# Multivariate First-Passage Models in Credit Risk

by

Adam Metzler

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Statistics and Actuarial Science

Waterloo, Ontario, Canada, 2008

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

**Abstract**

This thesis deals with credit risk modeling and related mathematical issues. In particular we study first-passage models for credit risk, where obligors default upon first passage of a "credit quality" process to zero.

The first passage problem for correlated Brownian motion is a mathematical structure which arises quite naturally in such models, in particular the seminal multivariate Black-Cox model. In general this problem is analytically intractable, however in two dimensions analytic results are available. In addition to correcting mistakes in several published formulae, we derive an exact simulation scheme for sampling the passage times. Our algorithm exploits several interesting properties of planar Brownian motion and conformal local martingales.

The main contribution of this thesis is the development of a novel multivariate framework for credit risk. We allow for both stochastic trend and volatility in credit qualities, with dependence introduced by letting these quantities be driven by systematic factors common to all obligors. Exploiting a conditional independence structure we are able to express the proportion of defaults in an asymptotically large portfolio as a path functional of the systematic factors. The functional in question returns crossing probabilities of time-changed Brownian motion to continuous barriers, and is typically not available in closed form. As such the distribution of portfolio losses is in general analytically intractable. As such we devise a scheme for simulating approximate losses and demonstrate almost sure convergence of this approximation. We show that the model calibrates well, across both tranches and maturities, to market quotes for CDX index tranches. In particular we are able to calibrate to data from 2006, as well as more recent "distressed" data from 2008.

**Acknowledgements**

To begin I would like to thank my thesis advisor, Professor Don L. McLeish, whose patience and generosity are apparently without limit. Completing this work under Don's guidance has been a pleasure, and I have benefited enormously from his incredible mathematical insight.

I would like to thank Professors David Saunders and Adam Kolkiewicz for their careful reading of this thesis, their valuable comments and suggestions, as well as their invaluable assistance in my job search. I would also like to thank Professors Ken Vetzal and Tom Hurd for their careful reading of this thesis and valuable comments.

I am very grateful for the financial support provided by NSERC, OGS, the Department of Statistics and Actuarial Science, the Faculty of Mathematics and the University of Waterloo.

Last but in no way least, I would like to thank my best friend and fiancée, the lovely Nicole Celian Mascarenhas. There is simply no way I would be where I am today without her unconditional love and support. Thank you for the best ten years of my life.

## Dedication

This accomplishment would not have been possible without the constant support and encouragement of the Adam, Mascarenhas and Metzler families. As such this thesis is dedicated to the wonderful memories of Joseph and Phoebe Adam, Lucas Mascarenhas and Marjorie Metzler.

# Contents

# List of Tables

# List of Figures

# Chapter 1

# Introduction

In many financial contracts, such as mortgages and corporate bonds, one of the parties involved has contractual obligations which extend over several years. The term "credit risk" refers to the fact that said party may prove unable to fulfill these obligations. As evidenced by the ongoing sub-prime mortgage crisis in the United States, the consequences of such failures, or "defaults," can be severe and widespread, extending far beyond those involved in the original transactions and even crossing international borders. This crisis has also demonstrated that industry-standard models for credit risk are simply not appropriate. These include models used by large financial institutions to value complex contracts, as well as those used by regulatory bodies to communicate the risk profiles of such contracts to less sophisticated investors.

Central to credit risk applications are "default times." These are instances when an obligor defaults on some or all of its obligations. Mathematical modeling of default times presents a challenging multivariate problem for several reasons. One reason is that dependence between default times of various firms is quite significant. The sources of this dependence could be a shared set of macroeconomic or industry factors which influence firm values, or business relationships such as that between an auto manufacturer and its suppliers. In addition quantities such as default rates and correlations tend to evolve dynamically through time, for instance both tend to be much higher during recessions. While a realistic dependence structure between default times is crucial, it is often complicated by practical concerns. Large financial institutions typically have exposure to a very large number of credits, and many popular credit derivatives are based on underlying portfolios consisting of up to 125 obligors. This means that the dimension of many real-world applications is quite high, and the trade-off between complexity of the model and computational

efficiency is a recurring theme in the literature.

The exact joint distribution of default times is not always required in credit risk applications. In many cases the distribution of the loss on a given portfolio is sufficient. For example a portfolio manager will typically be interested in quantities such as Value-at-Risk or expected losses. Mathematically the portfolio loss at a given time $t$ is represented as

$$L(t) = \sum_{i=1}^{N} E_i(1 - R_i)I\left(\tau_i \leq t\right)$$

where $N$ is the number of obligors in a given portfolio, $E_i$ is the principal owed by the $i^{th}$ obligor (often referred to as the exposure to the obligor), $R_i$ is the recovery rate for the $i^{th}$ obligor, $\tau_i$ is the default time for the $i^{th}$ obligor and $I\left(A\right)$ is the indicator of the event $A$. In practice recovery rates cannot be predicted with certainty and should be treated as random variables. In addition recovery rates tend to be correlated with default rates, which entails a further complication to the modeling process.

In recent years credit risk has become the most vibrant field of research in mathematical finance. A major impetus for this activity has been the rapid growth in the market for credit derivatives. At the present time the most liquid credit derivatives are credit default swaps (CDS) and CDS index tranches. Loosely speaking a CDS is an insurance contract against corporate defaults. In exchange for regular premium payments, a buyer of protection is reimbursed for losses incurred as a result of a specific default. Default swaps and their valuation are discussed in more detail in Appendix A.

A CDS index tranche is an example of a collateralized debt obligation (CDO). Loosely speaking a CDO allocates the cash flows accruing from a pre-specified portfolio of credits to different investors. These underlying credits could be individual mortgages, corporate bonds or even credit default swaps. The structure of CDOs can be quite complex, here we attempt the briefest of introductions. The procedure is to first define a series of "tranches" of varying levels of "seniority." Each tranche is defined by an "attachment" and "detachment" point. For example the "mezzanine" tranche might have attachment and detachment points of 3% and 7%, respectively. An investor in this tranche would not experience any losses until 3% of the portfolio has defaulted. Our investor would then be responsible for absorbing all subsequent portfolio losses, until their cumulative level has reached 7% of the portfolio prinicpal. Upon the occurence of this event our investor's contractual obligations are essentially terminated. As the individual tranches of a CDO are

subject to varying degrees of risk, they are compensated accordingly. For example our mezzanine investor might receive a regular coupon payment equal to 2% of the portfolio principal, while an investor in a tranche with an attachment point of 15% might only receive 0.5%. A CDS index tranche is simply a CDO where the underlying portfolio is a set of 125 individual default swaps on investment-grade corporations. CDOs and index tranches, as well as their valuation, are discussed in more detail in Appendix A.

The application of any model requires the specification of certain parameters, and appropriate methods for inferring or estimating the relevant parameters are crucial. Index tranches are the only liquid product whose values are directly influenced by dependence between defaults in large portfolios, as such their prices constitute an extremely valuable source of information regarding that dependence. On this issue Finger [55] notes that the availability of index tranche quotes

> . . . is the most direct indicator of the correlation of the names in the collateral pool. It is desirable, then, to leverage this information to value non-standard tranches on the index portfolios, as well as synthetic CDOs backed by other portfolios. Of course, there is a basis risk involved with applying correlations for the index portfolio to bespoke CDOs, but this risk appears preferable to any of the more indirect methods [such as historical default experience or time series of equity correlations] of establishing correlations.

It is no surprise, then, that market participants have begun to view the ability of a particular model to calibrate to market quotes for index tranches as a crucial property.

## 1.1 Credit Risk Modeling

There is a vast literature on the modeling of default times, portfolio losses and the valuation of credit derivatives. A good introduction is Bielecki and Rutkowski [18], though with the rapid pace of innovation in this area several new approaches are not included there. A fantastic source of cutting-edge research is the website www.defaultrisk.com. This site is managed by a professional in the field, and is essentially a repository working papers from both industry and academia. We begin with a brief description of the industry-standard model for pricing CDX index tranches.

### 1.1.1 The Industry-Standard Model

Li [87] is generally credited with introducing copulae to the finance community, and the Gaussian copula has now become the industry standard for pricing index tranches and other related products. Note that in general the $N$-dimensional Gaussian copula has $N(N-1)/2$ correlation parameters. The industry-standard model assumes that default times for the names in the portfolio have a Gaussian copula with equal correlation $\rho$.

Having specified the copula, the joint distribution becomes uniquely specified once marginal distributions are assigned. Construction of marginal distributions is typically accomplished in practice by using a term structure of default swap spreads. O'Kane and Turnbull [93] discuss a common procedure which assumes $\tau_i$ has the same distribution as the first event in a Poisson process with piecewise constant intensity. Intensities over various intervals are chosen to match the observed term structure of swap spreads for a particular name.

It is well known that the equicorrelated Gaussian copula model provides a poor fit to market quotes for CDS index tranches. In particular it is typically not possible to determine one correlation parameter which prices all tranches at a single maturity accurately. A common phenomenon when fitting the model to individual tranche quotes is the so-called "correlation skew." Given the spread on a particular index tranche, the *implied correlation* for that tranche is the value of $\rho$ which equates the model-implied spread with the observed spread. If one successively computes the implied correlation for each tranche, it is quite typical to find a pattern such as that in the first row of the following table, taken from D'Amato and Gyntelberg [36]

| Tranche | 0-3% | 3-7% | 7-10% | 10-15% | 15-30% |
|---|---|---|---|---|---|
| Implied Correlation | 0.21 | 0.06 | 0.15 | 0.22 | 0.26 |
| Base Correlation | 0.25 | 0.3 | 0.35 | 0.42 | 0.62 |

A problem with the implied correlation approach is that in some cases, there may be either multiple solutions (i.e. more than one correlation parameter which prices the tranche under consideration correctly) or no solution at all. To this end market participants tend to employ the *base correlation* approach, described nicely by Finger (see [55], which contains an excellent discussion of the industry implementation of the Gaussian copula)

> The key observation in the approach is that any arbitrary tranche can be represented as the difference between two first-loss tranches. For

instance, selling protection on the 3-7% tranche of the North American index is equivalent to buying protection of the 0-3% tranche and selling protection on a hypothetical 0-7% tranche. Under the base correlation approach, we fix a single correlation to price the 0-3% tranche, then look for a second correlation to price the 0-7% tranche such that the price difference is consistent with the observed 3-7% tranche. By construction, this approach resolves the problem of multiple correlation solutions, since first loss tranche prices are always monotonic with correlation.

We conclude this section with a simple observation. In the equicorrelated Gaussian copula model we may express (for positive correlation) the default time of obligor $i$ as $\tau_i = F_i^{-1}(\Phi(X_i))$, where $F_i$ is the marginal distribution function of $\tau_i$, $\Phi$ is the standard normal distribution function and

$$X_i = \sqrt{\rho} M + \sqrt{1-\rho^2} Y_i \tag{1.1}$$

Here $M, Y_1, \ldots, Y_N$ are independent standard normal variables. The common factor $M$ is a systematic component, common to all firms, and is typically interpreted as representing the "state of the market." The $Y_i$ represent firm-specific risk factors, often interpreted as representing idiosyncratic risk. The parameter $\rho \in (-1, 1)$ represents the sensitivity of obligors to the common factor and is equal to the correlation between credit qualities (though not necessarily the correlation between default times).

### 1.1.2 Copula and Factor Models

There is an abundance of available information on the marginal distributions of default times, in the form of both bond yields and default swap spreads. Though default risk certainly plays a major role in determining bond yields, several authors (see [49] and references therein) have noted that additional factors, such as liquidity, also play an important role. As such market participants have come to view swap spreads as a "more pure" form of information on default risk, and prefer these quotes as a source of data for inferring marginal distributions of default times.

Copulae provide the simplest method for constructing a multivariate model when marginal distributions are known (or at least when one has considerable information on marginal distributions). In a copula model one simply passes the ascribed margins through a given copula to arrive at the joint distribution of default

times. Hager and Schöbel [68] discuss how various observed correlation smiles can be obtained using more general correlation structures in the Gaussian copula, while Laurent and Gregory [65] investigate CDO valuation using the Clayton copula. We note in passing that the Archimedean copulae represent an important class due to the fact that they often admit a *conditional independence* structure, which can have enormous implication for efficient evaluation of the portfolio loss distribution.

In a factor model one explicitly models default times as a function of both systematic and idiosyncratic "factors." The typical approach would be to model the default time of obligor $i$ as

$$\tau_i = f_i\left(\mathbf{M}, \mathbf{Y}_i\right)$$

where $\mathbf{M}, \mathbf{Y}_i, \ldots, \mathbf{Y}_N$ are independent random vectors and $f_i$ is a deterministic function. Here $\mathbf{M}$ represents "systematic" risk factors common to all firms, while $\mathbf{Y}_i$ represents "idiosyncratic," or firm-specific risk factors. A simple one-factor structure would be

$$\tau_i = F_i^{-1}\left(H_i\left(X_i\right)\right) \tag{1.2}$$

$$X_i = \theta_i M + \sqrt{1-\theta_i^2}Y_i \tag{1.3}$$

where $\theta_i$ is a constant with $|\theta_i| \leq 1$, $H_i$ is the distribution function of $X_i$ and $F_i$ is an ascribed marginal distribution for $\tau_i$. By varying the distribution of the factors one can typically obtain reasonable fits to index tranche quotes for a particular maturity. Hull and White [74] obtain a satisfactory fit to iTraxx index quotes using using constant $\theta_i$ and a $t$-distribution with four degrees of freedom for the factors. Kalemanova et al. [79] compare five different one-factor models - Gaussian, $t$ with 3 and 4 degrees of freedom, and two different normal inverse Gaussian models. They find that the normal inverse Gaussian provides a somewhat better fit than the $t$ models, and that the $t$ models require nearly ten times as much computational time. Finally, Fabozzi et al. [51] consider a one-factor model where the factors have a distribution which is a mixture of a standard normal and $t$ with fractional degrees of freedom. They carry out an extremely comprehensive calibration exercise, and find that this model does a very good job of fitting index tranche quotes.

Andersen and Sidenius [6] investigate extensions of the basic structure (1.2),(1.3) and their implications for CDO valuation. In particular they allow recovery rates to be correlated with the factors, in addition to allowing for "random factor loadings." They find the impact of random recovery rates to be minor, however they find that models of the form

$$X_i = \theta_i(M)M + \sigma_i Y_i - c_i$$

can provide quite significant correlation skews. Here

$$\sigma_i = \sqrt{1 - Var(\theta_i(M)M)} \qquad\qquad c_i = E\left[\theta_i(M)M\right]$$

Introducing the function $\theta_i(M)$ allows for stochastic correlation between defaults, with correlations being stronger in bear markets (small $M$) than in bull markets (large $M$). The specific choice of $\theta_i$ used in that paper is of a two-point form

$$\theta_i(M) = \begin{cases} \alpha_i & M \leq m_i \\ \beta_i & M > m_i \end{cases}$$

When the portfolio is homogeneous, useful asymptotic approximations have been obtained in a wide variety of copula and factor models. These results typically focus on the large $N$ asymptotics of the proportion of defaults

$$\frac{1}{N}\sum_{i=1}^{N} I\left(\tau_i \leq t\right) \tag{1.4}$$

in particular the identification of its distribution. Vasicek [109] appears to be the first member of the finance community to embark on such an endeavour, obtaining the distribution of (1.4) in the context of the equicorrelated Gaussian copula model. Schönbucher [104] investigates the asymptotic distribution in the context of various Archimedean copula families, while O'Kane and Schloegl [92] obtain it for the $t$ copula. Similar results can be obtained for multi-factor models, and even for heterogeneous portfolios under certain conditions, see Frey and McNeil [59] for example. In Section 4.1.1 we provide rather general conditions for existence of this limit in *any* model with conditionally independent defaults.

