional marginal loss densities, that is, the dependence $P(L, t) = P[L_t = L]$ for all values of $t$ from 0 to a certain time horizon $T$ and for all possible levels of loss. We discuss how this assumption relates to the actual information available from the market in Section 3.2.

We now reduce Eq. (2) to a simpler forward Kolmogorov equation written on the density of loss, $P(L, t)$. This reduction is achieved by an integration of both parts of Eq. (2) over $\lambda$, which leads to

$$\frac{\partial P(L, t)}{\partial t} = 1_{L \geq h} \Lambda(L - h, t)P(L - h, t) - \Lambda(L, t)P(L, t)$$  \hspace{1cm} (6)

The quantity $\Lambda(L, t)$, which we call the *local* intensity, is defined by the equation

$$\Lambda(L, t)P(L, t) = \int_{0}^{\infty} \lambda p(\lambda, L, t)d\lambda$$  \hspace{1cm} (7)

and has the meaning of the expectation of the intensity $\lambda_t$ conditional on the loss $L$ accumulated by the time $t$,

$$\Lambda(L, t) = E[\lambda_t | L_t = L]$$  \hspace{1cm} (8)

We obtained Eq. (8) from a particular equation for the stochastic evolution of the intensity $\lambda_t$. It can be shown that the result is more general and is also valid for an adapted intensity process which is not necessarily Markovian. A more detailed discussion can be found in Appendix B.\textsuperscript{3}

Eq. (6) is the forward Kolmogorov equation of the jump process based on the intensity $\Lambda(L, t)$ considered as a stochastic process. This jump process is known as a continuous-time, non-homogeneous, one-step Markov chain. The state space of this Markov chain is given by the grid of possible loss values, 0, $h$, $2h$, ..., $N_{\text{max}}h$. For every $L<N_{\text{max}}h$, the intensity of the transition $L \rightarrow L + h$ at time $t$ is equal to $\Lambda(L, t)$, while the intensities of
all other transitions are 0. Viewed as an intensity-based jump process, the one-step Markov chain is a specific case with the intensity being a deterministic function of time and loss. By analogy with local volatility models of exchange rates or interest rates (see Appendix B), it is natural to call this model the local intensity model.

The local intensity model has recently been considered by van der Voort (2006) who applied it to the pricing of forward-starting CDOs. This model also appears in the works of Sidenius et al. (2008) and Schönbucher (2005) who use it as a part of their constructions of the dynamic framework. We regard the local intensity model as a very useful tool for the calibration of models with richer dynamics, but which, by itself, is generally insufficient as a dynamic model (see, e.g., the numerical results for tranche options in Section 4.2).

The local intensity \( \Lambda(\mathcal{L}, t) \) can be easily calibrated to a given density of loss \( P(\mathcal{L}, t) \), which is why van der Voort called it an implied loss model. Indeed, summing up Eq. (6) from \( \mathcal{L} = 0 \) to \( \mathcal{L} = K = kh \) for any \( k \leq N_{\text{max}} \), we obtain

\[
\frac{\partial}{\partial t} \sum_{\mathcal{L}=0}^{K} P(\mathcal{L}, t) = -\Lambda(K, t) P(K, t)
\]  (9)

which uniquely determines the local intensity \( \Lambda(K, t) \) provided \( P(K, t) \neq 0 \) (i.e., for all states which can be achieved by the process),

\[
\Lambda(K, t) = \frac{-1}{P(K, t)} \frac{\partial P[L_t \leq K]}{\partial t}
\]  (10)

This completes the first step of the calibration procedure.

The next step is to find the function \( \rho(\mathcal{L}, t) \) consistent with the local intensity \( \Lambda(\mathcal{L}, t) \). In order to accomplish this task, we take the time derivative of Eq. (7),

\[
\int_{0}^{\infty} \lambda \frac{\partial P(\lambda, L, t)}{\partial t} d\lambda = P(L, t) \frac{\partial \Lambda(L, t)}{\partial t} + \Lambda(L, t) \frac{\partial P(L, t)}{\partial t}
\]  (11)

and substitute the time derivatives of the densities \( p(\lambda, L, t) \) and \( P(L, t) \) from Eqs. (2) and (6), respectively. The resulting equation can be formally solved
for $\rho(L, t)$ (again, for all accessible states which obey $P(L, t) \neq 0$), to give

$$
\rho(L, t) = \Lambda(L, t) + \frac{1}{\kappa} \frac{\partial \Lambda(L, t)}{\partial t}
+ \frac{\Lambda(L, t)(1_{L \geq h} \Lambda(L - h, t)P(L - h, t) - \Lambda(L, t)P(L, t))}{\kappa P(L, T)}
+ \frac{M(L, t) - 1_{L \geq h} M(L - h, t)}{\kappa P(L, T)}
$$

(12)

where

$$
M(L, t) = \int_0^\infty \lambda^2 p(\lambda, L, t) d\lambda
$$

(13)

Eq. (12) is not an analytical solution for $\rho(L, t)$ because this function is implicitly present in the last term in the right-hand side via Eq. (2) that determines the density $p(\lambda, L, t)$. Nevertheless, Eq. (12) essentially solves the calibration problem. Indeed, a substitution of the function $\rho(L, t)$ from Eq. (12) into Eq. (2) leads to an integro-differential equation for the density $p(\lambda, L, t)$, which can be solved numerically. After that, the function $\rho(L, t)$ can be restored from Eq. (12).

Practically, instead of writing down the explicit integro-differential equation, it is more convenient to use Eqs. (2), (12), and (13) to set up a recursive procedure. We illustrate this procedure using a simple first-order scheme. We discretize the time dimension into a sequence of small intervals, $[0, t_1], [t_1, t_2], \ldots$, and introduce a discretization for $\lambda$ (the loss variable being already discrete). Starting at $t = 0$ with the density $p(\lambda, L, t)$ specified by the boundary condition (4), we find the corresponding $M(L, 0)$ from Eq. (13) and plug the result into Eq. (12) to obtain $\rho(L, 0)$. Eq. (2) is then used to propagate the density to $t = t_1$, and then the entire procedure is repeated for each $t = t_i$ until the desired maturity of the model is reached. We note that only one complete propagation through the three-dimensional lattice of the discretized values $t_{i_1}, \lambda_{i_2}$, and $L_{i_3} = i_3 h$ is required. We also note that the integration over $\lambda$ in Eq. (13) does not lead to any significant performance overhead. The overall computational effort is similar to that involved in the calibration of a typical two-factor interest rate model.

It is also possible to use higher-order discretization schemes. In our numerical experiments, second-order schemes performed best.
3.2. From Market Data to Local Intensity

We now turn to the calibration of the local intensity $\Lambda(L, t)$ to the actual market data, that is, to single tranche CDOs. (For a brief description of single tranche CDOs see Appendix A.) We see two alternatives for the local intensity calibration.

In the approach by van der Voort (2006), the relationship (10) is used to find the local intensity from the probability distribution of loss. In turn, the probability distribution of loss is found assuming a particular factor model with a sufficient number of degrees of freedom to fit all the tranches as well as the underlying credit index. For example, this could be the random factor loadings model of Andersen and Sidenius (2005a), or a mixture of several Gaussian copulas considered by Li and Liang (2006), or the implied copula model of Hull and White (2006). Dynamical properties of the auxiliary factor model are ignored, the only requirement being that the model should produce an arbitrage-free distribution of loss. (This requirement generally disqualifies the use of the base correlations model.)

Alternatively, one can assume a certain functional form for the local intensity $\Lambda(L, t)$ and do a parametric fit to the tranches and the index. In our opinion, this approach is more suitable for the purpose of dynamic model building since all assumptions about the time dependence of the local intensity are kept explicit. Such assumptions cannot be avoided because liquid tranche quotes are only available for a handful of maturity points. Consequently, we need to look for the most natural way to interpolate the local intensity. At present, we do not see any reliable way to control the time dependence of the local intensity within the static factor models approach. Therefore, we prefer dealing directly with a functional form of the local intensity.

We now proceed to parametric modeling of the local intensity. In doing so, we found it convenient to express the local intensity as a function $\Lambda(N, t)$ of the number of defaults $N$ and time $t$ instead of loss $L$ and time $t$. (With a deterministic LGD, this is an equivalent description because the number of defaults, $N$, is proportional to loss, $L = Nh$.) The challenge is to reconcile the behavior near $t = 0$, which as we will see turns out to be singular, with the more regular behavior away from $t = 0$. We address this issue by modeling the $N$-dependence of the local intensity as a power series in the following parameter,

$$x = \frac{N}{\hat{N}(t) + z(t)}$$

(14)
The function $\tilde{N}(t)$ is the average number of defaults occurred by time $t$, which is related to (but not fully fixed by) the quote for the credit index spread. The function $z(t)$ is introduced to regularize the singularity at $t \to 0$ in Eq. (14). Specifically, we used an exponential function,

$$z(t) = \exp(-\gamma t)$$

but any monotonic function that decays sufficiently fast with time could be used instead. A representation of the local intensity in terms of the parameter $x$ ensures that, for $t \gg \gamma^{-1}$, the local intensity becomes a function of the number of defaults, $N$, normalized by the expected number of defaults, $\tilde{N}(t)$. This normalization reflects the fact that the typical number of defaults naturally grow with time even in the case where the typical local intensity stays constant in time.