Factor models enjoy widespread popularity among practitioners, mainly due to their computational efficiency. The key to this efficiency is that, conditional upon the systematic factors, default times are independent. There are by now well-established techniques for evaluating the portfolio loss distribution which exploit this conditional independence structure, see for example Hull and White [74] or Andersen and Sidenius [6]. The primary deficiency of factor models is their lack of time dynamics. The economic environment in which obligors operate is ostensibly determined "today" and is not permitted to evolve over time. This is not consistent with empirical observations that both default rates and correlations are significantly affected by the state of the economy. See Andersen [5] for a discussion of this issue as well as some potential extensions of this framework incorporating time-dynamics of the underlying factors.

### 1.1.3 Intensity-Based Models

In the seminal credit risk model, namely the Black-Cox model to be discussed in the next section, the default time is *predictable* with respect to the filtration available to investors. Predictability of default is perceived as a major shortcoming here for several reasons. From a practical perspective this tends to produce corporate bond spreads which are unrealistically low for very short maturities. From an intuitive perspective, default can often come as a complete surprise (recall Enron and WorldCom) to market participants.

Intensity-based models draw on the general theory of counting processes and focus on the "hazard rate" of the process $N_t = I(\tau \leq t)$. In particular the intensity $\lambda(s)$ is often taken to be stochastic, typically as a function of some underlying process. The seminal paper in this area is Duffie and Singleton [45], who develop closed-form expressions for single-name defaultable contingent claims. In a multivariate setting the Cox process framework is often used, that is to say that conditional on some underlying covariate process $X_t$ the individual processes $N_t^i$ are independent, with jumps occurring at the first event time in a Poisson process with individual intensity of the form $\lambda_i(t, X_t)$. See Lando [85] for an early example using this approach. Duffie et al. [43] conduct an interesting empirical study of the applicability of this "doubly-stochastic" framework, while Bielecki and Rutkowski [18] provide an excellent and comprehensive introduction to the intensity framework in general.

Several authors have also investigated the use of counting process theory to model portfolio loss directly. In this "top-down" approach the focus is shifted from the characteristics of individual names to overall behaviour of portfolio losses. An interesting example here is Errais et al. [50] who treat portfolio loss as a marked point process, with marks representing random recovery. The intensity of the default process $D_t = \sum_{i=1}^{N} I(\tau_i \leq t)$ is of the form

$$\lambda_t = c_t + \delta \int_0^t e^{-\kappa(t-s)} dD_s$$

where $c$ is a deterministic function (the first-to-default intensity) and $\delta, \kappa$ are non-negative constants. In this way the default process becomes "self-exciting," in the sense the default of one name has an immediate impact (which dies out over time) on the default intensity of other names.

A natural advantage of this framework, as compared to copula and factor models, is that they *do* allow for significant time dynamics of default rates and corre-

lations. For example Errais et al. [50] find that their Hawkes process framework reproduces the tendency of defaults to cluster in time.

### 1.1.4 Structural and First-Passage Models

Structural models of credit risk occupy a special place in the hearts of both academics and practitioners. The primary reason for this affection appears to be the fact that these models provide a concrete link between the default event and economic fundamentals such as the value of a firm, its capital structure and the complex negotiations between debt and equity-holders. This is in contrast to alternative approaches such as factor or intensity models, where default times and intensities are simply treated as variables or processes to be modeled. While these approaches are able to incorporate macroeconomic and firm-specific covariates influencing default rates, these covariates influence the *likelihood* of the default event as opposed to precipitating the event directly.

Merton [91] is widely acknowledged as having pioneered this approach in what many consider to be the first recognizable credit risk model. The motivation for this seminal work was to develop a theory "for pricing bonds when there is a significant probability of default." Merton considers a firm whose outstanding debt consists of a zero-coupon bond with fixed maturity $T$. The value of a firm's assets, say $S_t$, is assumed to follow a geometric Brownian motion. One may think of this asset value as the liquidation value of the firm, should it choose to cease operations at time $t$. The firm is decreed to be in default at time $T$ if $S_T$ is less than the face value of debt. Simply put, the firm defaults if it cannot pay back its obligations.

A more general framework, which reflects the fact that firms may default at any time prior to the maturity of their debt, was put forth by Black and Cox [20]. These authors retained the assumption that the firms's asset value $S_t$ follows a geometric Brownian motion, but assume that corporate debt contains a safety covenant which allows bondholders to force bankruptcy and liquidation in the event of poor performance. The criteria for poor performance is given by a "default threshold" $B_t = Ke^{\lambda t}$ where $K < S_0$, and bondholders may force bankruptcy at the moment that $S_t$ falls below $B_t$. Thus the default time is given by the first passage time

$$\tau = \inf\{t \geq 0 : \ S_t \leq B_t\}$$

Mathematically default occurs upon first passage of a standard Brownian motion to a linear barrier.

There have been a staggering number of extensions of this basic framework. As we cannot possibly hope to do justice to this entire body of work, we mention that excellent references can be found in Duffie and Lando [44], Ericsson and Reneby [24], Embrechts et al. [48] and Bielecki and Rutkowski [18].

Several extensions of the Black-Cox framework have proceeded along "economic" grounds, with emphasis placed on theoretical underpinnings of the model. For instance Leland and Toft [86] retain the assumption of geometric Brownian motion for the firm value, and consider the optimal level and maturity of debt, as well as the optimal default threshold, from the shareholders' perspective. These quantities are determined with consideration of parameters such as bankruptcy costs and the tax shield on corporate interest payments. Collin-Dufresne and Goldstein [33] retain the geometric Brownian motion assumption for the firm's asset value and model the log-default threshold via

$$db_t = \lambda \left( s_t - \nu - b_t \right) dt$$

where $s_t$ is the logarithm of firm value. Interpreting the default threshold as being related (if not exactly equal) to the outstanding value of the firm's debt, this specification accounts for the fact that firms often have target leverage ratios (represented by the parameter $\nu$ here) and adjust debt levels accordingly over time. Huang et al. [71] extend this approach by allowing for a target leverage ratio which is time-dependent. As a final note, Yildirim [110] retains all assumptions of the Black-Cox model, but redefines the default event using an "occupational time," rather than first-passage, approach. In this approach default does not occur until firm value has spent a fixed amount of time below the default threshold, and the default time is defined as

$$\tau = \inf \left\{ t \geq 0 : \int_0^t I \left( S_u \leq B_u \right) du \geq c \right\}$$

where $c$ is a positive constant. This specification reflects the fact that, when a firm defaults on its obligations, it may often apply for bankruptcy protection. When a successful application is made, the firm is given a fixed amount of time to "right the ship," so to speak, before being forced into liquidation.

It is worth noting that the assumption that discounted firm value follows a martingale under the risk-neutral measure is virtually ubiquitous in the literature. It can be argued that this implicitly treats firm value as a traded asset, which is tenuous at best. The only example we are aware of which does *not* make this assumption is Buffet [30], who models the profit rate of a firm directly and treats

equity as the traded asset. The author then derives a risk-neutral measure under which discounted equity is a martingale, but the value of the firm's assets is not. The reader with even a passing interest in structural modeling is strongly encouraged to explore this reference.

Other extensions have taken a more "distributional" approach, with the emphasis placed on more appropriate models for the asset value $S_t$. As with many models for financial assets it is typically held that the geometric Brownian motion assumption is inappropriate. Unfortunately, unlike stock prices, asset values are not directly observable and empirical evidence for use of a particular process is typically unavailable. Nonetheless it is a widely-held belief that many empirically observed phenomena for stocks should also hold for asset values. As such many models used for stock price processes have also been used for firm value processes in the Black-Cox framework, for example Fouque et al. [57] incorporate stochastic volatility in the firm value process. Zhou [112] models $S_t$ as a jump diffusion (the continuous part of which is still assumed geometric Brownian motion) and finds that credit spreads increase dramatically in the presence of unanticipated jumps. In particular spreads for short-term bonds are no longer negligible, which has long been a major criticism of the Black-Cox model. Other models for $S_t$ have been put forth with an emphasis on valuation of credit derivatives. For example Luciano and Schoutens [89] model firm value as the exponential of a variance-gamma process (successfully calibrating the model to credit default swap spreads), while Kuen et al. [84] investigate the pricing of credit default swaps in a Markov-modulated regime-switching model.

In the Black-Cox model investors have "full information" in the sense that the firm's asset value is fully observable. This leads to predictable default times, which is often perceived as a serious deficiency for reasons noted in Section 1.1.3. Several extensions of the model have proceeded along "informational lines." This approach, pioneered by Duffie and Lando [44], is to assume imperfect information on the part of investors. In the case of a fixed (and known) default boundary this amounts to working with a sub-filtration of that generated by the firm's continuous asset value process. Assuming that investors only observe firm value at discrete intervals (for example through financial statements), Duffie and Lando [44] determine that defaults in this case are unpredictable with respect to the investor filtration and derive the intensity of the default process $N_t = I(\tau \leq t)$. Several recent papers have investigated similar methods for modeling information reduction and resulting intensities of default times in more general structural models. For example Guo et al. [67] study the problem in the context of general diffusions for firm value, including

regime switching models, while Jarrow et al. [77] determine default intensities for diffusion models where firm value is only observed at certain first passage times. Giesecke [62] investigates the case where the (time-independent) default boundary is a random variable unobserved by investors.

The discussion thus far has focused on the univariate (i.e. single-name) case, as there has been much less attention paid to multivariate structural modeling. Hull et al. [72] and Overbeck and Schmidt [94] investigate the basic multivariate extension of the Black-Cox model, namely a model in which the asset value of firm $i$ follows a geometric Brownian motion

$$dS_t^i = S_t^i \left[ \mu_i dt + \sigma_i dW_t^i \right] \tag{1.5}$$

where the $W^i$ are correlated Brownian motions. Here firm $i$ defaults upon first passage of $S^i$ to the barrier $K_i e^{\lambda_i t}$. We note that Overbeck and Schmidt [94] incorporate deterministic time-changes to the firm value processes, in an effort to obtain exact calibration to default swap data. There are two distinct similarities between these papers. To begin, both use Euler-type Monte Carlo schemes of the type described in Section 2.2.2 for valuation of multiname derivatives. In such a scheme one discretizes the time horizon under consideration and treats the correlated processes as if they were independent within each small time interval. Though useful as a first approximation the accuracy of this method can require a very small time step, making implementation prohibitively expensive.[1] A second common thread is that both studies find that model prices are quite similar to those obtained from a Gaussian copula model. Hull et al. [72] find this for index tranches, while Overbeck and Schmidt [94] observe the same phenomenon for basket default swaps. These observations provided the motivation for our work in Chapter 2.

While several alternative multivariate structural models have appeared in the literature (see Luciano and Schoutens [89], Fouque et al. [58] or Hurd [75] for example), the authors are aware of only one working paper (Baxter [12]) in which a multivariate structural model calibrates successfully to index tranche spreads. Unfortunately the model does not appear to calibrate to various maturities simultaneously. This observation provided the motivation for the material in Chapter 4, which in turn spawned Chapter 3. Before proceeding with an outline of this thesis, we find it prudent to mention that while we are not aware of successful attempts to calibrate multivariate structural models across both tranches and maturities, we

[1] We note that there are several approaches to reducing the bias in such a scheme. For example Schevchenko [106] employs Fréchet bounds while Huh and Kolkiewicz [69] employ large deviation techniques.

*are* aware of such endeavours in the context of intensity models, for example see DiGraziano and Rogers [41].

## 1.2 Outline of Thesis

Our original goals in this thesis were motivated by the experiences of Hull et al. [72] and Overbeck and Schmidt [94], discussed in Section 1.1.4. In particular we were interested in developing more efficient methods for implementing the multivariate Black-Cox model, as well as acquiring a deeper understanding of the nature of dependence in the model. The relevant mathematical structure here is first passage times of correlated Brownian motion to fixed levels, a problem which is still very far from being solved. In Chapter 2 we begin with the very simplest case, namely two dimensions and zero drift. Analytic results are possible here, unfortunately several published formulae are incorrect. In addition to correcting these mistakes, we derive a (nearly) exact simulation scheme for sampling the first passage times. Our algorithm exploits several interesting features of planar Brownian motion and conformal local martingales. We also compile a modest body of evidence concerning the similarity of the copula of these passage times to the Gaussian copula.

In Chapter 3 we investigate more general first-passage probabilities for one-dimensional Brownian motion. For "sensible" functions $f$ and $g$ we define $\tau_{f,g}$ as the first passage time of the time-changed Brownian motion $W(g_t)$ to the barrier $f$ and study the functional

$$\Psi(f,g) = P(\tau_{f,g} \leq T)$$

establishing its continuity over a "reasonable" product of function spaces. We also investigate the random variable $L = \Psi(A,B)$ were $A$ is a stochastic process and $B$ a stochastic time change. We devise approximate simulation schemes, and methods for assessing their accuracy, for approximating expectations $E[h(L)]$ for bounded, continuous functions $h$. Our motivation here is two-fold. To begin this provides a method for approximating crossing probabilities of a time-changed Brownian motion $W(B_t)$ to a stochastic barrier $A_t$, under very minimal conditions on the barrier and time-change. In addition, in Chapter 4 we propose a model for credit risk where $L$ represents the proportion of defaults on an asymptotically large portfolio. Quantities of the form $E[h(L)]$ are then crucial for implementing the model, in particular for pricing tranches on CDOs.

In virtually any structural model for credit risk, default is given by the first passage time of a "credit quality" process $X_t$ to the level zero. That is, with $\tau$

representing the default time of a particular obligor, we have

$$\tau = \inf \{t \geq 0 : \ X_t \leq 0\} \tag{1.6}$$

For instance in the Black-Cox model $X_t = \log(S_t/B_t)$ and is often referred to as the "log-leverage" ratio of the firm. Indeed one may treat this abstract notion of credit quality as the starting point of the modeling process. This is precisely the approach we take in Chapter 4, where we propose a general multivariate first-passage framework which models the credit quality of an obligor as an explicit combination of systematic and idiosyncratic "factor processes." As such we feel it provides a natural dynamic extension of the widely popular factor models. In addition we are able to calibrate several versions of the model to market quotes for index tranches. The model calibrates quite well across both tranches and maturities simultaneously. We are particularly pleased with the fact that the model continues to calibrate well in today's "distressed" environment.