Thus, we look for the local intensity in the form

$$\Lambda(N, t) = x_0(t) + x(N, t), \quad x(N, t) = \sum_{p=1}^{p_{\text{max}}} x_p(t)\lambda^p$$

To ensure the proper normalization of the probability density within the loss interval $[0, hN_{\text{max}}]$ we assume that the above definition holds for $N < N_{\text{max}}$, while at the boundary $N = N_{\text{max}}$ the local intensity is zero

$$\Lambda(N_{\text{max}}, t) = 0$$

The main dependence of the local intensity on time in Eq. (16) is contained in the parameter $x$. A residual dependence on time in the coefficients $x_p$ is included to ensure the matching with the initial condition at $t = 0$ (discussed below), and also for the fit of tranches with different maturities. The number of parameters $p_{\text{max}}$ should be less or equal to the number of tranches at a single maturity, $N_{\text{tr}}$. The best quality is usually achieved at $p_{\text{max}} = N_{\text{tr}}$.

The choice of the initial condition for the local intensity $\Lambda(N, 0)$ has to be consistent with the initial distribution $p_0(\lambda)$ assumed in Eq. (4) for the stochastic intensity model. Indeed, solving the Eq. (2) near $t = 0$, we obtain a family of Poisson distributions with constant intensities $\lambda$ distributed with the density $p_0(\lambda)$,

$$p(\lambda, N, t \approx 0) = \frac{(\lambda t)^N \exp(-\lambda t)p_0(\lambda)}{N!}$$
In this equation, \( \exp(-\lambda t) \) can be replaced by 1 in the same order of approximation. Using this expression for the density \( p(\lambda, L, t) \) with Eqs. (5) and (7), we obtain the following initial condition for the local intensity

\[
\Lambda(N, 0) = \frac{\int \lambda^{N+1} p_0(\lambda) d\lambda}{\int \lambda^N p_0(\lambda) d\lambda}
\]

(19)
in terms of the moments of the initial stochastic intensity distribution.

Setting \( N=0 \) in Eq. (19), we obtain a correct relationship for the instantaneous intensity of the first default,

\[
\Lambda(0, 0) = \int \lambda p_0(\lambda) d\lambda
\]

(20)

The relevance of the relationship (19) with \( N \geq 1 \) is not immediately obvious because neither the local intensity \( \Lambda(N, 0) \) nor the higher moments of the initial intensity distribution \( p_0(\lambda) \) can be extracted from the market data. Nevertheless, Eq. (19) is very important for the consistency of the calibration scheme. Indeed, it shows that a particular choice for the initial distribution \( p_0(\lambda) \) fully determines the initial condition for the local intensity \( \Lambda(N, 0) \), and vice versa. We note also that the distribution \( p_0(\lambda) \) is not guaranteed to exist for an arbitrary choice of \( \Lambda(N, 0) \); in particular, it is not possible to set \( \Lambda(N, 0) = 0 \) for \( N \geq 1 \), which might seem natural.

The easiest way to ensure that Eq. (19) holds is to choose a particular distribution \( p_0(\lambda) \) and restore \( \Lambda(N, 0) \). We used the simplest choice corresponding to a deterministic initial condition \( \lambda = \lambda_0 \) for the stochastic intensity,

\[
p_0(\lambda) = \delta(\lambda - \lambda_0)
\]

(21)
(Here, \( \delta(x) \) is the Dirac \( \delta \)-function, describing a unit weight localized at \( x = 0 \).) This corresponds to a constant initial condition for the local intensity,

\[
\Lambda(N, 0) = \lambda_0
\]

(22)
which leads to \( \Lambda_0(0) = \lambda_0 \) and the following set of initial conditions for the coefficients \( \alpha_p \) with \( p \geq 1 \) in Eq. (16),

\[
\alpha_p(0) = 0, \quad p \geq 1
\]

(23)
These initial conditions are satisfied by the following time dependence,

\[ a_p(t) = \tilde{a}_p(1 - z(t)), \quad p \geq 1 \]  

which interpolates between 0 and an asymptotic constant value. The numerical values of the coefficients \( \tilde{a}_p \) can be fitted to a set of tranches with a single maturity. Simultaneous fit to tranches with several different maturities can be achieved using an additional term structure of the coefficients \( \tilde{a}_p \).

We finally show that the term \( a_0(t) \) is fixed by the time dependence of the average number of defaults \( \bar{N}(t) \). Starting from the expression

\[ \bar{N}(t) = \sum_{N > 0} P(N, t) N \]  

we take the time derivative of both sides and use the Markov chain Eq. (6) to obtain

\[ \frac{d\bar{N}(t)}{dt} = \sum_{N > 0} N(\Lambda(N - 1, t)P(N - 1, t) - \Lambda(N, t)P(N, t)) \]  

Substituting the expression (16) for \( \Lambda(N, t) \), and using that \( \Lambda(N_{\text{max}}, t) = 0 \), we find

\[ a_0(t) = \frac{(d\bar{N}(t)/dt) - \sum_{N \geq 0}^{N_{\text{max}}} a(N, t)P(N, t)}{1 - P(N_{\text{max}}, t)} \]  

This equation is used to determine \( a_0(t) \) while solving the forward Eq. (6) for the local intensity model. We note also that the initial condition \( \lambda = \lambda_0 \) for the stochastic intensity is given by

\[ \lambda_0 = a_0(0) = \left. \frac{d\bar{N}(t)}{dt} \right|_{t=0} \]  

3.3. Numerical Results

We present numerical results for the calibration to a set of tranches on Dow Jones CDX.NA.IG.7 5Y quoted on January 12, 2007 (see Table 1). The index was quoted at 33.5 bp.
We first need to fit the average number of defaults $\bar{N}(t)$ to the index. To fix the time dependence $\bar{N}(t)$ completely, we would need to know a full-term structure of index quotes for all maturities until 5 years. In reality, the most one could currently get is a quote for the index at 3 years and quotes for CDS spreads for some of the assets in the basket at 1 year and 3 years.

In the absence of direct access to reliable information about the initial segment of the term structure, we are forced to introduce a parametric functional dependence for $\bar{N}(t)$. It turns out that a simple exponential decay of the fraction of survived assets, $1 - \bar{N}(t)/N_{\text{max}}$, with a constant hazard rate does not allow for a robust calibration to the tranches. Therefore, we introduced a slightly more complicated form

$$\frac{\bar{N}(t)}{N_{\text{max}}} = 1 - e^{-a(t+b^2)}$$

where $a$ and $b$ are fitting parameters. A positive coefficient $b$ takes into account the effect of an upward slope of the spread curve. We tried different values of $b$, in each case solving for the value of $a$ necessary to reproduce the spread of the index.

Once the dependence $\bar{N}(t)$ has been fixed, we fit the model to the tranches by adjusting the coefficients $\alpha_p'$ that determine the local intensity according to Eqs. (16) and (24). This is done using a multi-dimensional solver. For the calculation of tranche values, we used Eqs. (A.3) and (A.4) from Appendix

<table>
<thead>
<tr>
<th>Tranche (%)</th>
<th>Model</th>
<th>Market</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–3</td>
<td>500.2</td>
<td>500</td>
</tr>
<tr>
<td>3–7</td>
<td>71.8</td>
<td>71.8</td>
</tr>
<tr>
<td>7–10</td>
<td>13.3</td>
<td>13.3</td>
</tr>
<tr>
<td>10–15</td>
<td>5.3</td>
<td>5.3</td>
</tr>
<tr>
<td>15–30</td>
<td>2.8</td>
<td>2.6</td>
</tr>
</tbody>
</table>

The spread of the equity (0–3%) tranche assumes an upfront payment of 23.03% of the tranche notional. The coefficients in Eq. (24) found from the fit are $\tilde{a}_1 = 0.06677$, $\tilde{a}_2 = 0.07201$, $\tilde{a}_3 = -0.006388$, $\tilde{a}_4 = 0.00024$. 

Table 1. Market Data and Model Calibration Results for the Spreads of the Tranches on Dow Jones CDX.NA.IG.7 5Y Quoted on January 12, 2007.
A with $\alpha = \beta = 0.5$ and the standard assumption of 40% for the recovery rate. (The corresponding value for the LGD is $h = 0.6$.) The values of the stop-loss options in the local intensity model are obtained from the loss probability density, $P(L,t)$, found by integrating Eq. (6).

We observed that the quality of the fit was sensitive to the shape of the time dependence of the average number of defaults, $\bar{N}(t)$, controlled by the parameter $b$. In particular, we were not able to fit all five tranches with the required accuracy for $b=0$. However, increasing $b$ to the values of the order of 1.0 to 5.0 resulted in a dramatic improvement of the quality of the fit. For example, we were able to match all the spreads with the accuracy of 0.3 bp using $a = 0.00049$, $b = 4$ and four terms in the local intensity expansion series (16), as shown in Table 1. We used $\gamma = 1.6$ for the interpolation scale, which is defined by Eq. (15) and enters Eqs. (14) and (24). The quality of the fit was essentially insensitive to this parameter.

The surface of the resulting local intensity $\Lambda(N, t)$ is plotted in Fig. 1. There is a spike in the region of small values of $t$ and large values of $N$.

![Fig. 1. Local Intensity as a Function of the Number of Defaults $N$ and Time $t$ Measured in Years, Calibrated to the Data of Table 1.](image-url)
which, however, does not lead to any serious numerical difficulties because the probability of reaching this region is vanishingly small. For a fixed value of \( t \), the local intensity strongly increases with the number of defaults and has a positive convexity. It can be demonstrated that this shape is a signature of the skew in the Gaussian base correlations. For example, the local intensity surface derived from a Gaussian copula with constant correlations will have a much flatter shape (see Appendix B for an additional discussion).