# Chapter 2

# First Passage Times for Correlated Brownian Motion to Fixed Levels

Recall that in the multivariate Black-Cox model the credit quality of firm $i$ is given by

$$dX_t^i = \mu_i dt + \sigma_i dW_t^i \qquad X_0^i = x_0^i > 0$$

where the $W^i$ are correlated Brownian motions with $Cov(W_t^i, W_t^j) = \rho_{ij} t$. In addition firm $i$ defaults upon first passage of $X^i$ to zero, and we denote this default time as

$$\tau_i = \inf\left\{t \geq 0 : X_t^i = 0\right\}$$

At the present time the joint distribution of the $\tau_i$ is unknown, as such the "standard" implementation of the model in the finance literature is an Euler-based Monte Carlo scheme of the sort described in Section 2.2.2. Using this approach one discretizes the time horizon under consideration, conditions on the values of the process at the endpoints of each interval and treats the resulting bridged processes as if they were independent. Though a very reasonable first approximation, the accuracy of such an approach requires a very fine time step, as such its implementation can become incredibly time-consuming.

Our original motivation in this context was to develop efficient alternatives to the standard Monte Carlo implementation of the model. In Section 2.1 we begin with the very simplest case, namely two dimensions and zero drift. In addition to correcting published formulae for the joint density of the hitting times and exit

location[1], we derive a (nearly) exact simulation scheme for sampling $(\tau_1, \tau_2)$. Our algorithm is "indirect" in a sense, exploiting the strong Markov property of Brownian motion as well as known results for conformal local martingales. Unfortunately we are not able to extend our results to the case of non-zero drift, however our algorithm provides exactly the right output (i.e. Girsanov factors) required for unbiased estimation of various quantities in the presence of drift.

## 2.1 Exact Simulation in Two Dimensions

The problem we are interested in solving in this section is as follows. Let $W(t) = (W_1(t), W_2(t))$ be a correlated two-dimensional Brownian motion started at the origin. That is, each of $W_1(t)$ and $W_2(t)$ is a standard Brownian motion, with

$$\text{Cov}\left(W_1(t), W_2(t)\right) = \rho t \qquad\qquad |\rho| < 1$$

Note that we ignore the trivial cases $|\rho| = 1$. Given positive numbers $a_1$ and $a_2$, our goal is to simulate the random variables $(\tau_1, \tau_2)$, where

$$\tau_i = \inf\left\{t > 0 : \ W_i(t) = a_i\right\}$$

Iyengar [76] has found the joint density of this pair,[2] which means that in principle we could sample from the joint distribution of $(\tau_1, \tau_2)$ as follows

1. Generate $\tau_1$ from its marginal distribution. Since

$$P\left(\tau_1 \leq s\right) = 2\Phi\left(-\frac{a_1}{\sqrt{s}}\right)$$

   this is easily accomplished using inverse transform.

2. Having generated $\tau_1 = s$, generate $\tau_2$ from its conditional density

$$P\left(\tau_2 \in dt | \tau_1 \in ds\right) = \frac{\sqrt{2\pi s^3}}{a_1} e^{a_1^2/2s} f(s,t)dt$$

   where

$$P\left(\tau_1 \in ds, \ \tau_2 \in dt\right) = f(s,t)dsdt$$

[1] By exit location we mean the point at which the two-dimensional process first exits the region $(-\infty, b_1) \times (-\infty, b_2)$

[2] The formulae below are corrected versions of those found in Iyengar [76]. Details can be found in Section B.4 of Appendix B.

Unfortunately the density $f(s,t)$ is rather complicated. For $s < t$ it is given by

$$\begin{aligned} f(s,t) &= \frac{\pi \sin\alpha}{2\alpha^2\sqrt{s}(t-s)\sqrt{t-s\cos^2\alpha}} \exp\left(-\frac{r_0^2}{2s}\frac{t-s\cos 2\alpha}{(t-s)+(t-s\cos 2\alpha)}\right) \\ &\quad \times \sum_{n=1}^{\infty} n \sin\left(\frac{n\pi(\alpha-\theta_0)}{\alpha}\right) I_{n\pi/2\alpha}\left(\frac{r_0^2}{2s}\frac{t-s}{(t-s)+(t-s\cos 2\alpha)}\right) \end{aligned}$$

while for $s > t$ we have

$$\begin{aligned} f(s,t) &= \frac{\pi \sin\alpha}{2\alpha^2\sqrt{t}(s-t)\sqrt{s-t\cos^2\alpha}} \exp\left(-\frac{r_0^2}{2t}\frac{s-t\cos 2\alpha}{(s-t)+(s-t\cos 2\alpha)}\right) \\ &\quad \times \sum_{n=1}^{\infty} n \sin\left(\frac{n\pi\theta_0}{\alpha}\right) I_{n\pi/2\alpha}\left(\frac{r_0^2}{2t}\frac{s-t}{(s-t)+(s-t\cos 2\alpha)}\right) \end{aligned}$$

where $I_\nu(\cdot)$ is the modified Bessel function (of the first kind) of order $\nu$, and

$$r_0 = \sqrt{\tfrac{a_1^2+a_2^2-2\rho a_1 a_2}{1-\rho^2}}$$

$$\theta_0 = \begin{cases} \pi + \tan^{-1}\left(\frac{a_2\sqrt{1-\rho^2}}{a_1-\rho a_2}\right) & a_1 < \rho a_2 \\ \frac{\pi}{2} & a_1 = \rho a_2 \\ \tan^{-1}\left(\frac{a_2\sqrt{1-\rho^2}}{a_1-\rho a_2}\right) & a_1 > \rho a_2 \end{cases}$$

$$\alpha = \begin{cases} \pi + \tan^{-1}\left(-\frac{\sqrt{1-\rho^2}}{\rho}\right) & \rho > 0 \\ \frac{\pi}{2} & \rho = 0 \\ \tan^{-1}\left(-\frac{\sqrt{1-\rho^2}}{\rho}\right) & \rho < 0 \end{cases}$$

As this joint density is rather unwieldy, the conditional density in the second step is not amenable to common simulation techniques such as inverse transform or acceptance-rejection. An interesting feature of this density[3] for positive values of $\rho$ is that it explodes along the line $s = t$. That is, $f(s,t) \to \infty$ as $|s-t| \to 0$ for positive values of $\rho$. This implies that for a given value $s$, the conditional density in the second step explodes at the point $t = s$, which creates problems for acceptance-rejection. Using inverse transform would require integration of the conditional density and the inversion of this integral, which is an imposing endeavour.

In light of the difficulties involved with the application of standard Monte Carlo techniques, we have developed a simulation algorithm which exploits several interesting properties of planar (i.e. two-dimensional) Brownian motion. Our algorithm

[3]See Section B.4 of Appendix B.

actually solves an alternative formulation of the problem, involving a standard (i.e. uncorrelated) Brownian motion $Z(t)$ started inside the wedge

$$C_\alpha = \left\{(r\cos\theta, r\sin\theta) \in \mathbb{R}^2 :\ r \geq 0,\ 0 < \theta < \alpha\right\}$$

The next section re-states the problem in the context of such a process, and provides a brief outline of our algorithm.

### Re-Stating the Problem

With the same notation as in the previous section, consider the transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$T(\mathbf{w}) = \begin{bmatrix} \frac{a_1 - \rho a_2}{\sqrt{1-\rho^2}} \\ a_2 \end{bmatrix} - \begin{bmatrix} \frac{1}{\sqrt{1-\rho^2}} & -\frac{\rho}{\sqrt{1-\rho^2}} \\ 0 & 1 \end{bmatrix} \mathbf{w} \tag{2.1}$$

where $\mathbf{w} \in \mathbb{R}^2$. This transformation is best understood by considering it as the composition of the three individual transformations. Details are provided in Appendix B. The crucial facts here are that under $T(\cdot)$

- The vertical line at $w_1 = a_1$ becomes the line $z_2 = z_1 \tan\alpha$. Note that when $\alpha = \pi/2$ this is simply the vertical line with $z_1 = 0$.
- The horizontal line at $w_2 = a_2$ becomes the horizontal axis $z_2 = 0$.
- The process $Z(t) = T\left(W(t)\right)$ becomes a standard planar Brownian motion started at $z_0 = (r_0\cos\theta_0, r_0\sin\theta_0)$. Note that $z_0 \in C_\alpha$.

It follows that

- $\tau_1$ is the first passage time of $Z(t)$ to the line $z_2 = z_1\tan\alpha$
- $\tau_2$ is the first passage time of $Z(t)$ to the horizontal axis $z_2 = 0$
- $\tau = \min\left(\tau_1, \tau_2\right)$ is the first passage time of $Z(t)$ to $\partial C_\alpha$, the boundary of the wedge.

A brief outline of our algorithm for simulating $(\tau_1, \tau_2)$ is as follows

1. Simulate the exit location $Z(\tau)$. The exact distribution of this random variable is easily determined for the special case $\alpha = \pi/2$, while the conformal invariance of planar Brownian motion allows the general case to follow easily.

2. Simulate the exit time $\tau$ conditional on the exit location. This requires the joint distribution of the pair $(\tau, Z(\tau))$, which has been found by Iyengar [76]. We use acceptance-rejection to generate from the conditional density of $\tau$ given $Z(\tau)$.

3. Use the fact that $\{Z(t+\tau) - Z(\tau) : t \geq 0\}$ is a Brownian motion independent of $\{Z(s) : 0 \leq s \leq \tau\}$ to generate $\tau' = \max(\tau_1, \tau_2)$. This will only require simulating from an inverse Gaussian distribution.

This algorithm produces an exact drawing from the joint distribution of $(\tau, \tau', Z(\tau))$, which determines an exact drawing from $(\tau_1, \tau_2)$.

Having reformulated the problem, we now turn to some known results concerning Brownian motion in the wedge. Throughout, $P^{z_0}$ will denote the probability measure associated with standard planar Brownian motion beginning at $z_0$. The following expression has been derived by several authors, including Buckholtz and Wasan [29], Iyengar [76] and Rebholz [99], and has apparently been available in the literature since at least 1894. If $z = (r\cos\theta, r\sin\theta)$ is a point in $C_\alpha$, then

$$P^{z_0}(\tau > t,\ Z(t) \in dz) = \frac{2r}{t\alpha} e^{-(r^2+r_0^2)/2t} \sum_{n=0}^{\infty} \sin\frac{n\pi\theta}{\alpha} \sin\frac{n\pi\theta_0}{\alpha} I_{n\pi/\alpha}\left(\frac{rr_0}{t}\right) dr d\theta \qquad (2.2)$$

Integrating (2.2) over $r$ and $\theta$ yields the distribution of the first exit time from the wedge

$$P^{z_0}(\tau > t) = \frac{2r_0}{\sqrt{2\pi t}} e^{-r_0^2/4t} \sum_{n \text{ odd}}^{\infty} \frac{1}{n} \sin\left(\frac{n\pi\theta_0}{\alpha}\right) \left[I_{(\nu_n-1)/2}(r_0^2/4t) + I_{(\nu_n+1)/2}(r_0^2/4t)\right] \qquad (2.3)$$

where $\nu_n = n\pi/\alpha$. Equation (2.3) is a corrected version of the formula appearing in Iyengar [76], which corrected one of two mistakes in Buckholtz and Wasan [29]. The correct derivation is carried out in Section B.2 of Appendix B. A nearly correct version also appears in Rebholz [99], while a correct version also appears in Zhou [111].

We note here that Bañuelos and Smits [9] provide an expression analogous to (2.3) for general cones in $\mathbb{R}^n$, and investigate its asymptotics as $t \to \infty$. Unfortunately the expression seems rather impractical, as it involves eigenfunctions and eigenvalues of the Laplace-Beltrami operator on the $(n-1)$-dimensional sphere. In addition these authors verify that the series (2.3) is uniformly convergent in $t$ on any interval $[\epsilon, \infty)$.

### 2.1.1 Distribution of $Z(\tau)$

In this section we derive the distribution of the random variable $Z(\tau)$. It will be profitable to think of this random variable in terms of its polar co-ordinates $(R(\tau), \Theta(\tau))$, where $R(t)$, $\Theta(t)$ denote the radial and angular parts of $Z(t)$, respectively. Note that the range of $R(\tau)$ is $[0, \infty)$ while the range of $\Theta(\tau)$ is the two-point set $\{0, \alpha\}$. When we refer to the distribution of $Z(\tau)$ or the exit location distribution we are referring to probabilities of the form

$$P^{z_0}\left(R(\tau) \in dr, \Theta(\tau) = \delta\right) \qquad\qquad r \in [0, \infty) \quad \delta \in \{0, \alpha\}$$

Iyengar [76] purports to have obtained the distribution, however the formula given there is incorrect and does not integrate to unity. We note that this result was merely presented as a side remark, and does not affect any of the main results of that paper. The remainder of this section is dedicated to determining the distribution of $Z(\tau)$ and discussing its simulation.

#### Exit Location for the Positive Quadrant

In this section we derive the exit location distribution of planar Brownian motion from the positive quadrant $C_{\pi/2}$. We will often find it more convenient to work in rectangular co-ordinates, in which case we will write $z_0 = (x_0, y_0)$. We will need the following preliminary result, which is a problem in Karatzas and Shreve [80].