Now that the local intensity surface is known, we can proceed to the final step in the calibration of the stochastic model and find the function \( \rho \) using the method described in Section 3.1. The resulting function \( \rho \) depends on the values of the parameters \( \kappa \) and \( \sigma \) in Eq. (1). In Fig. 2, we present a typical surface plot of the function \( \rho(N, t) \), using \( \kappa = 1, \sigma = 1 \), and the number of defaults, \( N \), instead of loss, \( L \), as an argument. The qualitative behavior of \( \rho(N, t) \) is similar to that of the local intensity \( \Lambda(N, t) \), with a spike in the region of large \( N \) and small \( t \), which is again irrelevant because of a negligible probability of reaching this region.

\[ \text{Fig. 2. Dependence of the Function } \rho \text{ on the Number of Defaults } N \text{ and Time } t \]

\[ \text{Measured in Years for } \kappa = 1 \text{ and } \sigma = 1. \]
4. DYNAMIC APPLICATIONS

We now turn to the pricing of dynamics-sensitive financial instruments with the stochastic model defined by Eq. (1). An efficient implementation is possible both for the forward simulation and for the backward induction, the latter because the model is low-dimensional Markovian. In the present work, we focus our attention on the evaluation of tranche options using backward induction. Applications to forward-starting CDOs and other instruments that require a forward simulation are deferred to a separate work.

4.1. Backward Induction

We begin with a generic description of backward induction, assuming that the discounting rate is 0, so that all discount factors are equal to 1. (We will restore the proper discounting in Section 4.2.) Let \( F(\lambda, L, T) \) be an arbitrary payoff function of the pair of state variables \((\lambda, L)\) that can be achieved at time \(T\). The backward induction to an earlier time \(t\) is the procedure of going from \(F(\lambda, L, T)\) to another payoff function \(F(\lambda, L, t)\) defined as a conditional expectation with respect to the state achieved at time \(t\),

\[
F(\lambda, L, t) = E[F(\lambda_T, L_T, T)|L_t = L, \lambda_t = \lambda]
\]  

This expectation satisfies the backward Kolmogorov equation

\[
\frac{\partial F(\lambda, L, t)}{\partial t} = -\hat{A}_{\text{back}} F(\lambda, L, t)
\]  

where the action of the generator \(\hat{A}_{\text{back}}\) on an arbitrary function \(F(\lambda, L, t)\) is defined by

\[
\hat{A}_{\text{back}} F(\lambda, L, t) = \left(\kappa (\rho(L, t) - \lambda) \frac{\partial}{\partial \lambda} + \frac{1}{2} \sigma^2 \lambda \frac{\partial^2}{\partial \lambda^2} \right) F(\lambda, L, t) - \hat{\lambda} (F(\lambda, L + h, t) - F(\lambda, L, t))
\]  

This generator is a conjugate of the generator present in the right-hand side of the forward Kolmogorov Eq. (2). Correspondingly, our numerical solution of the discretized backward Kolmogorov equation is a conjugated
version of the solution of the forward Kolmogorov equation outlined in Appendix C.

It follows from the replication arguments presented in Appendix A that the payoff fundamental for the tranche valuation is that of the stop-loss option. For a stop-loss option with maturity $T$ and strike $X$, the payoff is a deterministic function of state at time $T$,

$$P_{T,X}(\lambda, L, T) = (L_T - X)^+$$

(33)

There is no dependence on $\lambda$ in the right-hand side of Eq. (33). Such dependence will appear after a backward induction to an earlier time $t$, the result of which represents the value of the stop-loss option as viewed from the prospective of the time $t$,

$$P_{T,X}(\lambda, L, t) = E[(L_T - X)^+ | L_t = L, \lambda_t = \lambda]$$

(34)

Taking $t$ to be the exercise time, the value of the entire tranche as of this time can be represented as a linear combination of the quantities (34) with different values of $X$ and $T$ (see Appendix A). In order to evaluate, for example, an option to enter the tranche, we only need to take the positive part and perform a final backward induction to $t=0$.

4.2. Numerical Results for the Tranche Option

We consider an option that gives the right to buy the protection leg of the tranche, selling the fee leg with a fixed value of spread, called strike, on a certain exercise date $T_{\text{ex}}$. As discussed above, the payoff from the exercise of the option can be represented as a function $V(\lambda, L, T_{\text{ex}})$ of state achieved at time $T_{\text{ex}}$. More specifically, the payoff is given by a linear combination of the elementary conditional expectations (34) for a portfolio of stop-loss options,

$$V(\lambda, L, T_{\text{ex}}) = \sum_{T_i \geq T_{\text{ex}}} w_i P_{T_i,X_i}(\lambda, L, T_{\text{ex}}) \frac{D(T_i)}{D(T_{\text{ex}})}$$

(35)

The weights $w_i$ are defined in Appendix A. Non-trivial discounting factors $D(t)$ have been restored under the assumption of deterministic interest rates. (Extending the model to include additional dimensions for stochastic interest rates is a subject for future work.)
The option exercise condition is taken into account by replacing negative values of \( V(\lambda, L, T_{\text{ex}}) \) by zero,

\[
V^+ (\lambda, L, T_{\text{ex}}) = \max(V(\lambda, L, T_{\text{ex}}), 0)
\]  

(36)

The value of the option is finally obtained by applying the backward induction from \( T_{\text{ex}} \) to 0 to \( V^+ (\lambda, L, T_{\text{ex}}) \) and multiplying by the discount factor \( D(T_{\text{ex}}) \).

We note that the same backward induction procedure can also be implemented for the local intensity model. The only difference is that the states of the local intensity model include loss \( L \), but not intensity \( \lambda \). The local intensity model can also be regarded as a limit of the two-dimensional model at \( \sigma \to 0, \kappa \to \infty \). We will give option pricing results produced by this model to compare with those obtained within the full model.

The dependence of the option value on the strike spread is shown in Fig. 3 for the case of a mezzanine tranche 3–7% with 159 days to exercise. The at-the-money (ATM) strike is defined as the model-independent forward spread \( S_F \), which can be obtained by dividing the forward value of the protection leg by the basis point value of the fee leg. In our case, \( S_F = 79.7 \) bp.

![Fig. 3. Dependence of the Value of the Option on the Mezzanine 3–7% Tranche on the Strike Spread. The Time to Exercise is 159 Days, Which Corresponds to the Exercise on June 20, 2007. Solid Lines Represent the Results from the Two-Dimensional Stochastic Model with \( \kappa = 1 \) and \( D = \sigma^2/2 = 0.1, 0.8, 1.4 \). Dashed Line is the Result from the Local Intensity Model. Option Value is Measured as a Percentage of the Tranche Notional.](image-url)
Solid curves correspond to \( \kappa = 1 \) and different values of the parameter \( D = \sigma^2 / 2 \). The value of the option in the local intensity model is shown by the dashed line. One can see that the change of the strength of the diffusion term in the stochastic intensity model leads to a noticeable change in option values, thereby providing some freedom to fit market quotes for the options. This is in contrast to the local intensity model, which does not have any free parameters remaining after the calibration to the tranches.

Fig. 4 provides an equivalent representation of the option prices in terms of implied Black volatilities. The order of magnitude 80–120% for the volatilities in the ATM region is consistent with typical values used in heuristic estimates.

The hockey-stick-like dependence of the option price generated by the local intensity model is in agreement with the general intuition about the zero-volatility limit. We note, however, that the local intensity model retains stochastic degrees of freedom of the jump process even though it can be obtained as a degenerate limit of the full two-dimensional model. The appearance of two straight lines can be understood by taking into account that the probability of a default before the exercise date is low, so that the main contribution to the option price comes from the scenarios with either 0 or 1 default before \( T_{\text{ex}} \). In each of the two scenarios, the option is either worthless or depends linearly on the strike. The initial steep segment comes from the no-default scenario. The nearly flat segment corresponds to the scenario where the strike is large but the option remains valuable because of

![Fig. 4. Value of the Option on the Mezzanine 3–7% Tranche Expressed in Terms of Implied Black Volatilities. All Option and Model Parameters are the Same as in Fig. 3. Dashed Line Corresponds to the Local Intensity Model.](image)
the default before exercise. (We assume here that the option does not knock out on default.)

We can conclude that the local intensity model is too inflexible to provide a good description of tranche options. The complete two-dimensional model does not suffer from the type of degeneracy exhibited by the local intensity model because of the smoothing provided by an integration over a continuous range of local intensities $\lambda$.

It is interesting to note that the value of the option decreases with increasing parameter $D = \sigma^2/2$. This behavior is not intuitive as the value of the option usually grows with volatility. One should keep in mind, however, that the stochastic model is calibrated to remain consistent with the same surface of loss for any value of $\sigma$. An increased volatility of the diffusion term is compensated by a decrease in the strength of the back-action term driven by the function $\rho(L, t)$. The direction of the total effect on the option value is not obvious.

5. CONCLUSIONS

We suggested a new intensity-based Markovian model for the dynamics of the aggregate credit loss and developed an efficient method for the calibration to the distribution of loss implied by the market data for CDO tranches. The calibration method is based on the technique of Markovian projection, which in our case allows us to associate the original two-dimensional model with a Markov chain generated by a local surface of default intensity. The Markov chain model is used on the first step of the calibration procedure to find the local intensity and the distribution of loss consistent with the market spreads of CDO tranches. After that, the full two-dimensional stochastic model is calibrated to the local intensity, which already includes all the necessary market information.

Apart from the ability to match a generic distribution of loss, our model has additional parametric freedom to control the fit to more complicated dynamics-sensitive instruments. Specifically, the parameter $\sigma$ controls the strength of diffusive fluctuations of default intensity, while the parameter $\kappa$ sets the time scale of reversion in the drift term. The SDE for the intensity is similar to that for the short rate in the CIR model. The similarity, however, should be explored with caution because the drift term includes a back action of the loss process onto the intensity process. Changing the relative value of the coefficients $\kappa$ and $\sigma$, we can go from an intensity process...
dominated by diffusion to one dominated by the back action of loss, while maintaining the calibration to the same distribution of loss.

The model can be used for pricing of different financial instruments via standard methods developed for Markovian stochastic processes. In the present paper, we focused on applications to tranche options. This instrument can be conveniently evaluated using the backward induction technique. We found that the model can produce a wide range of option prices corresponding to different values of $\sigma$.

We note that our approach is not limited to the specific CIR-like intensity dynamics (1). Other equations, for example, those based on BK-like evolution, may turn out to be more suitable for the purpose of credit portfolio modeling. The current evidence, however, indicates that a sufficiently flexible form of a back action term is essential for a model’s ability to match the market of CDO tranches in a robust way.

NOTES

1. The replication of super-senior tranches requires a similar set of options on recovery also (see Appendix A).
2. A solution alternative to ours was independently developed by Arnsdorf and Halperin (2007) and released when the first revision of our paper was in circulation.
3. See also the forthcoming work by Giesecke (2007) which contains a systematic exposition and further applications of the method of Markovian projection to basket credit modeling.
4. To avoid too big values of the local intensity arising from high powers of $x$ in the expansion (16) we used the regularization $\alpha(N, t) \rightarrow \alpha(N, t)\alpha_{\text{max}}/(\alpha(N, t) + \alpha_{\text{max}})$, with $\alpha_{\text{max}} \approx 100$.
5. The proof given in the first revision of our paper was applicable only to non-self-affecting doubly stochastic processes. We thank Kay Giesecke for pointing this out.

ACKNOWLEDGMENTS

We are grateful to the organizers and the participants of the 5th Annual Advances in Econometrics conference at LSU, Baton Rouge (November 3–5, 2006) for the opportunity to present and discuss the results of this work. We would like to thank Alexandre Antonov, Leonid Ryzhik, Serguei Mechkov, Ren-Jie Zhang for useful discussions and especially Kay Giesecke for his valuable comments on the generalization of Markovian projection technique to the case of processes which are not doubly stochastic. We
are grateful to Gregory Whitten and our colleagues at NumeriX for support of our work.

REFERENCES


APPENDIX A. SINGLE TRANCHE CDO

The purpose of this section is to introduce single tranche CDOs and justify the replication of a tranche by a portfolio of stop-loss options. A single tranche CDO is a synthetic basket credit instrument, which involves two parties and references a portfolio of credit names. One party is the buyer of the protection, the other is the seller of the protection. A single tranche CDO contract defines two bounds, $k < K$, called attachment points and usually quoted as percentage points of the total original reference notional $A$ of the underlying portfolio. The lowest tranche, 0–3% or similar, is customarily called the equity tranche. The highest tranche, 30–100% or alike, is called super-senior. The other tranches are called mezzanine and senior. The difference of the bounds $K - k$ is the original notional of the tranche, which is the cap of the liability held by the seller of the protection. Additionally, the single tranche CDO contract defines a schedule of accrual and payment dates, a fixed annualized periodic rate $S$, called tranche spread, and, in the case of equity tranches, an upfront payment to be made by the buyer of the protection. Par spread of a single tranche CDO is defined as the value of $S$ that makes the present value of the tranche equal to 0.

The cash flows are driven by the slice of the loss of the reference portfolio within the segment $[k, K]$. The total loss sustained by the tranche between its inception at time 0 and time $T$ is given by a difference of two stop-loss options,

$$L_{k,K}(T) = (L(T) - k)^+ - (L(T) - K)^+$$

(by definition $(x)^+ = x$ if $x > 0$ and $(x)^+ = 0$ otherwise). As soon as a positive jump $\Delta L_{k,K}$ in the quantity $L_{k,K}$ is reported, the seller of the protection must pay the amount $\Delta L_{k,K}$ to the buyer of the protection. This is the only source of the payments made by the seller of the protection. The payments made by the buyer of the protection are determined by the outstanding notional of the tranche $A_{k,K}(T)$ as a function of time $T$. The initial notional of the tranche is $A_{k,K}(0) = K - k$. The notional of the tranche at time $T$ is given by

$$A_{k,K}(T) = A_{k,K}(0) - L_{k,K}(T)$$

The outstanding notional of the tranche is monitored every day of each payment period, and the fee is accrued on the outstanding notional of the tranche with the rate equal to the tranche spread $S$. The total accrued fee is paid by the buyer of the protection on the payment date.
Let the payment periods be \([0, T_1], [T_1, T_2], \ldots, [T_{f-1}, T_f]\). Introducing the risk-free discount curve \(D(t)\), the leg of the payments made by the protection seller (protection leg) can be approximated as

\[
\begin{align*}
\text{P}_{\text{prot}} & = \sum_i (E[L_{k,K}(T_i)] - E[L_{k,K}(T_{i-1})])D(T_i) \\
& \quad \quad \text{(A.3)}
\end{align*}
\]

Here we ignored the exact timings of defaults. This approximation can be refined by introducing a schedule of default observations, which is more frequent than the payment schedule.

The leg of the payments made by the protection buyer (fee leg) can be approximated as

\[
\begin{align*}
\text{P}_{\text{fee}} & = \sum_i S \cdot \tau(T_{i-1}, T_i)(\alpha A_{k,K}(T_{i-1}) + \beta A_{k,K}(T_i))D(T_i) \\
& \quad \quad \text{(A.4)}
\end{align*}
\]

Here \(\tau(T_{i-1}, T_i)\) is the day count fraction from \(T_{i-1}\) to \(T_i\), and \(\alpha, \beta = 0.5\) we effectively assume that the defaults on the average happen in the middle of the payment period. Again, it is possible to use a more frequent grid of observations to improve the accuracy of the calculation.

The present value of the tranche, \(P_{\text{tr}}\), is equal to the difference of the legs, that is, \(\text{P}_{\text{prot}} - \text{P}_{\text{fee}}\) for protection buyer, and \(\text{P}_{\text{fee}} - \text{P}_{\text{prot}}\) for protection seller. It is easy to see that the final expression can be represented as a linear combination of stop-loss expectations,

\[
\begin{align*}
P_{\text{tr}} & = \sum_j w_j E[(L_j - X_j)^+]D(t_j) \\
& \quad \quad \text{(A.5)}
\end{align*}
\]

Here \(t_j\) is either one of the payment dates or one of the dates of a more frequent grid introduced to improve the accuracy of the calculation; the strike \(X_j\) is one of the two attachment points, \(k\) or \(K\), and \(w_j\) is a weight that can be positive or negative. We assume that the interest rates are deterministic and, where necessary, include the ratios of the discount factors into the definition of the weights \(w_j\) to obtain the replication in form of Eq. (A.5).

The formula (A.5) is given in terms of unconditional expectations and, strictly speaking, does not express the static replication which has to hold at every moment in the life of the instrument. However, exactly the same derivation can be repeated with conditional expectations, leading to a static replication of the tranche by a portfolio of short and long positions in stop-loss options with the weights \(w_j\).
In the case of super-senior tranches, it is also necessary to take into account the amortization provision. The obligatory amortization begins as soon as the cumulative recovery amount \( R(T) \) exceeds \( A - K \) and can extend to the total original notional of the tranche, \( K - k \). The reduction of the tranche notional due to amortization is given by

\[
R_{k,K}(T) = (R(T) - (A - K))^+ - (R(T) - (A - k))^+ \tag{A.6}
\]

This quantity should be subtracted from the right-hand side of Eq. (A.2). It follows that a static replication of super-senior tranches requires recovery options in addition to stop-loss options. This complication does not limit the applicability of static factor models because the recovery options are also insensitive to the dynamics. We note, furthermore, that in the case of a deterministic LGD, the ratio of recovery and loss is a constant factor, so that the recovery options can be rewritten in terms of stop-loss options, which removes the need to model the process for the recovery separately.

We finally note that the index itself can be treated as a tranche with attachment points 0 and 100%. As with any tranche with a large value of the upper attachment point, it is necessary to take into account the contribution from recovery. The value of the index is fully determined by the term structure of expected loss, \( E[L(T)] \), and expected recovery, \( E[R(T)] \). Under the assumption of a deterministic LGD, the value of the tranche is fully determined by the term structure of the expected number of defaults \( \tilde{N}(t) \) (see Eq. (25)).

**APPENDIX B. LOCAL VOLATILITY AND LOCAL INTENSITY**

Here we discuss in more detail the technique of Markovian projection for jump processes and establish a relationship between the stochastic intensity and the local intensity. We also elaborate on the analogy between the local intensity and the local volatility.