**Lemma 2.1.1.** *Let $W(t)$ be a (one-dimensional) Brownian motion started at $x > 0$, and let $T_0$ denote the first passage time of $W$ to zero. Then*

$$P\left(W(t) \in dy,\ T_0 > t\right) = p(t; x, y) - p(t; x, -y)$$

*where*

$$p(t; x, y) = \frac{1}{\sqrt{2\pi t}} \exp\left(-(y-x)^2/2t\right) \qquad y > 0$$

*Proof.* Note first that $B(t) = x - W(t)$ is a Brownian motion beginning at zero. It is well known (see Shepp [105] for example) that the joint density of $B(t)$ and its running maximum $M^B(t)$ is given by

$$f(b, m) = \frac{2(2m-b)}{\sqrt{2\pi t^3}} \exp\left(-(2m-b)^2/2t\right) \quad b \leq m \quad m > 0$$

If $m^W(t)$ denotes the running minimum of $W(t)$, it is clear that $M^B(t) = x - m^W(t)$. Hence for $y > 0$ we have

$$\begin{aligned} P\left(W(t) \in dy,\ T_0 > t\right) &= P\left(W(t) \in dy,\ m^W(t) > 0\right) \\ &= P\left(W(t) - x \in dy - x,\ m^W(t) - x > -x\right) \\ &= P\left(x - W(t) \in x - dy,\ x - m^W(t) < x\right) \\ &= P\left(B(t) \in x - dy,\ M^B(t) < x\right) \\ &= \int_{x-y}^{x} f(x-y, z)dz \\ &= p(t; x, y) - p(t; x, -y) \end{aligned}$$

where we have used the fact that if $B(t) \in x - dy$, it must be the case that $M^B(t)$ is greater than $x - y$. □

We are now in a position to determine the exit distribution, which is given in the following

**Proposition 2.1.2.** *Let $R$, $\Theta$ denote the radial and angular parts of a planar Brownian motion $Z$, and let $\tau$ denote the first exit time of $Z$ from the interior of the positive quadrant. Then for $z_0 \in C_{\pi/2}$ we have*

$$P^{z_0}\left(R(\tau) \in dr,\ \Theta(\tau) = 0\right) = \frac{1}{\pi}\frac{y_0}{(r - x_0)^2 + y_0^2} - \frac{1}{\pi}\frac{y_0}{(r + x_0)^2 + y_0^2} \quad (2.4)$$

$$P^{z_0}\left(R(\tau) \in dr,\ \Theta(\tau) = \frac{\pi}{2}\right) = \frac{1}{\pi}\frac{x_0}{(r - y_0)^2 + x_0^2} - \frac{1}{\pi}\frac{x_0}{(r + y_0)^2 + x_0^2} \quad (2.5)$$

*Proof.* Note first that

$$\tau_i = \inf\left\{t > 0 :\ Z_i(t) = 0\right\}$$

Using Lemma 2.1.1 we have that for any $r > 0$

$$P^{z_0}\left(Z_1(t) \in dr,\ \tau_1 > t\right) = p(t; x_0, r) - p(t; x_0, -r)$$

Note also that $P^{z_0}(\tau_2 \in dt) = (y_0/t)p(t; 0, y_0)$. Now, since $\tau_2$ is independent of both

$\tau_1$ and $Z_1(t)$, we obtain

$$\begin{aligned}
P^{z_0}\left(R(\tau) \in dr, \Theta(\tau) = 0\right) &= P^{z_0}\left(Z_1(\tau_2) \in dr, \tau_2 < \tau_1\right) \\
&= \int_0^\infty P^{z_0}\left(Z_1(\tau_2) \in dr, \tau_2 < \tau_1 | \tau_2 \in dt\right) P^{z_0}\left(\tau_2 \in dt\right) \\
&= \int_0^\infty P^{z_0}\left(Z_1(t) \in dr, t < \tau_1 | \tau_2 \in dt\right) P^{z_0}\left(\tau_2 \in dt\right) \\
&= \int_0^\infty P^{z_0}\left(Z_1(t) \in dr, t < \tau_1\right) P^{z_0}\left(\tau_2 \in dt\right) \\
&= \left[\int_0^\infty \left[p(t; x_0, r) - p(t; x_0, -r)\right] (y_0/t) p(t; 0, y_0) dt\right] dr \\
&= \left[\frac{1}{\pi} \frac{y_0}{(r - x_0)^2 + y_0^2} - \frac{1}{\pi} \frac{y_0}{(r + x_0)^2 + y_0^2}\right] dr
\end{aligned}$$

which is the difference of two Cauchy densities. An analogous argument shows that

$$P^{z_0}\left(R(\tau) \in dr, \Theta(\tau) = \frac{\pi}{2}\right) = \left[\frac{1}{\pi} \frac{x_0}{(r - y_0)^2 + x_0^2} - \frac{1}{\pi} \frac{x_0}{(r + y_0)^2 + x_0^2}\right] dr$$

which concludes the proof. □

It is interesting to note that the probability $Z$ hits the vertical axis before it hits the horizontal axis is equal to $2\theta_0/\pi$, and does not depend on the initial radius. This can be seen by integrating (2.5) over $(0, \infty)$.

In the next section we find it profitable to express the exit location distribution in polar co-ordinates. Tedious manipulation of (2.4) and (2.5) yields

$$P^{z_0}\left(R(\tau) \in dr, \Theta(\tau) = 0\right) = \frac{2}{\pi r_0} \frac{(r/r_0)\sin(2\theta_0)}{\sin^2(2\theta_0) + [(r/r_0)^2 - \cos(2\theta_0)]^2} \quad (2.6)$$

$$P^{z_0}\left(R(\tau) \in dr, \Theta(\tau) = \frac{\pi}{2}\right) = \frac{2}{\pi r_0} \frac{(r/r_0)\sin(2\theta_0)}{\sin^2(2\theta_0) + [(r/r_0)^2 + \cos(2\theta_0)]^2} \quad (2.7)$$

**Exit Location for the General Wedge**

Recall that for complex $z$ and real $a$ we may define $z^a$ as $r^a e^{ia\theta}$, where $(r, \theta)$ are the radial and angular parts of $z$, respectively. Hence the complex mapping $z \mapsto z^{\pi/2\alpha}$ "folds the wedge up" into the positive quadrant, while the mapping $z \mapsto z^{2\alpha/\pi}$ "unfolds" the positive quadrant into the wedge. This suggests that in order to study the exit location of our planar Brownian motion $Z$, we consider the transformed process $Z^{\pi/2\alpha}$. Note that we are viewing $Z$ as a $\mathbb{C}$-valued process here, and some care is required in defining $Z^{\pi/2\alpha}$ in order to ensure it is a continuous process.

The proof of Theorem 2.1.3 demonstrates that, while not a Brownian motion itself, the process $Z^{\pi/2\alpha}$ is in fact the *time-change* of a Brownian motion. As such the exit location for the general wedge may be recovered from the exit location given by (2.6) and (2.7). Before proceeding with the statement and proof of Theorem 2.1.3 we collect several known facts concerning conformal local martingales, each of which may be found in Rogers and Williams [101]. Given a probability space $(\Omega, \mathcal{F}, P)$, we recall that a $\mathbb{C}$-valued process $Y = U + iV$ is said to be a conformal local martingale (CLM), relative to the filtration $\mathcal{F}_t$, if $U$ and $V$ are local martingales, relative to $\mathcal{F}_t$, such that $[U] = [V]$ and $[U, V] = 0$, where $[\cdot]$, $[\cdot, \cdot]$ denote quadratic variation and quadratic co-variation, respectively. The results used in the proof of Theorem 2.1.3 are as follows

(C1) **Analytic transformations preserve the CLM property.** *If $Y$ is a CLM and $f : \mathbb{C} \to \mathbb{C}$ is analytic, then $f(Y)$ is a CLM.*

(C2) **CLMs are time-changed Brownian motion.** *If $Y = U + iV$ is a CLM, then there exists a complex Brownian motion $B$ such that $Y(t) = B([U](t))$.* Note that $B$ is a Brownian motion relative to $\mathcal{G}_t = \mathcal{F}_{A_t}$, where

$$A_t = \inf \{s \geq 0 : [U](s) > t\}$$

(C3) *If $Z$ is a complex Brownian motion, then $\log(R(t)) + i\Theta(t)$ is a CLM, where $R$ and $\Theta$ denote the radial and angular parts of $Z$, respectively.*

We now proceed with the main result of the section

**Theorem 2.1.3.** *Let $Z$ be a planar Brownian motion beginning at $z_0 \in C_\alpha$. Then there exists a planar Brownian motion $B$, beginning at $b_0 = z_0^{\pi/2\alpha} \in C_{\pi/2}$, such that*

$$Z(\tau) = [B(\tau_B)]^{2\alpha/\pi}$$

*where*

- $\tau = \inf \{t \geq 0 : \ Z(t) \in \partial C_\alpha\}$
- $\tau_B = \inf \left\{t \geq 0 : \ B(t) \in \partial C_{\pi/2}\right\}$

*Proof.* Let $R(t), \Theta(t)$ denote the radial and angular parts of $Z(t)$, respectively. By (C3) we know that the process $U = \log(R) + i\Theta$ is a CLM. Using the properties of quadratic and co-variation, this means that for any real $\beta$ the process $\beta U =$

$\log\left(R^\beta\right) + i\beta\Theta$ is also a CLM. Finally, since $f(z) = e^z$ is analytic, (C1) allows us to conclude that the process

$$V = R^\beta \exp\left(i\beta\Theta\right)$$

is a CLM. Now, set $\beta = \frac{\pi}{2\alpha}$ and let $\Theta_V$, $R_V$ denote the angular and radial parts of $V$, respectively. Note that $V$ begins at the point $v_0 = z_0^{\pi/2\alpha} \in C_{\pi/2}$. In addition since $\Theta_V = \pi\Theta/2\alpha$ we have

$$\tau = \inf\left\{t \geq 0 :\ \Theta(t) \notin (0,\alpha)\right\} = \inf\left\{t \geq 0 :\ \Theta_V(t) \notin (0,\pi/2)\right\} =: \tau_V$$

Moreover we may recover $Z(\tau)$ from $V\left(\tau_V\right)$ via

$$R(\tau) = \left(R_V(\tau_V)\right)^{2\alpha/\pi} \qquad \Theta(\tau) = \frac{2\alpha}{\pi}\Theta(\tau_V) \tag{2.8}$$

Now, since $V$ is a conformal local martingale it follows from (C2) that there exists a Brownian motion $B$ (beginning at $b_0 = v_0$) such that $V = B([V])$. Thus $V$ and $B$ will trace out exactly the same path in the plane, indeed they will simply move along this path at different speeds. As such they will strike the boundary of $C_{\pi/2}$ at exactly the same point, that is $V(\tau_V) = B(\tau_B)$. Recalling (2.8) we obtain the desired result. □

Recall that if $X$ is a random variable with pdf $f_X(x)$, then for $\gamma \neq 0$ the pdf of $Y = X^\gamma$ is given by

$$f_Y(y) = \frac{y^{(1/\gamma)-1}}{\gamma} f_X(y^{1/\gamma}) \tag{2.9}$$

This observation provides us with the general exit location distribution, as seen in the following

**Corollary 2.1.4.** *Let $R$, $\Theta$ denote the radial and angular parts of a planar Brownian motion $Z$, and let $\tau$ denote the first exit time of $Z$ from the interior of $C_\alpha$. Then for $z_0 \in C_\alpha$ we have*

$$\begin{aligned} P^{z_0}\left(R(\tau) \in dr, \Theta(\tau) = 0\right) &= \frac{1}{\alpha r_0} \frac{(r/r_0)^{(\pi/\alpha)-1} \sin\left(\pi\theta_0/\alpha\right)}{\sin^2\left(\pi\theta_0/\alpha\right) + \left[(r/r_0)^{\pi/\alpha} - \cos\left(\pi\theta_0/\alpha\right)\right]^2} \\ P^{z_0}\left(R(\tau) \in dr, \Theta(\tau) = \alpha\right) &= \frac{1}{\alpha r_0} \frac{(r/r_0)^{(\pi/\alpha)-1} \sin\left(\pi\theta_0/\alpha\right)}{\sin^2\left(\pi\theta_0/\alpha\right) + \left[(r/r_0)^{\pi/\alpha} + \cos\left(\pi\theta_0/\alpha\right)\right]^2} \end{aligned}$$

*Proof.* In the notation of Theorem 2.1.3, substitute $b_0$ into (2.6) and (2.7) to obtain the distribution of $(R_B(\tau_B), \Theta_B(\tau_B))$. Since

$$(R(\tau), \Theta(\tau)) = \left(\left(R_B(\tau_B)\right)^{2\alpha/\pi}, 2\alpha\Theta_B(\tau_B)/\pi\right)$$

the result follows from setting $\gamma = 2\alpha/\pi$ in (2.9). □

We may also use Theorem 2.1.3 to demonstrate that the probability $Z$ hits a given line (passing through the origin) before it hits the horizontal axis does not depend on the initial radius.

**Corollary 2.1.5.** *Let $Z$ be a planar Brownian motion beginning at $z_0 \in C_\alpha$. Then the probability $Z$ hits the line $z_2 = z_1 \tan\alpha$ before it hits the horizontal axis is equal to $\frac{\theta_0}{\alpha}$.*

*Proof.* $Z$ hits the line before the axis if and only if $\Theta(\tau_\alpha) = \alpha$, where $\tau_\alpha$ denotes the first exit time of $Z$ from the wedge $C_\alpha$. By Theorem 2.1.3 we obtain

$$P^{z_0}\left(\Theta(\tau_\alpha) = \alpha\right) = P^{z_0^*}\left(\Theta(\tau_{\pi/2}) = \pi/2\right) = \frac{\theta_0}{\alpha}$$

where $z_0^* = z_0^{\pi/2\alpha}$. The result follows upon integrating (2.5) (with respect to $r$ over $(0,\infty)$. □

Before proceeding with simulation of $Z(\tau)$, we would like to point out that the distribution obtained in Corollary 2.1.4 can be used to obtain numerical solutions for the Dirichlet problem in the wedge. To this end suppose we seek a continuous function $u$, harmonic inside the wedge, which satisfies the boundary conditions $u(x) = f(x)$ for $x \in \partial C_\alpha$. Here $f : \partial C_\alpha \to \mathbb{R}$ is continuous. According to Karatzas and Shreve [80] $u$ has the stochastic representation

$$u(x) = E^x\left[f\left(Z_\tau\right)\right]$$

for any $x \in \overline{C_\alpha}$, provided of course $E^x\left[\left|f\left(Z_\tau\right)\right|\right] < \infty$. Thus the distribution obtained in Corollary 2.1.4 could be used to numerically ascertain the value of $u$ at an arbitrary point in the closure of our wedge.

### Simulating the Exit Location

In light of the relationship provided by Theorem 2.1.3, we may obtain the exit location of a Brownian motion $Z$ beginning at a point $z_0$ in the interior of $C_\alpha$ as follows - begin a Brownian motion $B$ at the point $b_0 = r_0^{\pi/2\alpha}\left(\cos(\pi\theta_0/2\alpha), \sin(\pi\theta_0/2\alpha)\right)$, run it until it exits the interior of the positive quadrant for the first time, and set

$$(R(\tau), \Theta(\tau)) = \left(\left(R_B(\tau_B)\right)^{2\alpha/\pi}, 2\alpha\Theta_B(\tau_B)/\pi\right)$$

Thus simulation of the general exit location only requires that we be able to simulate from the distribution given by (2.4) and (2.5). If $R$ and $\Theta$ have this distribution,

it is easily seen that we may generate $\Theta$ as follows

$$\Theta = \frac{\pi}{2} I\left(V \geq \frac{2\theta_0}{\pi}\right)$$

where $V$ is a random number.[4] Conditional on $\Theta = 0$ the distribution function of $R$ is given by

$$F_0(r) = P\left(R \leq r | \Theta = 0\right) = 1 + \frac{1}{\pi - 2\theta_0}\left[\tan^{-1}\left(\frac{r - x_0}{y_0}\right) - \tan^{-1}\left(\frac{r + x_0}{y_0}\right)\right]$$

It is shown in Section B.3 in Appendix B that the inverse of this function is given by

$$F_0^{-1}(u) = r_0\sqrt{\cos(2\theta_0) - \frac{\sin(2\theta_0)}{\tan\left((\pi - 2\theta_0)(u-1)\right)}}$$

Similarly the conditional distribution function of $R$, given $\Theta = \pi/2$, is given by

$$F_{\pi/2}(r) = P\left(R \leq r | \Theta = \pi/2\right) = 1 + \frac{1}{2\theta_0}\left[\tan^{-1}\left(\frac{r - y_0}{x_0}\right) - \tan^{-1}\left(\frac{r + y_0}{x_0}\right)\right]$$

and its inverse is given by

$$F_{\pi/2}^{-1}(u) = r_0\sqrt{-\cos(2\theta_0) - \frac{\sin(2\theta_0)}{\tan\left((2\theta_0)(u-1)\right)}}$$

These facts allow us to inverse transform to generate $R$, conditional on the value of $\Theta$. The following proposition summarizes the algorithm for generating from this distribution

**Proposition 2.1.6.** *Let $U$, $V$ be independent random numbers, and define*

$$\begin{aligned}
\Theta &= \frac{\pi}{2} \cdot I\left(V \geq \frac{2\theta_0}{\pi}\right) \\
R &= r_0\sqrt{\cos(2\theta_0) - \frac{\sin(2\theta_0)}{\tan\left((\pi - 2\theta_0)(U-1)\right)}} \cdot I\left(\Theta = 0\right) \\
&\quad + r_0\sqrt{-\cos(2\theta_0) - \frac{\sin(2\theta_0)}{\tan\left(2\theta_0(U-1)\right)}} \cdot I\left(\Theta = \pi/2\right)
\end{aligned}$$

*Then the distribution of $(R, \Theta)$ is given by (2.6) and (2.7).*

In order to simulate the exit location for a general wedge $C_\alpha$ and initial point $z_0$ we use the following recipe

[4] By a random number we mean a random variable uniformly distributed on the interval $(0, 1)$.