A stochastic volatility model involves a filtered probability space \((\Omega, \mathcal{F}, \{\mathcal{F}_t\})\) and an equation

\[
dX_t = \alpha_t dt + \beta_t dW_t \tag{B.1}
\]

where the drift \( \alpha_t \) and the volatility \( \beta_t \) are random processes adapted to the filtration \( \{\mathcal{F}_t\} \). In a local volatility model, the processes \( \alpha_t \) and \( \beta_t \) are...
deterministic functions of $X_t$ and $t$. The local volatility model was introduced by Dupire (1994) in the form

$$\frac{dX_t}{X_t} = \mu(X_t, t)dt + \sigma(X_t, t)dW_t$$  \hspace{1cm} (B.2)

Local volatility models are regarded as a degenerate case of stochastic volatility models. They are easier to solve and calibrate to European options, but they do not generate very realistic dynamics.

There is a remarkable reduction of a stochastic volatility model to a local volatility model with the preservation of all one-dimensional marginal distributions, due to Gyöngy (1986). A non-technical statement of the claim is that the process $X_t$ defined by Eq. (B.1) has the same one-dimensional distributions as the process $Y_t$ defined by the equation

$$dY_t = a(Y_t, t)dt + b(Y_t, t)dW_t$$  \hspace{1cm} (B.3)

where $Y_0 = X_0$, and the local coefficients $a, b$ are given by

$$a(x, t) = E[\alpha_t | X_t = x]$$  \hspace{1cm} (B.4)

$$b^2(x, t) = E[\beta^2_t | X_t = x]$$  \hspace{1cm} (B.5)

The mapping from the process $X_t$ to the process $Y_t$ is called the Markovian projection.

Piterbarg (2007) gave an intuitive proof and numerous applications of this result to various problems of mathematical finance. In such applications, the Markovian projection is typically used for fast calculation of European options in a stochastic volatility model. Fast calculation of European options is often critical to ensure adequate performance at the stage of model calibration. The method works because the European options only depend on the one-dimensional marginals of the underlying rate and can be computed in the effective local volatility model.

To extend the methodology of Markovian projection to stochastic intensity models of credit basket loss, we need a counterpart of Gyöngy’s theorem for jump processes. Omitting the technical conditions, the statement is that a counting process $N_t$ with an adapted stochastic intensity $\lambda_t$ and $N_0 = 0$ has the same one-dimensional marginal distributions as the process $M_t$ with the intensity $\Lambda(M, t)$ given by

$$\Lambda(M, t) = E[\lambda_t | N_t = M]$$  \hspace{1cm} (B.6)
This is the same as Eq. (8), which was derived from the Kolmogorov equation for a specific stochastic intensity process. For a general proof, we start with the expression (10) for the local intensity in terms of the probability distribution \( P(N,t) = P[N_t = N] \),

\[
\Lambda(M, t) = -\frac{\partial P[N_t \leq M]}{\partial t} \quad (B.7)
\]

and write the derivative term as

\[
\frac{d}{dt} P[N_t \leq M] = \frac{d}{dt} E[1_{N_t \leq M}] = \lim_{\epsilon \to +0} \frac{1}{\epsilon} E[1_{N_{t+\epsilon} \leq M} - 1_{N_t \leq M}] \quad (B.8)
\]

Denote \( \delta N = N_{t+\epsilon} - N_t \). Since \( \delta N \geq 0 \), the expression under the expectation in Eq. (B.8) can be written as

\[
1_{N_{t+\epsilon} \leq M} - 1_{N_t \leq M} = -1_{M-\delta N < N_t \leq M} \quad (B.9)
\]

We obtain

\[
\frac{d}{dt} P[N_t \leq M] = -\lim_{\epsilon \to +0} \frac{1}{\epsilon} E[1_{M-\delta N < N_t \leq M}] \quad (B.10)
\]

The leading contribution in \( \epsilon \) comes from the realizations with \( \delta N = 0, 1 \). Thus, one can write

\[
\frac{d}{dt} P[N_t \leq M] = -\lim_{\epsilon \to +0} \frac{1}{\epsilon} E[\delta N 1_{N_t = M}] \\
= -\lim_{\epsilon \to +0} \frac{1}{\epsilon} E[\delta N | N_t = M] P[N_t = M] \\
= -\Lambda(M, t) P[N_t = M] \quad (B.11)
\]

which leads to Eq. (B.6).

We use the local intensity as a key element in the calibration procedure for the two-dimensional Markovian model. In concluding this section, we note that the local intensity calibrated to the market bears a distinctive signature of the correlation skew, as shown in Fig. B1. The Gaussian copula with any constant correlation value leads to a nearly linear dependence of the local intensity on the number of defaults. This is in contrast with the behavior of the local intensity calibrated to the actual market data, which shows a convex segment before saturating at a very large number of defaults.
APPENDIX C. DISCRETIZATION OF INTENSITY

Numerical integration of Eq. (2) by means of a finite difference scheme requires discretization of time $t$ and intensity $\lambda$. The discretization of time does not pose any conceptual difficulties. The discretization of $\lambda$ is more subtle because it needs to be done in a way preserving the key ingredients of the calibration method presented in Section 3.1, including Eqs. (6) and (12). Here we present a simple scheme that satisfies this requirement.

We use a uniform grid, $\lambda_i = i\Delta$, and introduce the finite difference operators $\hat{D}_\pm$ as

\[
\hat{D}_+ f(\lambda_i) = \frac{f(\lambda_i + \Delta) - f(\lambda_i)}{\Delta}, \quad \hat{D}_- f(\lambda_i) = \frac{f(\lambda_i) - f(\lambda_i - \Delta)}{\Delta}
\]  

(C.1)

In the limit $\Delta \to 0$, these converge to the continuous derivative operator $d/d\lambda$. The discrete counterpart of the second-order derivative $d^2/d\lambda^2$ reads

\[
\hat{D}^2 = \hat{D}_+ \hat{D}_- = \hat{D}_- \hat{D}_+
\]  

(C.2)

Fig. B1. Dependence of the Local Intensity on the Number of Defaults at Maturity for Different Values of Flat Gaussian Correlations and for the Market Correlations Skew.
The discretized forward Kolmogorov Eq. (2) takes the form

\[
\frac{\partial}{\partial t} p(\lambda_i, L, t) = \left( -\kappa \hat{D}_-(\rho(L, t) - \lambda_i) + \frac{1}{2} \sigma^2 \hat{D}_+^2 \lambda_i \right) p(\lambda_i, L, t) \\
+ \lambda_i (p(\lambda_i, L - h, t) - p(\lambda_i, L, h))
\]  
(C.3)

(Here and below we omit the indicator \(1_{L \geq h}\) and assume that \(p(\lambda_i, -h, t) = 0\).) Note that the term containing the first-order derivative with respect to \(\lambda\) in Eq. (2) can be replaced either with \(\hat{D}_+\) or with \(\hat{D}_-\). With the choice of \(\hat{D}_-\), we avoid the appearance of boundary terms after the summation over \(\lambda\) (see below). It is convenient to append \(\hat{\lambda}_{-1} = -\Delta\) to the range of allowed intensity values so that the boundary condition can be set as

\[ p(\hat{\lambda}_{-1}, L, t) = 0 \]  
(C.4)

The probability density of loss and the local intensity in the discrete setting are defined similarly to Eqs. (5) and (7),

\[ P(L, t) = \sum_{i=0}^{i_{\text{max}}} p(\lambda_i, L, t) \]  
(C.5)

\[ \Lambda(L, t) P(L, t) = \sum_{i=0}^{i_{\text{max}}} \lambda_i p(\lambda_i, L, t) \]  
(C.6)

Summing both sides of Eq. (C.3) over \(\lambda_i\) from zero to the chosen limit \(i_{\text{max}}\), the forward Kolmogorov Eq. (6) is recovered. The boundary terms at the lower limit of summation disappear because of the condition (C.4).

We now proceed to the derivation of Eq. (12) in the discrete setting. Taking the derivative of both sides of Eq. (C.6) with respect to time, we get

\[
\sum_{i=0}^{i_{\text{max}}} \lambda_i \frac{\partial}{\partial t} p(\lambda_i, L, t) = P(L, t) \frac{\partial}{\partial t} \Lambda(L, t) + \Lambda(L, t) \frac{\partial}{\partial t} P(L, t) 
\]  
(C.7)

After that, we insert the time derivatives of the distribution functions \(p(\lambda_i, L, t)\) and \(P(L, t)\) from Eqs. (6) and (C.3) into Eq. (C.7). We recover Eq. (12) using Eqs. (C.5) and (C.6) and the definition for the second moment of intensity,

\[ M(L, t) = \sum_{i=0}^{i_{\text{max}}} \lambda_i^2 p(\lambda_i, L, t) \]  
(C.8)
Eq. (C.3) represents a system of coupled ordinary differential equations that can be solved by any suitable method. We used the second-order Runge–Kutta scheme.

We note that the absence of the probability flow to the region with negative intensities in the presented numerical scheme is guaranteed by construction. Thus, occasional appearances of negative values of $\rho(L, t)$ will not break the calibration algorithm.

The choice of the step, $\Delta$, and the upper limit for the intensity, $i_{\text{max}}\Delta$, is dictated by accuracy requirements. In our numerical experiments, we achieved the error of less than $1e-5$ using $\Delta \sim 0.07-0.25$ and $i_{\text{max}} \sim 1,000$. 
CREDIT DERIVATIVES AND RISK AVERSION

Tim Leung, Ronnie Sircar and Thaleia Zariphopoulou

ABSTRACT

We discuss the valuation of credit derivatives in extreme regimes such as when the time-to-maturity is short, or when payoff is contingent upon a large number of defaults, as with senior tranches of collateralized debt obligations. In these cases, risk aversion may play an important role, especially when there is little liquidity, and utility-indifference valuation may apply. Specifically, we analyze how short-term yield spreads from defaultable bonds in a structural model may be raised due to investor risk aversion.

1. INTRODUCTION

The recent turbulence in the credit markets, largely due to overly optimistic valuations of complex credit derivatives by major financial institutions, highlights the need for an alternative pricing mechanism in which risk aversion is explicitly incorporated, especially in such an arena where liquidity is sporadic and has tended to dry up. A number of observations suggest that utility-based valuation may capture some
common market phenomena better than the traditional risk-neutral (expectation) valuation:

- Short-term yield spreads from single-name credit derivative prices decay slowly and seem to approach a nonzero limit, suggesting significant anticipation (or phobia) of credit shocks over short horizons.
- Among multi-name products, the premia paid for senior CDO tranches have often been on the order of a dozen or so basis points (e.g., for CDX tranches), ascribing quite a large return for providing protection against the risk of default of 15–30% investment grade companies over a few years. On the other hand, market models seem to have underestimated the risks of less senior tranches of CDOs associated with mortgage-backed securities in recent years.
- The current high yields attached to all credit-associated products in the absence of confidence, suggest that risk averse quantification might presently be better used for securities where hitherto there had been better liquidity.

It is also clear that rating agencies, perhaps willingly neglectful, have severely underestimated the combined risk of basket credit derivatives, especially those backed by subprime mortgages. In a front page article about the recent losses of over $8 billion by Merrill Lynch, the Wall Street Journal (on October 25, 2007) reported: “More than 70% of the securities issued by each CDO bore triple-A credit ratings. . . . But by mid-2006, few bond insurers were willing to write protection on CDOs that were ultimately backed by subprime mortgages . . . Merrill put large amounts of AAA-rated CDOs onto its own balance sheet, thinking they were low-risk assets because of their top credit ratings. Many of those assets dived in value this summer.”

In this article, we focus on the first point mentioned above to address whether utility valuation can improve structural models to better reproduce observed short-term yield spreads. While practitioners have long since migrated to intensity-based models where the arrival of default risk inherently comes as a surprise, hence leading to nonzero spreads in the limit of zero maturity, there has been interest in the past in adapting economically preferable structural models toward the same effect. Some examples include the introduction of jumps (Hilberink & Rogers, 2002; Zhou, 2001), stochastic interest rates (Longstaff & Schwartz, 1995), imperfect information on the firm’s asset value (Duffie & Lando, 2001), uncertainty in the default threshold (Giesecke, 2004a), and fast mean-reverting stochastic volatility (Fouque, Sircar, & Solna, 2006). In related work, utility-based valuation
has been applied within the framework of intensity-based models for both single-name derivatives (Bielecki & Jeanblanc, 2006; Shouda, 2005; Sircar & Zariphopoulou, 2007) and, in addressing the second point, for multi-name products (Sircar & Zariphopoulou, 2006). The mechanism of utility valuation quantifies the investor’s risk aversion and translates it into higher yield spreads.

In a complete market setting, the payoffs of any financial claims can be replicated by trading the underlying securities, and their prices are equal to the value of the associated hedging portfolios. However, in market environments with credit risks, the risks associated with defaults may not be completely eliminated. For instance, if the default of a firm is triggered by the firm’s asset value falling below a certain level, then perfect replication for defaultable securities issued by the firm requires that the firm’s asset value be liquidly traded. While the firm’s stock is tradable, its asset value is not, and hence the market completeness assumption breaks down. The buyer or seller of the firm’s defaultable securities takes on some unhedgeable risk that needs to be quantified in order to value the security. In the Black and Cox (1976) structural model, the stock price is taken as proxy for the firm’s asset value (see Giesecke, 2004b for a survey), but we will focus on the effect of the incomplete information provided by only being able to trade the firm’s stock, which is imperfectly correlated with its asset value.

We will apply the technology of utility-indifference valuation for defaultable bonds in a structural model of Black–Cox type. The valuation mechanism incorporates the bond holder’s (or seller’s) risk aversion, and accounts for investment opportunities in the firm’s stock to optimally hedge default risk. These features have a significant impact on the bond prices and yield spreads (Figs. 1 and 2).

2. INDIFFERENCE VALUATION FOR DEFAULTABLE BONDS

We consider the valuation of a defaultable bond in a structural model with diffusion dynamics. The firm’s creditors hold a bond promising payment of $1 on expiration date $T$, unless the firm defaults. In the Merton (1974) model, default occurs if the firm’s asset value on date $T$ is below a prespecified debt level $D$. In the Black and Cox (1976) generalization, the firm defaults the first time the underlying asset value hits the lower boundary

$$
\tilde{D}(t) = De^{-\beta(T-t)}, \quad t \in [0, T]
$$
where $\beta$ is a positive constant. This boundary represents the threshold at which bond safety covenants cause a default, so the bond becomes worthless if the asset value ever falls below $\bar{D}$ before expiration date $T$.

Let $Y_t$ be the firm’s asset value at time $t$, which we take to be observable. Then, the firm’s default is signaled by $Y_t$ hitting the level $\bar{D}(t)$. The firm’s stock price ($S_t$) follows a geometric Brownian motion, and the firm’s asset value is taken to be a correlated diffusion:

$$dS_t = \mu S_t dt + \sigma S_t dW^1_t$$

(1)

$$dY_t = \nu Y_t dt + \eta Y_t (\rho dW^1_t + \rho' dW^2_t)$$

(2)

The processes $W^1$ and $W^2$ are independent Brownian motions defined on a probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$, where $(\mathcal{F}_t)_{0 \leq t \leq T}$ is the augmented

---

Fig. 1. The Defaultable Bond Buyer’s and Seller’s Yield Spreads. The Parameters are $\nu = 8\%$, $\eta = 20\%$, $r = 3\%$, $\mu = 9\%$, $\sigma = 20\%$, $\rho = 50\%$, $\beta = 0$, Along with Relative Default Level $D/Y = 0.5$. The Curves Correspond to Different Risk-Aversion Parameters $\gamma$ and the Arrows Show the Direction of Increasing $\gamma$ Over the Values $(0.01, 0.1, 0.5, 1)$. 
filtration generated by these two processes. The instantaneous correlation coefficient $r_{A}/C_{0}$ measures how closely changes in stock prices follow changes in asset values and we define $r_{0} = \sqrt{1/C_{0}}r_{2}$. It is easy to accommodate firms that pay continuous dividends, but for simplicity, we do not pursue this here.

2.1. Maximal Expected Utility Problem

We assume that the holder of a defaultable bond dynamically invests in the firm’s stock and a risk less bank account which pays interest at constant rate $r$. Note that the firm’s asset value $Y$ is not market-traded. The holder can partially hedge against his position by trading in the company stock $S$, but not the firm’s asset value $Y$. The investor’s trading horizon $T<\infty$ is
chosen to coincide with the expiration date of the derivative contracts of interest. Fixing the current time $t \in [0, T)$, a trading strategy $\{\theta_u; t \leq u \leq T\}$ is the cash amount invested in the market index $S$, and it is deemed admissible if it is self-financing, non-anticipating, and satisfies the integrability condition $E\{\int_t^T \theta_u^2 \, du\} < \infty$. The set of admissible strategies over the period $[t, T]$ is denoted by $\Theta_{t,T}$. The employee’s aggregate current wealth $X$ then evolves according to

$$dX_s = [\theta_s(\mu - r) + rX_s]ds + \theta_s\sigma dW_s^1, \quad X_t = x \quad (3)$$

Considering the problem initiated at time $t \in [0, T]$, we define the default time $\tau_t$ by

$$\tau_t := \inf\{u \geq t : Y_u \leq D(u)\}$$

If the default event occurs prior to $T$, the investor can no longer trade the firm’s stock. He has to liquidate holdings in the stock and deposit in the bank account, reducing his investment opportunities. (Throughout, we are neglecting other potential investment opportunities, but a more complex model might include these; in multi-name problems, such as valuation of CDOs, this is particularly important: see Sircar & Zariphopoulou, 2006.) For simplicity, we also assume that he receives full pre-default market value on his stock holdings on liquidation. One might extend to consider some loss at the default time, but at a great cost in complexity, since the payoff would now depend explicitly on the control $\theta$. Therefore, given $\tau_t < T$, for $t \in (\tau_t, T]$, the investor’s wealth grows at rate $r$:

$$X_t = X_{\tau_t}e^{r(t-\tau_t)}$$

The investor measures utility (at time $T$) via the exponential utility function $U : \mathbb{R} \mapsto \mathbb{R}_-$ defined by

$$U(x) = -e^{-\gamma x}, \quad x \in \mathbb{R}$$

where $\gamma > 0$ is the coefficient of absolute risk aversion. The indifference pricing mechanism is based on the comparison of maximal expected utilities from investments with and without the credit derivative. We first look at the optimal investment problem of an investor who dynamically invests in the firm’s stock as well as the bank account, and does not hold any derivative. In the absence of the defaultable bond, the investor’s
value function is given by

\[
M(t, x, y) = \sup_{\Theta, T} \mathbb{E}\left\{ -e^{-\gamma T} I_{\{\tau_T > T\}} + (-e^{-\gamma T} e^{(\tau_T - t)}) I_{\{\tau_T \leq T\}} \mid X_t = x, Y_t = y \right\}
\]

(4)

which is defined in the domain \( \mathcal{I} = \{(t, x, y) : t \in [0, T], x \in \mathbb{R}, y \in [\hat{D}(t), +\infty)\} \).