- Use Proposition 2.1.6 with parameters $r_0^{\pi/2\alpha}$ and $\pi\theta_0/2\alpha$ to generate variables $R^*$ and $\Theta^*$.
- Set $\Theta = 2\alpha\Theta^*/\pi$ and $R = (R^*)^{2\alpha/\pi}$.

Having generated the exit location, we now discuss conditional simulation of the exit time given the exit location. This requires the joint distribution of the triplet $(\tau, \Theta(\tau), R(\tau))$ which has been obtained by Iyengar.

### 2.1.2 Conditional Distribution of $\tau$

According to Iyengar [76] the joint distribution of $(Z(\tau), \tau)$ is given by

$$P^{z_0}\left(\tau \in dt,\ Z(\tau) \in dz\right) = \frac{1}{2}\frac{\partial}{\partial n} f(t, z, z_0)$$

where $f$ is the transition density

$$f(t, z, z_0)dz = P^{z_0}\left(\tau > dt,\ Z(t) \in dz\right)$$

and for a point $z$ on the boundary of the wedge

$$\frac{\partial}{\partial n} f(t, z, z_0)$$

denotes the derivative of $f$ with respect to $z$, in the direction of the inward normal to the boundary at the point $z$. Note that the inward normal $n$ is either $(0, 1)$ or $(\sin\alpha, -\cos\alpha)$, according as $z$ lies on the horizontal axis or line $z_2 = z_1 \tan\alpha$. For a point $z = (r, 0)$ on the horizontal axis it follows easily that

$$P^{z_0}\left(\tau \in dt,\ R(\tau) \in dr,\ \Theta(\tau) = 0\right) = \frac{\pi}{\alpha^2 tr} e^{-(r^2+r_0^2)/2t} \sum_{n=0}^{\infty} g_n\left(r_0, \theta_0, \alpha, r\right)\ dr\ dt$$

where

$$g_n\left(r_0, \theta_0, \alpha, r, t\right) = n \sin\left(\frac{n\pi\theta_0}{\alpha}\right) I_{n\pi/\alpha}\left(\frac{rr_0}{t}\right) \tag{2.10}$$

For a point $z = (r\cos\alpha, r\sin\alpha)$ on the line $z_2 = z_1 \tan\alpha$ we have

$$P^{z_0}\left(\tau \in dt,\ R(\tau) \in dr,\ \Theta(\tau) = \alpha\right) = \frac{\pi}{\alpha^2 tr} e^{-(r^2+r_0^2)/2t} \sum_{n=0}^{\infty} g_n\left(r_0, \alpha - \theta_0, \alpha, r\right)\ dr\ dt$$

This latter formula can be obtained either by direct calculation, or by using the first formula and noting that the process $\tilde{Z}$ obtained by reflecting $Z$ through the line $z_2 = \tan(\alpha/2)z_1$ is also a planar Brownian motion. The radial components of the two

processes are equal and the angular parts are related by $\tilde{\Theta}(t) = \alpha - \Theta(t)$ and the exit times from the wedge are equal (i.e. $\tilde{\tau} = \tau$). With $\tilde{z}_0 = r_0(\cos(\alpha - \theta_0), \sin(\alpha - \theta_0))$ it immediately follows that

$$P^{z_0}\left(\tau \in dt,\ R(\tau) \in dr,\ \Theta(\tau) = \alpha\right) = P^{\tilde{z}_0}\left(\tau \in dt,\ R(\tau) \in dr,\ \Theta(\tau) = 0\right)$$

Using these expressions we see that the conditional distribution of $\tau$, given $Z(\tau)$ is given by

$$P^{z_0}\left(\tau \in dt | R(\tau) \in dr, \Theta(\tau) = 0\right) = \frac{\pi e^{-(r^2+r_0^2)/2t}}{c(\theta_0, r_0, r, 0)\alpha^2 t r} \sum_{n=0}^{\infty} g_n\left(r_0, \theta_0, \alpha, r, t\right)\ dr\ dt \tag{2.11}$$

$$P^{z_0}\left(\tau \in dt | R(\tau) \in dr, \Theta(\tau) = \alpha\right) = \frac{\pi e^{-(r^2+r_0^2)/2t}}{c(\theta_0, r_0, r, \alpha)\alpha^2 t r} \sum_{n=0}^{\infty} g_n\left(r_0, \alpha - \theta_0, \alpha, r\right)\ dr\ dt \tag{2.12}$$

where

$$c(\theta_0, r_0, r, 0) = P^{z_0}\left(R(\tau) \in dr,\ \Theta(\tau) = 0\right)$$

and $c(\theta_0, r_0, r, \alpha)$ is defined analogously.

Now suppose that we have simulated the exit location $Z(\tau)$ and obtained the values $R(\tau) = r$ and $\Theta(\tau) = \delta$. Suppressing its dependence on all other parameters, let $f(t)$ denote the associated conditional density of $\tau$, given these values. As there is no analytic expression for the integral of this density (or its inverse), conditional simulation via inverse transform is not an option. Recall that if we can find a density $g(t)$ (supported on $(0, \infty)$) which is easy to simulate from, and for which

$$\sup_{t \geq 0} \frac{f(t)}{g(t)} < \infty$$

then we may simulate observations from $f(t)$ using the acceptance-rejection algorithm. This algorithm would begin by first determining an upper bound on the ratio, that is a value of $c$ such that $f(t) \leq cg(t)$ for all $t$, and then proceeding as follows

- Simulate a random variable $T$ from the density $g$
- Simulate a random number $U$, independent of $T$
- If $U \leq f(T)/cg(T)$, set $\tau = T$. Repeat otherwise.

In the present situation $f(t)$ is only expressible as a series and this creates a slight problem in the last step, namely computing the ratio $f(t)/g(t)$. We have found

that this series tends to converge quite rapidly, as such we use an approximate acceptance-rejection scheme by replacing $f(t)$ with $f_N(t)$. Here $f_N(t)$ denotes the function obtained by summing the first $N$ terms in the series representation of $f(t)$.

The only remaining issue now is identifying densities $g$ which dominate $f_N(t)$ in both tails, and for which reasonable approximations to the upper bound of the ratio $f_N(t)/g(t)$ can be determined. In what follows we use the following asymptotics for Bessel functions, each of which can be found in Abramowitz and Stegun [2]

$$\begin{aligned} I_\nu(z) &\sim (z/2)^\nu/\Gamma(\nu+1) && z \to 0 \\ I_\nu(z) &\sim \tfrac{1}{\sqrt{2\pi z}} e^z && z \to \infty \end{aligned}$$

Our first candidate density is that of the random variable $T = X^{-1}$, where $X$ is a gamma variate. This candidate works quite well for simulated values of $R(\tau) = r$ for which $r$ is near $r_0$. In addition we are able to derive what we believe to be an explicit upper bound on the ratio $f_N(t)/g(t)$. In Proposition 2.1.7 we establish that the ratio *is* bounded, which clearly only requires the ratio to have a finite limit in both tails (i.e. as $t \to 0$ and $t \to \infty$). In addition, Proposition 2.1.7 implicitly assumes that the simulated value for $\Theta(\tau)$ is zero.

**Proposition 2.1.7.** *Let $g(t)$ be the inverse gamma density*

$$g(t) = \frac{1}{\Gamma(\beta)\lambda^\beta} \frac{1}{t^{\beta+1}} e^{-1/t\lambda}$$

*Then for each $N$*

*(i) For $\lambda \geq \frac{2}{(r-r_0)^2}$ and $\beta \leq \pi/\alpha$ we have*

$$\lim_{t\to 0} \frac{f_N(t)}{g(t)} = 0 \qquad \lim_{t\to\infty} \frac{f_N(t)}{g(t)} < \infty$$

*(ii) For $\lambda = \frac{2}{(r-r_0)^2}$ and $\beta = \pi/\alpha$ we have*

$$\lim_{t\to\infty} \frac{f_N(t)}{g(t)} = \left(\frac{r_0}{r-r_0}\right)^{2\pi/\alpha} \left[\sin^2(\pi\theta_0/\alpha) + \left((r/r_0)^{\pi/\alpha} - \cos(\pi\theta_0/\alpha)\right)^2\right]$$

*Proof.* The tail behaviour of the ratio is determined by terms of the form

$$t^\beta \exp\left(\frac{2-\lambda(r^2+r_0^2)}{2\lambda t}\right) I_{n\pi/\alpha}\left(\frac{rr_0}{t}\right)$$

As $t \to 0$ the first term tends to unity, while the product of the latter terms behave as

$$\begin{aligned} \exp\left(\frac{2-\lambda(r^2+r_0^2)}{2\lambda t}\right) I_{n\pi/\alpha}\left(\frac{rr_0}{t}\right) &\sim \exp\left(\frac{2-\lambda(r^2+r_0^2)}{2\lambda t}\right)\sqrt{\frac{t}{2\pi rr_0}}e^{rr_0/t} \\ &= \sqrt{\frac{t}{2\pi rr_0}}\exp\left(\frac{2-\lambda(r-r_0)^2}{2\lambda t}\right) \end{aligned}$$

These terms will explode if $2-\lambda(r-r_0)^2 > 0$ and tend to zero otherwise. Now as $t \to \infty$ the ratio is governed by the behaviour of

$$t^\beta I_{n\pi/\alpha}(rr_0/t)$$

the latter tending to zero and the former exploding. Asymptotically we have

$$\begin{aligned} t^\beta I_{n\pi/\alpha}(rr_0/t) &\sim t^\beta\left(\frac{rr_0}{2t}\right)^{n\pi/\alpha}\frac{1}{\Gamma((n\pi/\alpha)+1)} \\ &= \frac{(rr_0/2)^{n\pi/\alpha}}{\Gamma((n\pi/\alpha)+1)}t^{\beta-(n\pi/\alpha)} \end{aligned}$$

and these terms will explode unless $\beta \leq \pi/\alpha$. Now, setting $\beta = \pi/\alpha$ and $\lambda = 2/(r-r_0)^2$ the ratio becomes

$$\frac{\pi\Gamma(\pi/\alpha)2^{\pi/\alpha}}{r\alpha^2 c(\theta_0,r_0,r,0)(r-r_0)^{2\pi/\alpha}}t^{\pi/\alpha}\exp(-rr_0/t)\sum_{n=1}^{\infty} n\sin\left(\frac{n\pi\theta_0}{\alpha}\right)I_{n\pi/\alpha}\left(\frac{rr_0}{t}\right)$$

The ratio tends to zero as $t \to 0$, since

$$t^{\pi/\alpha}e^{-rr_0/t}I_{n\pi/\alpha}\left(\frac{rr_0}{t}\right) \sim \sqrt{\frac{t}{2\pi rr_0}}t^{\pi/\alpha}$$

As $t \to \infty$ the behaviour of the ratio is governed by the terms

$$\begin{aligned} t^{\pi/\alpha}I_{n\pi/\alpha}(rr_0/t) &\sim t^{\pi/\alpha}\left(\frac{rr_0}{2t}\right)^{n\pi/\alpha}/\Gamma((n\pi/\alpha)+1) \\ &= \frac{(rr_0)^{n\pi/\alpha}}{2^{n\pi/\alpha}\Gamma((n\pi/\alpha)+1)}t^{(1-n)\pi/\alpha} \end{aligned}$$

and we see that such terms tend to zero for $n \geq 2$ and the term for $n = 1$ tends to

$$\frac{(rr_0)^{\pi/\alpha}}{2^{\pi/\alpha}\Gamma((\pi/\alpha)+1)}$$

A little tedious algebra finally shows that

$$\lim_{t\to\infty}\frac{f_N(t)}{g(t)} = \frac{\sin(\pi\theta_0/\alpha)(rr_0)^{\pi/\alpha}}{r\alpha(r-r_0)^2 c(\theta_0,r_0,r,0)}$$

which can be further simplified to yield

$$\left(\frac{r_0}{r-r_0}\right)^{2\pi/\alpha}\left[\sin^2(\pi\theta_0/\alpha)+\left((r/r_0)^{\pi/\alpha}-\cos(\pi\theta_0/\alpha)\right)^2\right]$$

□

The efficiency of the acceptance-rejection algorithm is determined by the maximum value of the ratio $f_N(t)/g(t)$. Since $g$ depends on the parameters $\lambda$ and $\beta$, it follows that the efficiency of the algorithm depends on the chosen values for these parameters. In particular for maximum efficiency one should choose these parameters in such a way is to minimize the maximum value of $f_N(t)/g(t)$, a quantity which does not admit a closed form expression. Extensive numerical evidence (i.e. numerically approximating this quantity for a large number of parameter values) indicates that the maximum value of the ratio here is minimized when $\lambda$ is set equal to its lower bound and $\beta$ is set equal to its upper bound. The same evidence also indicates that in this case the ratio is increasing, which would imply that maximum efficiency in the conditional simulation is obtained by setting the rejection constant $c$ equal to the limit found in (ii). As a final note, Proposition 2.1.7 implicitly assumes the simulated value of the exit location lies on the horizontal axis, that is $\Theta(\tau)=0$. Thankfully the analogous result remains true when $\Theta(\tau)=\alpha$, we need only change the "-" to a "+" in front of the cosine in (ii).

Inspection of (ii) in Proposition 2.1.7 reveals that the rejection constant becomes quite large for values of $r$ near $r_0$. Indeed for parameter values $\theta_0=\pi/4$, $\alpha=3\pi/4$ and $r_0=1$ the constant exceeds 10 for $0.4<r<1.8$. This deficiency motivates our second choice for a dominating density, which tends to perform quite well precisely when $r$ is close to $r_0$.

**Proposition 2.1.8.** *Let $g(t)$ denote the "absolute Cauchy" density*

$$g(t)=\frac{1}{rr_0\pi}\frac{1}{1+(t/rr_0)^2}$$

*Then*

$$\lim_{t\to 0}\frac{f_N(t)}{g(t)}=\lim_{t\to\infty}\frac{f_N(t)}{g(t)}=0$$

*Proof.* Letting $z=rr_0/t$ we may express the ratio as

$$\frac{\pi^2}{c(\theta_0,r_0,r,0)\alpha^2 r}z\left[1+z^{-2}\right]\exp(-bz)\sum_{n=1}^{\infty}n\sin\left(\frac{n\pi\theta_0}{\alpha}\right)I_{n\pi/\alpha}(z)$$

where $b = \frac{r^2+r_0^2}{2rr_0} > 1$. Denoting this function by $h(z)$, we see that $z \to \infty$ we have

$$\begin{aligned} ze^{-bz}I_\nu(z) &\sim ze^{-bz}\frac{1}{\sqrt{2\pi z}}e^z \\ &= \sqrt{\frac{z}{2\pi}}e^{(1-b)z} \end{aligned}$$

This quantity tends to zero for $b > 1$, and it follows that $h(z) \to 0$ as $z \to \infty$. Now, as $z \to 0$ we have

$$\begin{aligned} z(1+z^{-2})I_\nu(z) &\sim (z+z^{-1})z^\nu \frac{1}{2^\nu \Gamma(\nu+1)} \\ &= \left[z^{\nu+1} + z^{\nu-1}\right] \frac{1}{2^\nu \Gamma(\nu+1)} \end{aligned}$$

which tends to zero provided $\nu > 1$, which is satisfied for $\nu = n\pi/\alpha$ and $n \geq 1$. Therefore

$$\lim_{z\to 0} z(1+z^{-2})I_{n\pi/\alpha}(z) = 0$$

for $n \geq 1$, and it follows that $h(z) \to 0$ as $z \to \infty$. □

Unfortunately we have not been able to derive an exact formula for the maximum of the ratio in this case, and in our simulations we typically use a value of 15. This is nearly twice what our numerical experimentation has indicated is acceptable.

We now have a (nearly) exact algorithm to simulate the triplet $(\tau, R(\tau), \Theta(\tau))$. Due to the Markov property of Brownian motion it is a simple matter to simulate $\tau' = \max(\tau_1, \tau_2)$ conditional on this triplet. This is discussed in the next section.

### 2.1.3 Distribution of $\max(\tau_1, \tau_2) - \min(\tau_1, \tau_2)$

In this section we show that the distribution of $\tau' - \tau$ is an inverse Gaussian mixture, where the mixing variable is $R(\tau)$. Recall that $\tau' = \max(\tau_1, \tau_2)$ and $\tau = \min(\tau_1, \tau_2)$. In addition we show how the joint density of $(\tau_1, \tau_2)$ may be obtained, with details provided in Section B.4 of Appendix B.

To begin we note that by the strong Markov property of Brownian motion, the process

$$\{Z(t+\tau) - Z(\tau) : t \geq 0\}$$

is a Brownian motion independent of $\{Z(s) : 0 \leq s \leq \tau\}$. Now suppose that $Z$ strikes the boundary of the wedge for the first time at the point $(r\cos\alpha, r\sin\alpha)$. In

this case we have that $X(t) = \{Z(t+\tau) : t \geq 0\}$ is a Brownian motion beginning at $(r\cos\alpha, r\sin\alpha)$, and that $\tau' - \tau = \tau_2 - \tau_1$ is simply the first passage time of $X$ to the horizontal axis. It is clear, then, that conditional on the this value of $Z(\tau)$ we have

$$\tau' - \tau = \inf\{t \geq 0 : X_2(t) = 0\}$$

which has an inverse Gaussian distribution with distribution function

$$F(t) = 2\Phi\left(-\frac{r\sin\alpha}{\sqrt{t}}\right)$$

Next suppose that $Z$ strikes the boundary of the wedge for the first time at the point $(r, 0)$, so that $X(t) = \{Z(t+\tau) : t \geq 0\}$ is a Brownian motion beginning at $(r, 0)$ and $\tau' - \tau = \tau_1 - \tau_2$ is the first passage time of $X$ to the line $x_2 = x_1\tan\alpha$. Letting $\tilde{X}$ denote the counter-clockwise rotation of $X$ by the angle $\pi - \alpha$, we see that $\tilde{X}$ is a Brownian motion beginning at $(-r\cos\alpha, r\sin\alpha)$. In addition, under this rotation the boundary $x_2 = x_1\tan\alpha$ becomes the horizontal axis, so that

$$\tau' - \tau = \inf\left\{t \geq 0 : \tilde{X}_2(t) = 0\right\}$$

which again has the inverse Gaussian distribution with distribution function

$$F(t) = 2\Phi\left(-\frac{r\sin\alpha}{\sqrt{t}}\right)$$

We have now demonstrated that

$$P^{z_0}\left(\tau' - \tau \leq t | \tau, R(\tau), \Theta(\tau)\right) = 2\Phi\left(-\frac{R(\tau)\sin\alpha}{\sqrt{t}}\right)$$

Therefore having generated $(\tau, R(\tau), \Theta(\tau))$ we may simulate $\tau' - \tau$ as

$$\tau' - \tau = \left(\frac{R(\tau)\sin\alpha}{\Phi^{-1}(U/2)}\right)^2$$

where $U$ is a random number independent of the triplet. We now have a complete algorithm for generating $(\tau_1, \tau_2)$ as follows

- Generate the pair $(R(\tau), \Theta(\tau))$.
- Conditional on this pair, generate $\tau$ using acceptance-rejection and $\tau' - \tau$ using inverse transform, and set $\tau' = (\tau' - \tau) + \tau$.
- We now have the four variables $(R(\tau), \Theta(\tau), \tau, \tau')$ which we use to obtain $(\tau_1, \tau_2)$ by setting

$$\begin{aligned}\tau_1 &= \tau' I\left(\Theta(\tau) = 0\right) + \tau I\left(\Theta(\tau) = \alpha\right)\\ \tau_2 &= \tau' I\left(\Theta(\tau) = \alpha\right) + \tau I\left(\Theta(\tau) = 0\right)\end{aligned}$$

We may also use the distribution of $\tau' - \tau$ derived in this section to derive an expression for the joint density of $(\tau_1, \tau_2)$. Details of this endeavour are carried out in Section B.4 of Appendix B

### 2.1.4 Expectations in the Presence of Drift

In this section we began with first passage times of a correlated Brownian motion $W_t$, and transformed the problem into one involving the first passage times of standard Brownian motion $Z_t = \mathbf{a} - AW_t$, where $\mathbf{a}$ and $A$ are given in (2.1). If the original process had drift $\gamma$, say, then the transformed process $Z = T(W)$ would be a standard Brownian motion with drift $\mu = -A\gamma$. Unfortunately the presence of drift destroys the conformal local martingale property of $Z$, and we are not able to extend our results to this more general case.

We might hope that, since a Girsanov factor provides the ratio between the two distributions, it is possible to simply simulate in the zero drift case and use acceptance-rejection to obtain samples from the non-zero drift distribution. Unfortunately for financial applications the original process will drift *away* from the barriers, meaning that the transformed process drifts away from the boundary of the wedge. That is, for virtually all financial applications we would have both components of $\mu$ being positive. To illustrate the problems this creates for acceptance-rejection, consider the ratio

$$\frac{P_{\mu}^{z_0}\left(\tau \in dt,\ R(\tau) \in dr,\ \Theta(\tau) = 0\right)}{P_{0}^{z_0}\left(\tau \in dt,\ R(\tau) \in dr,\ \Theta(\tau) = 0\right)} = \exp\left(\mu_1 r - \mu' z_0 - |\mu|^2 t/2\right) \tag{2.13}$$

For $\mu_1 > 0$ this ratio does not remain bounded over the region $[0, \infty) \times [0, \infty)$, and it is therefore not possible to generate the triplet in the non-zero drift case by generating the triplet $(\tau, R(\tau), \Theta(\tau))$ in the zero drift case, and accepting with a certain probability. A similar phenomenon holds for the ratio between the densities of $(\tau_1, \tau_2)$. These phenomena can be traced to the fact that if the process drifts away from a fixed level, it becomes possible that this level is never breached. As such the hitting times $(\tau_1, \tau_2)$ are defective in the sense that $P(\tau_i = \infty) > 0$.

Despite the fact that we may not use our zero-drift abilities to simulate exactly in the presence of drift, we *can* use these abilities to approximate expected values in the presence of drift. Indeed the outputs of our algorithm are exactly what we need in order to be able to simulate the Girsanov factor. To see this note that

$$E_{\mu}\left[f(\tau)\right] = E_0\left[\exp\left(\mu'(Z(\tau) - z_0) - |\mu|^2\tau/2\right) f(\tau)\right]$$

A similar expression would hold for functions $f(\tau, Z(\tau))$ of exit time *and* location. This would be very useful in the valuation of certain complex credit derivatives such as an option on a second-to-default swap (on two names) which can be exercised as soon as the first default occurs. Similarly, to evaluate expectations of functions $f(\tau_1, \tau_2)$ of both hitting times, we would have

$$E_\mu \left[ f(\tau_1, \tau_2) \right] = E_0 \left[ \exp \left( \mu'(Z(\tau') - z_0) - |\mu|^2 \tau'/2 \right) f(\tau_1, \tau_2) \right]$$

where $\tau'$ is the maximum of $\tau_1$ and $\tau_2$.

## 2.2 Approximation of the Survivor Function in More Than Two Dimensions

Unfortunately we have not been able to extend our results for exact simulation to more than two dimensions. As alluded to in the Introduction, however, the exact distribution (or ability to simulate from the exact distribution) of default times is not always required for credit risk applications. In this section we examine an approximation, devised by Bhansali and Wise [16], to the survivor function

$$P(t) = P\left( \tau_1 > t, \ldots, \tau_N > t \right)$$

where the $\tau_i$ are first passage times of correlated Brownian motion to fixed levels.

There are at least two problems for which a good approximation to $P(t)$ would be useful. For a homogeneous portfolio of $N$ names the portfolio loss is determined by the distribution of the number of survivors

$$S^N(t) = \sum_{i=1}^{N} I\left( \tau_i > t \right)$$

Known results for exchangeable Bernoulli variables (see [81] for example) can be used to express the distribution of this variate in terms of probabilities of the form

$$q(k) = P\left( \tau_1 > t, \ldots, \tau_k > t \right) \qquad k = 1, \ldots, N$$

Thus a valid approximation to the $q(k)$ would provide a much more efficient alternative to the standard Monte Carlo scheme. Unfortunately the expression relating the distribution of $S^N$ to the $q(k)$ can become numerical unstable as $N$ gets reasonably large, usually in the neighbourhood of about 30. Fortunately it is known (see [81]) that the limit $S(t) = \lim_{N \to \infty} S^N(t)$ exists almost surely and has a distribution

whose $k^{th}$ moment is given by $q(k)$. One might hope that good approximations to $q(k)$ would provide a reasonable approximation to the Laplace transform of $S(t)$, which could then be inverted to obtain an approximate distribution for $S^N(t)$ in cases where $N$ is large.

As a second application, we note that Brasch [22] shows how a $k^{th}$-to-default swap on a basket of $N$ names can be exactly replicated with first-to-default swaps on sub-baskets. Valuation of a first-to-default swap only requires knowledge of the distribution of the first default, and this is precisely $1 - P(t)$. Again this would provide an alternative to Monte Carlo valuation of basket default swaps, though it would only provide an *efficient* alternative for reasonably small baskets (the number of first-to-default swaps needed to replicate a $k^{th}$-to-default swap grows quite rapidly in both $k$ and $N$). We note that the approximation suggested by Bhansali and Wise [16] is not restricted to the homogeneous situation, so that it could be used to price swaps on baskets with arbitrary characteristics.

The basic model in this section is correlated Brownian motion - that is, a multivariate process

$$dX_t^i = \mu_i dt + \sigma_i dW_t^i$$

where the $W^i$ are standard Brownian motion with $Cov(W_t^i, W_t^j) = \rho_{ij} t$. We are interested in the minimum of the first passage times

$$\tau_i = \inf\left\{t \geq 0 : X_t^i = b_i\right\}$$

That is, we are interested in the random variable

$$\tau = \min\left\{\tau_1, \ldots, \tau_N\right\}$$

which is simply the first exit time of the multivariate process $X_t = \left(X_t^1, \ldots, X_t^N\right)$ from the interior of the region

$$B = (-\infty, b_1] \times \cdots \times (-\infty, b_N]$$

Without loss of generality we may assume that $\sigma_i = 1$ for all $i$, and note that the transition density of the process $X$, defined as

$$f\left(t, x_0, x\right) dx = P^{x_0}\left(X_t \in dx\right)$$

satisfies the forward equation

$$\left[\frac{\partial}{\partial t} + \sum_{i=1}^{N} \mu_i \frac{\partial}{\partial x_i} - \frac{1}{2} \sum_{i=1}^{N} \frac{\partial^2}{\partial x_i^2} - \sum_{i<j} \rho_{ij} \frac{\partial^2}{\partial x_i \partial x_j}\right] p(t, x_0, x) = 0$$

subject to the initial condition

$$f(0, x_0, x) = \prod_{i=1}^{N} \delta(x_0^i - x^i)$$

where $\delta(\cdot)$ is the Dirac delta function.

The approximation begins with the observation that the transition sub-density of the absorbed process $X_{t \wedge \tau}$, defined as

$$p(t, x_0, x)dx = P^{x_0}\left(X_t \in dx,\ \tau > t\right)$$

satisfies the *same* forward equation as $f$, subject to the same initial condition in addition to the boundary conditions

- $p(t, x_0, x) = 0$ for $x \notin B^o$, where $B^o$ denotes the interior of $B$.
- $p(t, x_0, x) \to 0$ as $x^i \to -\infty$ for each $i$.

If one knew the sub-density $p$, the survivor function of $\tau$ could be recovered as

$$P^{x_0}\left(\tau > t\right) = \int_B p(t, x_0, x)dx$$

Bhansali and Wise [16] perform a regular perturbation in each correlation parameter to obtain an approximation to the sub-density, which is then integrated to arrive the following approximation to the survivor function $P(t) = P^0\left(\tau > t\right)$

$$P(t) \approx \prod_{i=1}^{N} P_i(t) \left[1 + \frac{1}{2} \sum_{i \neq j} \left( \frac{P_{i,j}(t)}{P_i(t) P_j(t)} - 1 \right) \right]$$

where $P_i(t)$ is the marginal survivor function of $\tau_i$ and $P_{i,j}(t)$ is the bivariate survivor function of $\min\left(\tau_i, \tau_j\right)$. The beauty of this approximation is that semi-explicit expressions are available for each of these terms. With zero drift $P_{i,j}(t)$ can be obtained from (2.3) and will involve an infinite series of Bessel functions. The series tends to be rapidly convergent, with 10 to 15 terms typically sufficient, and its evaluation does not present a computational burden. With non-zero drift one must integrate a Girsanov factor against (2.3), which involves a series of double integrals of Bessel functions.