**Proposition 1.** The value function \( M : \mathcal{I} \mapsto \mathbb{R}_+ \) is the unique viscosity solution in the class of function that are concave and increasing in \( x \), and uniformly bounded in \( y \) of the HJB equation

\[
M_t + \mathcal{L}_y M + rx M_x + \max_r \left( \frac{1}{2} \sigma^2 \theta^2 M_{xx} + \theta(\rho \sigma \eta M_{xy} + (\mu - r)M_x) \right) = 0
\]

(5)

where the operator \( \mathcal{L}_y \) is defined as

\[
\mathcal{L}_y = \frac{1}{2} \eta^2 y^2 \frac{\partial^2}{\partial y^2} + vy \frac{\partial}{\partial y}
\]

The boundary conditions are given by

\[
M(T, x, y) = -e^{-\gamma x}, \quad M(t, x, De^{-\beta(T-t)}) = -e^{-\gamma xe^{(T-t)}}
\]

**Proof.** The proof follows the arguments in Theorem 4.1 of Duffie and Zariphopoulou (1993), and is omitted. \( \blacksquare \)

Intuitively, if the firm’s current asset value \( y \) is very high, then default is highly unlikely, so the investor is likely to be able to invest in the firm’s stock \( S \) till time \( T \). Indeed, as \( y \to +\infty \), we have \( \tau_T \to +\infty \) and \( I_{\{\tau_T > T\}} = 1 \) a.s. Hence, in the limit, the value function becomes that of the standard (default-free) Merton investment problem (Merton, 1969) which has a closed-form solution. Formally,

\[
\lim_{y \to +\infty} M(t, x, y) = \sup_{\Theta, T} \mathbb{E}\left\{ -e^{-\gamma T} \mid X_T = x \right\}
\]

\[
= -e^{-\gamma xe^{(T-t)}} e^{-(\mu - r)^2/2\sigma^2(T-t)}
\]

(6)

2.2. Bond Holder’s Problem

We now consider the maximal expected utility problem from the perspective of the holder of a defaultable bond who dynamically invests in the firm’s
stock and the bank account. Recall that the bond pays $1 on date $T$ if the firm has survived till then. Hence, the bond holder’s value function is given by

$$V(t, x, y) = \sup_{\Theta_t,T} \mathbb{E} \left\{ -e^{-\gamma(X_{T+1})} 1_{\{s \geq T\}} + (-e^{-\gamma X_s} e^{(T-t)}) 1_{\{s \leq T\}} \right\}$$

(7)

We have a HJB characterization similar to that in Proposition 1.

**Proposition 2.** The valuation function $V : \mathcal{I} \mapsto \mathbb{R}_+$ is the unique viscosity solution in the class of function that are concave and increasing in $x$, and uniformly bounded in $y$ of the HJB equation

$$V_t + \mathcal{L}_y V + r x V_x + \max_{\theta} \left\{ \frac{1}{2} \sigma^2 \theta^2 V_{xx} + \theta (\rho \sigma y V_{xy} + (\mu - r) V_x) \right\} = 0$$

(8)

with terminal and boundary conditions

$$V(T, x, y) = -e^{-\gamma(1+1)} , \quad V(t, x, D e^{-\beta(T-t)}) = -e^{-\gamma x e^{(T-t)}}$$

If the firm’s current asset value $y$ is far away from the default level, then it is very likely that the firm will survive through time $T$, and the investor will collect $1$ at maturity. In other words, as $y \to +\infty$, the value function (formally) becomes

$$\lim_{y \to +\infty} V(t, x, y) = \sup_{\Theta_t,T} \mathbb{E} \left\{ -e^{-\gamma(t+1)} 1_{\{s \geq T\}} \right\} \left\{ X_t = x \right\}$$

$$= -e^{-\gamma(1+xe^{(T-t)})} e^{-((\mu - r)^2/2\sigma^2)(T-t)}$$

(9)

### 2.3. Indifference Price for the Defaultable Bond

The buyer’s indifference price for a defaultable bond is the reduction in his initial wealth level such that the maximum expected utility $V$ is the same as the value function $M$ from investment without the bond. Without loss of generality, we compute this price at time zero.

**Definition 1.** The buyer’s indifference price $p_{0,T}(y)$ for a defaultable bond with expiration date $T$ is defined by

$$M(0, x, y) = V(0, x - p_{0,T}, y)$$

(10)

where $M$ and $V$ are given in (4) and (7).
It is well known that the indifference price under exponential utility does not depend on the investor’s initial wealth $x$. This can also be seen from Proposition 3 below. When there is no default risk, then the value functions $M$ and $V$ are given by (6) and (9). From the above definition, we have the indifference price for the default-free bond as $e^{-rT}$, which is just the present value of the $1 to be collected at time $T$, and is independent of the holder’s risk aversion and the firm’s asset value.

2.4. Solutions for the HJB Equations

The HJB equation (5) can be simplified by the familiar distortion scaling

$$M(t, x, y) = -e^{-\gamma x e^{(T-t)}} u(t, y)^{1/(1-\rho^2)} $$

(11)

The non-negative function $u$ is defined over the domain $J = \{(t, y) : t \in [0, T], y \in [\tilde{D}(t), +\infty)\}$. It solves the linear (Feynman–Kac) differential equation

$$u_t + \tilde{\mathcal{L}}_y u - (1 - \rho^2) \frac{(\mu - r)^2}{2\sigma^2} u = 0,$$

$$u(T, y) = 1,$$

$$u(t, De^{-\beta(T-t)}) = 1$$

(12)

where

$$\tilde{\mathcal{L}}_y = \mathcal{L}_y - \rho \frac{(\mu - r)}{\sigma} \eta y \frac{\partial}{\partial y}$$

For the bond holder’s value function, the transformation

$$V(t, x, y) = -e^{-\gamma (xe^{(T-t)} + 1)} w(t, y)^{1/(1-\rho^2)}$$

(13)

reduces the HJB equation (8) to the linear PDE problem

$$w_t + \tilde{\mathcal{L}}_y w - (1 - \rho^2) \frac{(\mu - r)^2}{2\sigma^2} w = 0$$

$$w(T, y) = 1,$$

$$w(t, De^{-\beta(T-t)}) = e^{\gamma(1-\rho^2)}$$

(14)

which differs from (12) only by a boundary condition. By classical comparison results (Protter & Weinberger, 1984), we have

$$u(t, y) \leq w(t, y), \text{ for } (t, y) \in J$$

(15)
Furthermore, \( u \) and \( w \) admit the Feynman–Kac representations

\[
\begin{align*}
  u(t, y) &= \mathbb{E} \left\{ e^{-(1-\rho^2)((\mu-r)^2/2\sigma^2)(\tau_t \wedge T-t)} | Y_t = y \right\} \\
  w(t, y) &= \mathbb{E} \left\{ e^{-(1-\rho^2)((\mu-r)^2/2\sigma^2)(T-t)} 1_{\{\tau_t > T\}} \right. \\
  & \quad + \left. e^{-(1-\rho^2)((\mu-r)^2/2\sigma^2)(\tau_t - t)} e^{-(1-\rho^2)((\mu-r)^2/2\sigma^2)(\tau_t - t)} 1_{\{\tau_t \leq T\}} | Y_t = y \right\}
\end{align*}
\]  

where the expectations are taken under the measure \( \mathbb{P} \) defined by

\[
\mathbb{P}(A) = \mathbb{E} \left\{ \exp \left( -\frac{\mu-r}{\sigma} W_t - \frac{1}{2} \left( \frac{\mu-r}{\sigma} \right)^2 T \right) 1_A \right\}, \quad A \in \mathcal{F}_T
\]  

Hence, under \( \mathbb{P} \), the firm’s stock price is a martingale, and the dynamics of \( Y \) are

\[
dY_t = \left( v - \rho \frac{(\mu-r)}{\sigma} \eta \right) Y_t dt + \eta Y_t d\tilde{W}_t, \quad Y_0 = y
\]

where \( \tilde{W} \) is a \( \mathbb{P} \)-Brownian motion. The measure \( \mathbb{P} \) is the equivalent martingale measure that has the minimal entropy relative to \( \mathbb{P} \) (Fritelli, 2000). This measure arises frequently in indifference pricing theory.