An alternative form for the approximation is given by

$$P(t) \approx Q(t) + \sum_{i<j} \left[Q_{i,j}(t) - Q(t)\right] \tag{2.14}$$

where

$$Q(t) = \prod_{i=1}^{N} P_i(t) \quad Q_{i,j}(t) = P_{i,j}(t) \prod_{k \neq i,j} P_k(t)$$

Note that $Q(t)$ would be the survivor function of $\tau$ if all components were uncorrelated, and $Q_{i,j}(t)$ would be the survivor function of $\tau$ if only component $i$ and $j$ were correlated. As Bhansali and Wise [16] (henceforth BW) are rather terse in their development, we now provide a brief outline of their argument.

### 2.2.1 Outline of Bhansali and Wise's Argument

For ease of exposition we will assume that $\mu_i = 0$ for each $i$, and note that the argument put forth here extends easily to the case of non-zero drift. We begin by noting that if $Z_t$ is a standard (i.e. uncorrelated) $N$-dimensional Brownian motion beginning at a point $x \in B^o$, then the transition sub-density of the absorbed process $Z_{t \wedge \tau}$ defined as

$$q(t, x, z)dz = P^x \left( Z_t^1 \in dz_1, \ldots, Z_t^N \in dz_N, \tau > t \right)$$

is equal to 0 for $z \notin B^o$ and is equal to

$$\prod_{i=1}^{N} \frac{1}{\sqrt{2\pi t}} \left[ e^{-(z_i - x_i)^2/2t} - e^{-(2b_i - z_i - x_i)^2/2t} \right]$$

for $z \in B^o$. This latter expression follows from Lemma 2.1.1 and the independence of the components. As discussed in the previous section, for a given $x \in B^o$ this function solves the equation

$$\left[ \frac{\partial}{\partial t} - \frac{1}{2} \sum_{i=1}^{k} \frac{\partial^2}{\partial z_i^2} \right] q(t, x, z) = 0 \tag{2.15}$$

subject to the initial and boundary conditions

- $q(0, x, z) = \prod_{i=1}^{N} \delta(z_i - x_i)$
- $q(t, x, z) = 0$ for $z \notin B^o$.
- $q(t, x, z) \to 0$ as $z_i \to -\infty$ for each $i$

It appears that we may interpret $q$ as the Green's function for the problem

$$\left[ \frac{\partial}{\partial t} - \frac{1}{2} \sum_{i=1}^{k} \frac{\partial^2}{\partial z_i^2} \right] g(t, z) = k(t, z) \quad (t, z) \in [0, \infty) \times B \tag{2.16}$$

subject to

- $g(0,x) = \prod_{i=1}^{N} \delta(x_i)$
- $g(t,x) = 0$ for $x \notin B^o$.
- $g(t,x) \to 0$ as $x_i \to -\infty$ for each $i$

We are led to believe that for a given function $k$ the solution to (2.16) satisfying the given initial and boundary conditions can be obtained via

$$g(t,x) = \int_0^t \int_B q(t-s,x,z)k(z,s)dzds$$

Returning to our original problem with correlated Brownian motion $X_t$, fix a pair $(i,j)$ with $i \neq j$ and assume that only these components are correlated. That is, assume that $\rho_{k\ell} = 0$ for $(k,\ell) \neq (i,j)$. In addition we assume that the process begins at the origin and define

$$p_{ij}(t,x) = P^0\left(X_t \in dx, \tau > t\right)$$

so that $p_{ij}$ satisfies

$$\left[\frac{\partial}{\partial t} - \frac{1}{2}\sum_{i=1}^{N}\frac{\partial^2}{\partial x_i^2} - \rho_{ij}\frac{\partial^2}{\partial x_i \partial x_j}\right] p_{ij}(t,x) = 0 \tag{2.17}$$

subject to the appropriate initial and boundary conditions. Expanding $p_{ij}$ in a perturbation series

$$p_{ij}(t,x) = p_{ij}^{(0)}(t,x) + \rho_{ij}p_{ij}^{(1)}(t,x) + \rho_{ij}^2 p_{ij}^{(2)}(t,x) + \ldots$$

and substituting into (2.17) we find that after equating powers of $\rho_{ij}$ the $p_{ij}^{\ell}$ are obtained recursively as

$$\begin{aligned}
\left[\frac{\partial}{\partial t} - \frac{1}{2}\sum_{k=1}^{N}\frac{\partial^2}{\partial x_k^2}\right] p_{ij}^{(0)}(t,x) &= 0 \\
\left[\frac{\partial}{\partial t} - \frac{1}{2}\sum_{k=1}^{N}\frac{\partial^2}{\partial x_k^2}\right] p_{ij}^{(1)}(t,x) &= \frac{\partial^2}{\partial x_i \partial x_j}p^{(0)}(t,x)
\end{aligned}$$

and in general

$$\left[\frac{\partial}{\partial t} - \frac{1}{2}\sum_{k=1}^{N}\frac{\partial^2}{\partial x_k^2}\right] p^{(\ell+1)}(t,x) = \frac{\partial^2}{\partial x_i \partial x_j}p^{(\ell)}(t,x)$$

again subject to the appropriate initial and boundary conditions. The "linear approximation" (more precisely, an approximation at linear order in $\rho_{ij}$) to $p_{ij}(t,x)$ is then

$$p_{ij}(t,x) \approx p_{ij}^{(0)}(t,x) + \rho_{ij} p_{ij}^{(1)}(t,x) \tag{2.18}$$

Based on our discussion at the beginning of this section we see that these first two terms are given by

$$p_{ij}^{(0)}(t,x) = q(t,0,x) \tag{2.19}$$

$$p_{ij}^{(1)}(t,x) = \int_0^t \int_B q(t-s,w,x) \frac{\partial^2}{\partial w_i \partial w_j} q(s,0,w) dw ds \tag{2.20}$$

It is interesting to note that the integrand in (2.20) can be expressed as the product

$$\left[ \prod_{k \neq i,j} \psi(t-s,w_k,x_k) \psi(s,0,w_k) \right] \psi(t-s,w_i,x_i) \psi(t-s,w_j,x_j) \frac{\partial \psi}{\partial w_i}(s,0,w_i) \frac{\partial \psi}{\partial w_j}(s,0,w_j)$$

where

$$\psi(t,w_\ell,x_\ell) = \frac{1}{\sqrt{2\pi t}} \left[ e^{-(x_\ell - w_\ell)^2/2t} - e^{-(2b_\ell - x_\ell - w_\ell)^2/2t} \right] = P^{w_\ell} \left( X_t^\ell \in dx_\ell,\ \tau_\ell > t \right)$$

One can verify with tedious calculation that

$$\int_{-\infty}^{b_\ell} \psi(t-s,w_\ell,x_\ell) \psi(s,0,w_\ell) dw_\ell = \psi(t,0,x_\ell)$$

Alternatively one can verify this by noting that

$$\int_{-\infty}^{b_\ell} \psi(t-s,w_\ell,x_\ell) \psi(s,0,w_\ell) dw_\ell = E \left[ P \left( X_t^\ell \in dx_\ell,\ \tau_\ell > t | X_s^\ell,\ \tau_\ell > s \right) \right]$$

All of this means that

$$p_{ij}^{(1)}(t,x) = \left[ \prod_{k \neq i,j} \psi(t,0,x_k) \right] K(t,x_i,x_j) \tag{2.21}$$

where $K(t,x_i,x_j)$ is given by

$$\int_0^t \int_{-\infty}^{b_i} \int_{-\infty}^{b_j} \psi(t-s,w_i,x_i) \psi(t-s,w_j,x_j) \frac{\partial \psi}{\partial w_i}(s,0,w_i) \frac{\partial \psi}{\partial w_j}(s,0,w_j) ds dw_i dw_j$$

We note at this point that one can obtain a linear (in $\rho_{ij}$) approximation to the survivor function $Q_{ij}(t)$ by integrating the linear approximation to the sub-density (2.18)

$$Q_{ij}(t) \approx \int_B \left( p_{ij}^{(0)}(t,x) + \rho_{ij} p_{ij}^{(1)}(t,x) \right) dx \tag{2.22}$$

$$= Q(t) + \rho_{ij} \int_B p_{ij}^{(1)}(t,x) dx \tag{2.23}$$

Unfortunately BW are a little unclear as to exactly how this simple method is extended to the more general situation of an arbitrary correlation structure. In the case of arbitrary correlation the transition sub-density $p(t,x)$ satisfies

$$\left[\frac{\partial}{\partial t} - \frac{1}{2}\sum_{i=1}^{N}\frac{\partial^2}{\partial x_i^2} - \sum_{i<j}\rho_{ij}\frac{\partial^2}{\partial x_i \partial x_j}\right] p(t,x) = 0 \tag{2.24}$$

subject to the appropriate boundary and initial conditions. They begin by "expanding the joint survival probability density [i.e. the sub-density of the absorbed process] in powers of the off diagonal correlation matrix elements" and write

$$p(t,x) = p^{(0)}(t,x) + p^{(1)}(t,x) + \ldots$$

with the statement "where $p^{(0)}(t,x)$ is the solution with the correlations set to zero, $p^{(1)}(t,x)$ contains all the terms linear in the off diagonal elements of the correlation matrix, etc." They then claim that the "contribution to the joint survival probability density that is linear in the asset correlations $p^{(1)}(t,x)$ satisfies the differential equation"

$$\left[\frac{\partial}{\partial t} - \frac{1}{2}\sum_{k=1}^{N}\frac{\partial^2}{\partial x_k^2}\right] p^{(1)}(t,x) = \sum_{i<j}\frac{\partial^2}{\partial x_i \partial x_j}p^{(0)}(t,x) \tag{2.25}$$

These characterizations of $p^{(0)}$ and $p^{(1)}$ appear to be based on a "formal" expansion of the form

$$p(t,x) = h^{(0)}(t,x) + \sum_{i<j}\rho_{ij}h_{ij}^{(1)}(t,x) + \sum_{i<j}\sum_{k<\ell}\rho_{ij}\rho_{k\ell}h_{ijk\ell}^{(2)}(t,x) + \ldots$$

If we insert this expression into (2.24) and equate coefficients we obtain the system of equations

$$\begin{aligned}
\left[\frac{\partial}{\partial t} - \frac{1}{2}\sum_{k=1}^{N}\frac{\partial^2}{\partial x_k^2}\right] h^{(0)}(t,x) &= 0 \\
\left[\frac{\partial}{\partial t} - \frac{1}{2}\sum_{k=1}^{N}\frac{\partial^2}{\partial x_k^2}\right] h_{ij}^{(1)}(t,x) &= \frac{\partial^2}{\partial x_i \partial x_j}h^{(0)}(t,x) \\
\left[\frac{\partial}{\partial t} - \frac{1}{2}\sum_{k=1}^{N}\frac{\partial^2}{\partial x_k^2}\right] h_{ijk\ell}^{(2)}(t,x) &= \frac{\partial^2}{\partial x_i \partial x_j}h_{k\ell}^{(1)}(t,x)
\end{aligned}$$

and this would imply that

$$\begin{aligned}
h^{(0)}(t,x) &= q(t,0,x) \\
h_{ij}^{(1)}(t,x) &= p_{ij}^{(1)}(t,x)
\end{aligned}$$

where $p_{ij}^{(1)}$ is given in (2.20). It is easily verified that if $p^{(1)} = \sum_{i<j} \rho_{ij} p_{ij}^{(1)}$ then $p^{(1)}$ satisfies (2.25). The linear approximation is therefore given by

$$p(t,x) \approx q(t,0,x) + \sum_{i<j} \rho_{ij} p_{ij}^{(1)}(t,x)$$

Integrating this over $B$ would provide what BW call a first-order approximation to the survivor function $P(t)$

$$P(t) \approx Q(t) + \sum_{i<j} \rho_{ij} \int_B p_{ij}^{(1)}(t,x)dx \tag{2.26}$$

which they compute by numerically integrating the individual terms given by (2.21). We make one small adjustment in the zero-drift case to eliminate the need for these integrals. Using (2.22) and (2.23) we note that

$$Q_{ij}(t) - Q(t) \approx \rho_{ij} \int_B p_{ij}^{(1)}(t,x)$$

Plugging this into (2.26) we obtain

$$P(t) \approx Q(t) + \sum_{i<j} [Q_{ij}(t) - Q(t)]$$

In the next section we provide selected results from extensive Monte Carlo investigation into the accuracy of this approximation.

### 2.2.2 Monte Carlo Assessment

In order to get some insight with respect to the accuracy of the approximation we compared (2.14) to a simulated empirical survivor function based on an Euler-type Monte Carlo scheme. For given correlation parameters and barrier levels we fixed a terminal time horizon $T$ and partitioned the interval $[0,T]$ into $M$ equally spaced intervals $[0,t_1],[t_1,t_2],\ldots,[t_{M-1},T]$ with $t_k = k\delta$ and $\delta = T/M$. We then generated an exact "skeleton path" $X_{t_1}, X_{t_2}, \ldots, X_T$ using the facts that the increments are multivariate normal and distinct increments are independent. Conditional on the values of the process at each of these times we then treated the processes

$$X^{i,k} = \left\{ X_u^i : u \in [t_{k-1}, t_k], X_{t_{k-1}}^i = x_{k-1}^i, X_{t_k}^i = x_{k-1}^i \right\}$$

as if they were independent. Note that each $X^{i,k}$ is a Brownian bridge with $X^{i,k}$ and $X^{i,\ell}$ being independent for $k \neq \ell$. In general $X^{i,k}$ and $X^{j,k}$ will *not* be independent for $i \neq j$, however if $M$ is large the dependence should be negligible. For each

$X^{i,k}$ we can generate a random variable $\tau^{i,k}$ having the same distribution as the first passage time to the level $b_i$ of a Brownian bridge from $x_{k-1}^i$ to $x_k^i$ over the interval $[0, \delta]$. Note that this distribution is defective in the sense that $\tau^{i,k} = \infty$ if the bridge does not attain the level $b_i$, and $P\left(\tau^{i,k} = \infty\right) > 0$. Proposition 3.18 in Section 3.3.1 illustrates that the cumulative distribution function for the first passage time to the level $b$ of a Brownian bridge from 0 to $y$ over the interval $[0, T]$ is given by

$$F(t) = 1 - \Phi\left(\frac{b - \frac{t}{T}y}{\sqrt{t\left(1 - \frac{t}{T}\right)}}\right) + \exp\left(-\frac{2b(b-y)}{T}\right)\Phi\left(\frac{\frac{t}{T}(b-y) - b\left(1 - \frac{t}{T}\right)}{\sqrt{t\left(1 - \frac{t}{T}\right)}}\right)$$

for $t \in [0, T]$. Simulation from this distribution using inverse transform is easily accomplished using Newton's method. The final step is to set

$$\tau^i = \min\left\{\tau^{i,k} : \ 1 \leq k \leq M\right\}$$

This algorithm provides one observation from an approximate distribution of the random vector

$$(\tau_1, \ldots, \tau_N)$$

and we repeat this procedure $n$ times. The marginal distributions of the resulting $\tau_i$ are correct.[5] Our aim is to compare the resulting empirical distribution function of

$$\tau = \min\{\tau_1, \ldots, \tau_N\}$$

to the approximation, and in order to ensure that we have chosen sufficiently large $M$ we compare empirical distributions of $\min\{\tau_i, \tau_j\}$ with the exact formulae $P_{i,j}(t)$.