The representations (16) and (17) are useful in deriving closed-form expressions for the functions \( u(t,y) \) and \( w(t,y) \). First, we notice that

\[
\begin{align*}
  u(t, y) &= e^{-(1-\rho^2)((\mu-r)^2/2\sigma^2)(\tau_t \wedge T-t)} \mathbb{E} \{ \tau_t > T | Y_t = y \} \\
  & \quad + \mathbb{E} \left\{ e^{-(1-\rho^2)((\mu-r)^2/2\sigma^2)(\tau_t - t)} 1_{\{\tau_t \leq T\}} | Y_t = y \right\}
\end{align*}
\]

Under the measure \( \mathbb{P} \), the default time \( \tau_t \) is given by

\[
\tau_t = \inf \left\{ u \geq t : \left( v - \rho \frac{(\mu-r)}{\sigma} \eta - \frac{\eta^2}{2} - \beta \right)(u-t) \\
  + \eta(\tilde{W}_u - \tilde{W}_t) \leq \log \left( \frac{D}{Y_t} \right) - \beta(T-t) \right\}
\]

Then, we explicitly compute the representations using the distribution of \( \tau_t \), which is well known (Karatzas & Shreve, 1991). Standard yet tedious
calculations yield the following expression for \( u(t,y) \):

\[
u(t,y) = e^{-\sigma(T-t)} \left[ \Phi \left( \frac{-b + \psi(T-t)}{\sqrt{T-t}} \right) - e^{2\psi b} \Phi \left( \frac{b + \psi(T-t)}{\sqrt{T-t}} \right) \right] + e^{b(\psi-c)} \left[ \Phi \left( \frac{b - c(T-t)}{\sqrt{T-t}} \right) + e^{2bc} \Phi \left( \frac{b + c(T-t)}{\sqrt{T-t}} \right) \right]
\]

Here \( \Phi(\cdot) \) is the standard normal cumulative distribution function, and

\[
\begin{align*}
x &= (1 - \rho^2) \frac{(\mu - r)^2}{2\sigma^2}, \quad b = \frac{\log(D/y) - \beta(T-t)}{\eta}, \\
\psi &= \frac{v - \beta}{\eta} - \rho \left( \frac{\mu - r}{\sigma} \right) - \frac{\eta}{2}, \quad c = \sqrt{\psi^2 + 2x}
\end{align*}
\]

A similar formula can be obtained for \( w(t,y) \):

\[
w(t,y) = e^{-\sigma(T-t)} \left[ \Phi \left( \frac{-b + \psi(T-t)}{\sqrt{T-t}} \right) - e^{2\psi b} \Phi \left( \frac{b + \psi(T-t)}{\sqrt{T-t}} \right) \right] + e^{\gamma(1-\rho^2)}e^{b(\psi-c)} \left[ \Phi \left( \frac{b - c(T-t)}{\sqrt{T-t}} \right) + e^{2bc} \Phi \left( \frac{b + c(T-t)}{\sqrt{T-t}} \right) \right]
\]  \(19\)

3. THE YIELD SPREAD

Using (11) and (13), we can express the indifference price and the yield spread (at time zero), which can be computed using the explicit formulas for \( u(0,y) \) and \( w(0,y) \) above.

**Proposition 3.** The indifference price \( p_{0,T}(y) \) defined in (10) is given by

\[
p_{0,T}(y) = e^{-rT} \left( 1 - \frac{1}{\gamma(1-\rho^2)} \log \frac{w(0,y)}{u(0,y)} \right)
\]  \(20\)

It satisfies \( p_{0,T}(y) \leq e^{-rT} \) for every \( y \geq D e^{-\beta T} \). The yield spread, defined by

\[
\gamma_{0,T}(y) = -\frac{1}{T} \log(p_{0,T}(y)) - r
\]  \(21\)

is non-negative for all \( y \geq D e^{-\beta T} \) and \( T > 0 \).
Proof. The fact that $p_{0,T} \leq e^{-rT}$ follows from the inequality $u \leq w$. To show that $\mathcal{Y}_{0,T}$ is well defined, we need to establish that $p_{0,T} \geq 0$. For this, consider $w^* := e^{-\gamma(1-\rho^2)} w$, and observe that it satisfies the same PDE as $u$, as well as the same condition on the boundary $\{ y = De^{-\beta(T-t)} \}$ and terminal condition $w^*(T,y) = e^{-\gamma(1-\rho^2)} \leq 1$. Therefore $w^* \leq u$, which gives $w \leq e^{\gamma(1-\rho^2)} u$, and the assertion follows. ■

3.1. The Seller’s Price and Yield Spread

We can construct the bond seller’s value function by replacing $+1$ by $-1$ in the definition (7) of $V$, and the corresponding transformation (13). If we denote the seller’s indifference price by $\tilde{p}_{0,T}(y)$, then

$$
\tilde{p}_{0,T}(y) = e^{-rT} \left( 1 - \frac{1}{\gamma(1-\rho^2)} \log \frac{\mu(0,y)}{\tilde{w}(0,y)} \right)
$$

where $\tilde{w}$ solves

$$
\tilde{w}_t = \tilde{L} \tilde{w} - (1-\rho^2) \frac{(\mu - r)^2}{2\sigma^2} \tilde{w} = 0 \\
\tilde{w}(T,y) = 1, \\
\tilde{w}(t, De^{-\beta(T-t)}) = e^{-\gamma(1-\rho^2)}
$$

The comparison principle yields

$$
u(t,y) \geq \tilde{w}(t,y), \text{ for } (t,y) \in \mathcal{J} \quad (23)$$

Therefore, $\tilde{p}_{0,T}(y) \leq e^{-rT}$, and the seller’s yield spread, denoted by $\tilde{Y}_{0,T}(y)$, is also non-negative for all $y \geq De^{-\beta T}$ and $T > 0$, as follows from a similar calculation to that in the proof of Proposition 3. We obtain a closed-form expression for $\tilde{w}$ by replacing $e^{\gamma(1-\rho^2)}$ by $e^{-\gamma(1-\rho^2)}$ in (19) for $w$.

3.2. The Term-Structure of the Yield Spread

The yield spread term-structure is a natural way to compare zero-coupon defaultable bonds with different maturities. The plots of the buyer’s and seller’s yield spreads for various risk aversion coefficients and Sharpe ratios of the firm’s stock are shown, respectively, in Figs. 1 and 2. While risk
aversion induces the bond buyer to demand a higher yield spread, it reduces the spread offered by the seller. On the other hand, a higher Sharpe ratio of the firm’s stock, given by \((\mu-r)/\sigma\), entices the investor to invest in the firm’s stock, resulting in a higher opportunity cost for holding or selling the defaultable bond. Consequently, both the buyer’s and seller’s yield spreads increase with the Sharpe ratio.

It can be observed from the formulas for \(u\) and \(w\) that the yield spread depends on the ratio between the default level and the current asset value, \(D/y\), rather than their absolute levels. As seen in Fig. 3, when the firm’s asset value gets closer to the default level, not only does the yield spread increase, but the yield curve also exhibits a hump. The peak of the curve moves leftward, corresponding to shorter maturities, as the default-to-asset ratio increases. In these figures, we have taken \(\beta=0\): the curves with \(\beta>0\) are qualitatively the same.

Fig. 3. The Defaultable Bond Buyer’s and Seller’s Yield Spreads for Different Default-to-Asset Ratios \((D/y)\). The Parameters are \(v=8\%, \eta=20\%, r=3\%, \gamma=0.5\), \(\mu=9\%, \sigma=20\%, \rho=50\%, \beta=0\).
3.3. Comparison with the Black–Cox Model

We compare our utility-based valuation with the complete market’s Black–Cox price. In the Black–Cox setup, the firm’s asset value is assumed tradable and evolves according to the following diffusion process under the risk-neutral measure $\mathbb{Q}$:

$$dY_t = rY_t dt + \eta Y_t dW_t^Q$$  \hspace{1cm} (24)

where $W_t^Q$ is a $\mathbb{Q}$-Brownian motion. The firm defaults as soon as the asset value $Y$ hits the boundary $\tilde{D}$. In view of (24), the default time is then given by

$$\tau = \inf \left\{ t \geq 0 : \left( r - \frac{\eta^2}{2} - \beta \right) t + \eta W_t^Q = \log \left( \frac{D}{y} \right) - \beta T \right\}$$

The price of the defaultable bond (at time zero) with maturity $T$ is

$$c_{0,T}(y) = \mathbb{E}^Q \{ e^{-rT} 1_{\{\tau > T\}} \} = e^{-rT} \mathbb{Q} \{ \tau > T \}$$

which can be explicitly expressed as

$$c_{0,T}(y) = e^{-rT} \left[ \Phi \left( \frac{-b + \phi T}{\sqrt{T}} \right) - e^{2\phi b} \Phi \left( \frac{b + \phi T}{\sqrt{T}} \right) \right]$$

with

$$\phi = \frac{r}{\eta} - \frac{\eta}{2} - \frac{\beta}{\eta}$$

Of course the defaultable bond price no longer depends on the holder’s risk aversion parameter $\gamma$, the firm’s stock price $S$, nor the drift of the firm’s asset value $\nu$.

In Fig. 4, we show the buyer’s and seller’s yield spreads from utility valuation for two different values of $\nu$, and low and moderate risk aversion levels (left and right graphs, respectively), and compare them with the Black–Cox yield spread. From the bond holder’s and seller’s perspectives, since defaults are less likely if the firm’s asset value has a higher growth rate, the yield spread decreases with respect to $\nu$. Most strikingly, in the top-right graph, with moderate risk aversion, the utility buyer’s valuation enhances short-term yield spreads compared to the standard Black–Cox valuation. This effect is reversed in the seller’s curves (bottom-right). We observe therefore that the risk averse buyer is willing to pay a lower price for short-term
defaultable bonds, so demanding a higher yield. We highlight this effect in Fig. 5 for a more highly distressed firm, and plotted against log maturities.

4. CONCLUSIONS

Utility valuation offers an alternative risk aversion based explanation for significant short-term yield spreads observed in single-name credit spreads. As in other approaches which modify the standard structural approach for default risk, the major challenge is to extend to complex multi-name credit derivatives. This may be done if we assume independence between default times and “effectively correlate” them through utility valuation (see Fouque, Wignall, & Zhou, 2008 for small correlation expansions around the independent case with risk-neutral valuation). Another possibility is to assume a large degree of homogeneity between the names...
ACKNOWLEDGMENTS

The work of Tim Leung was partially supported by NSF grant DMS-0456195 and a Charlotte Elizabeth Procter Fellowship. The work of Ronnie Sircar was partially supported by NSF grant DMS-0456195, and the work of Thaleia Zariphopoulou was partially supported by NSF grants DMS-0456118 and DMS-0091946.

REFERENCES