Our results indicate that the approximation to the survivor function is remarkably accurate for reasonably small values of $N$, and quite robust against the correlation structure. In all cases discussed here we use $T = 10$, $M = 1000$ (so that $\delta = .01$) and 10,000 skeleton paths. Figure 2.1 shows the results for a correlation matrix with off-diagonal entries of $(.5, .11, .09)$ and barrier levels $(1, 1.1, 1.2)$. It is perhaps surprising that the approximation works this well when one of the correlations is as high as 50%. Figure 2.2 indicates that $M$ is sufficiently large that we are generating $\tau_{ij} = \min(\tau_i, \tau_j)$ from the correct distribution. The exact distributions plotted in this figure are given by (2.3).

Figure 2.3 illustrate the approximation breaking down as $N$ increases. Both plots are based on a homogeneous group with pairwise correlation of 25% and

[5]Note that we actually draw from the distribution of $\tau_i I(\tau_i \leq T) + \infty I(\tau_i > T)$.

Figure 2.1: Empirical and Approximate Distributions for $N = 3$

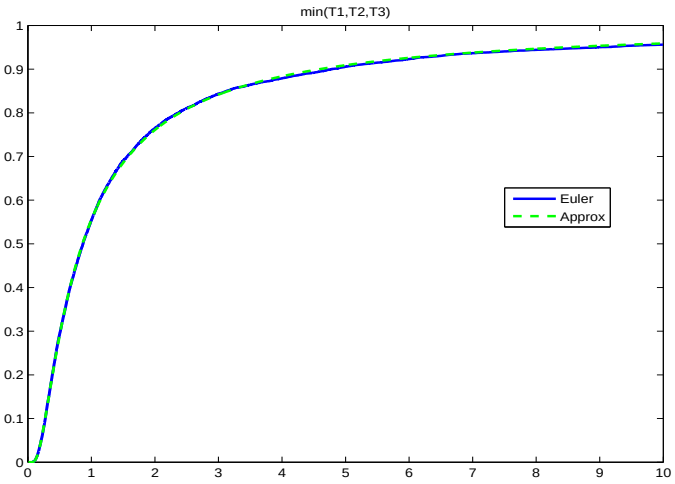



Figure 2.2: Empirical and Exact Distributions of $\tau_{ij} = \min(\tau_i, \tau_j)$ for $N = 3$

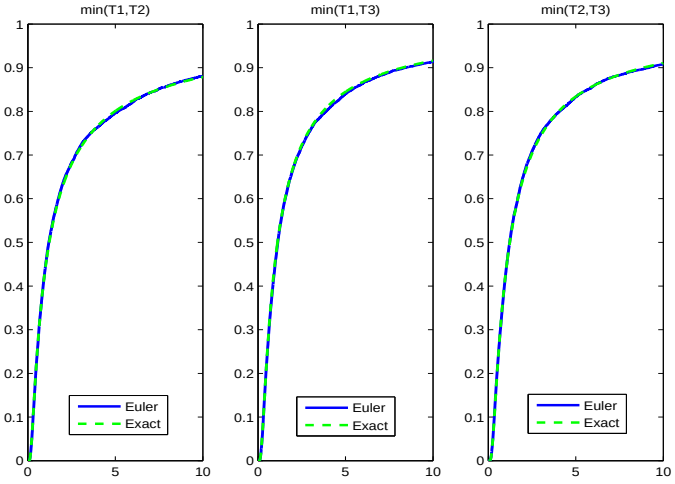
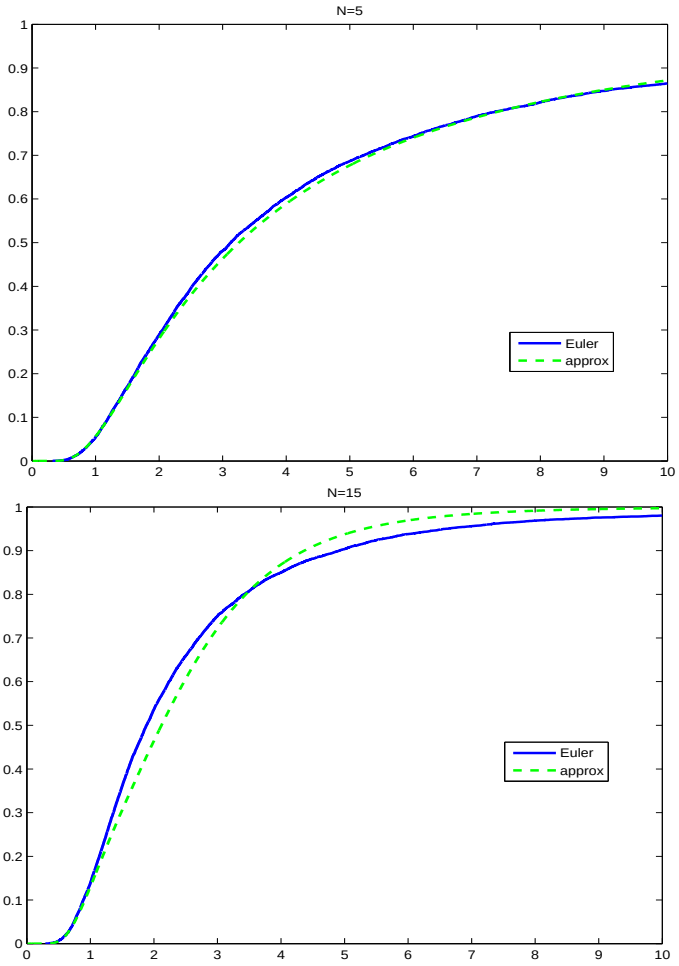


Figure 2.3: Empirical and Approximate Distributions for $N = 5$ and $N = 15$




individual barrier levels of 2.5. The approximation appears quite reasonable when $N = 5$, but when $N = 15$ it leaves something to be desired, and the situation only gets worse from there.

## 2.3 Similarity to the Gaussian Copula

Suppose that the copula of the pair $(\tau_1, \tau_2)$ were Gaussian, with correlation parameter $\theta$. Then the joint density of this pair would be given by

$$h(s,t) = \frac{\phi_\theta\left(\Phi^{-1}\left(F_1(s)\right), \Phi^{-1}\left(F_2(t)\right)\right)}{\phi\left(\Phi^{-1}\left(F_1(s)\right)\right)\phi\left(\Phi^{-1}\left(F_2(t)\right)\right)} f_1(s) f_2(t) \qquad s,t \geq 0 \tag{2.27}$$

where $f_i$ and $F_i$ are the density and distribution function of $\tau_i$, respectively, $\phi$ and $\Phi$ are the density and distribution function of a standard normal variate and $\phi_\theta$ is

the standard bivariate normal density

$$\phi_\theta(x,y) = \frac{1}{2\pi\sqrt{1-\theta^2}} \exp\left(-\frac{x^2+y^2-2\theta xy}{2\left(1-\theta^2\right)}\right) \qquad (x,y) \in \mathbb{R}^2$$

Since the actual joint density of $(\tau_1, \tau_2)$ explodes along the line $s = t$, and since (2.27) is bounded, it follows that the copula of $(\tau_1, \tau_2)$ is not Gaussian. However several authors have provided evidence that the true copula may in fact be quite similar to the Gaussian. For example Overbeck and Schmidt [94] find that basket swap spreads computed using first passage times of correlated Brownian motion are quite similar to those computed using a Gaussian copula. In addition McLeish [90] devises a bivariate normal approximation to the joint distribution of the maxima of two correlated Brownian motion which appears to be remarkably accurate. In this section we investigate this similarity in more detail.

### 2.3.1 Goodness-of-Fit in Two Dimensions

In this section we apply our ability to simulate the pair exactly to the problem of assessing the similarity between a Gaussian copula and the true copula of the pair $(\tau_1, \tau_2)$. We do so by applying a goodness-of-fit test to simulated observations from the exact joint distribution. The test was designed by Genest et al. [61] to test a hypothesis of the form

$$H_0 : C \in \mathcal{C} \tag{2.28}$$

where $C$ is the copula of a given pair of random variables $(X, Y)$ and

$$\mathcal{C} = \{C_\theta(u,v) : \theta \in \Theta\}$$

is a given parametric family of copulas. We will see that the test has a hard time rejecting this null hypothesis when $\mathcal{C}$ is the Gaussian family, and take these results as an indication that while the true copula is *not* Gaussian, it *is* similar enough to the Gaussian that this test has a hard time distinguishing between the two.

We will not provide details of the test, rather we will simply give an outline. Let $(X, Y)$ be a given pair of random variables with copula $C$ and marginals $F$ and $G$, and suppose we are interested in testing the hypothesis (2.28) for some given parametric family. To any parametric family $\mathcal{C} = \{C_\theta(u,v) : \theta \in \Theta\}$ there is an associated parametric family of univariate distribution functions $\mathcal{K} = \{K_\theta(t) : \theta \in \Theta\}$ where $K_\theta(t)$ is defined as via a bivariate probability integral transform

$$K_\theta(t) = P\left(C_\theta(U,V) \leq t\right)$$

and the pair $(U, V)$ have joint distribution $C_\theta$. The idea of Genest et al. [61] is to convert the two-dimensional goodness-of-fit test on the copula of $(X, Y)$ to a univariate test on the distribution of

$$T = H(X, Y) = C(F(X), G(Y))$$

Given independent observations $(X_1, Y_1), \ldots, (X_n, Y_n)$ from a distribution with copula $C$ and marginals $F$, $G$ they construct "pseudo-observations"

$$\hat{T}_i = \hat{H}(X_i, Y_i) = \hat{C}(\hat{F}(X_i), \hat{G}(Y_i))$$

where $\hat{H}$, $\hat{F}$ and $\hat{G}$ are the empirical bivariate and marginal distributions, and $\hat{C}$ is the empirical copula. Either a one-dimensional Kolmogorov-Smirnov or Cramer-von Mises test is then performed on the pseudo-observations to test the hypothesis that $K \in \mathcal{K}$. See Genest et al. [61] for an investigation of the asymptotic properties of the test statistics and the power of the test.

Table 2.1 presents $P$-values for the test obtained using different parameters values. In each case we simulated 1,000 observations from the exact distribution of $(\tau_1, \tau_2)$ and applied the test to the simulated population. The columns CVS and KS refer to $P$-values for the Cramer-von Mises and Kolmogorov-Smirnov statistics, respectively. The test requires an estimate of the copula parameter based on the simulated data, which is reported in the last column. Note that this is not an estimate of the correlation between the Brownian motions, rather it is an estimate of the copula parameter under the assumption that the copula is Gaussian. We see that in no cases would the test reject the hypothesis that the simulated values of $(\tau_1, \tau_2)$ were drawn from a distribution with a Gaussian copula. While we *know* that the true copula is *not* Gaussian, we see that it *is* similar enough that this test has trouble distinguishing between the two, even for as many as 1,000 observations. It is interesting to note that when the two barriers are very far apart, the dependence between the hitting times is not significantly affected by the correlation between the processes.

### 2.3.2 Tail Dependence in Two Dimensions

It is known that the Gaussian copula does not possess dependence in either its upper or lower tails. In this section we verify that the pair $(\tau_1, \tau_2)$ has no upper tail dependence.

To begin, let $\rho$ be the correlation between $(W^1, W^2)$ and assume for the moment that the barrier levels are equal, say $b_1 = b_2 = b > 0$. We recall that the marginal

Table 2.1: $P$-Values for Goodness-of-Fit Test

| $b_1$ | $b_2$ | $\rho$ | CVS | KS | $\hat{\theta}$ |
|---|---|---|---|---|---|
| 8 | 8 | .25 | .1710 | .1950 | .1651 |
| 8 | 8 | .50 | .3020 | .2960 | .4177 |
| 2 | 2 | .25 | .2140 | .3500 | .1984 |
| 2 | 2 | .50 | .1190 | .2030 | .4162 |
| 2 | 8 | .25 | .6490 | .7320 | .2467 |
| 2 | 8 | .50 | .5750 | .6380 | .2054 |

distribution of $\tau_i$ is $2\Phi\left(-b/\sqrt{t}\right)$ and that the survivor function of $\tau = \min(\tau_1, \tau_2)$ is given by the uniformly convergent (see Bañuelos and Smits [9]) series

$$P(\tau > t) = \frac{2r_0}{\sqrt{2\pi t}} e^{-r_0^2/4t} \sum_{n \text{ odd}}^{\infty} \frac{1}{n} \sin\left(\frac{n\pi\theta_0}{\alpha}\right) \left[I_{(\nu_n-1)/2}(r_0^2/4t) + I_{(\nu_n+1)/2}(r_0^2/4t)\right]$$

where $\alpha < \pi$, $r_0$ and $\theta_0$ are parameters depending on the correlation $\rho$ and the barrier level $b$.

The coefficient of upper tail dependence in this case is

$$\begin{aligned}
\lambda_U &= \lim_{u \nearrow 1} P\left(\tau_1 > F^{-1}(u) | \tau_2 > F^{-1}(u)\right) \\
&= \lim_{t \to \infty} \frac{P(\min(\tau_1, \tau_2) > t)}{P(\tau_1 > t)} \\
&= \lim_{t \to \infty} \frac{P(\tau > t)}{1 - F(t)}
\end{aligned}$$

For large $t$ the denominator behaves proportionally to $t^{-1/2}$, while the numerator behaves proportionally to $t^{-\pi/2\alpha}$. This means that the ratio behaves as $t^{(\alpha-\pi)/2\alpha}$, which tends to zero since $\alpha < \pi$. Therefore $\lambda_U = 0$ when the barriers are equal. To see these individual behaviours, first note that

$$\Phi(x) = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} x + o(x)$$

where $o(x)/x$ tends to zero as $x$ tends to zero. Therefore as $x \to 0$ we have

$$1 - 2\Phi(x) \sim -\sqrt{\frac{2}{\pi}} x$$

and this means that as $t \to \infty$ we have

$$1 - F(t) = 1 - 2\Phi\left(-b/\sqrt{t}\right) \sim \sqrt{\frac{2}{\pi}} \frac{b}{\sqrt{t}}$$

Next we show that

$$\lim_{t\to\infty} t^{\pi/2\alpha} P\left(\tau > t\right) = c$$

for a constant $c$, which will demonstrate that

$$P\left(\tau > t\right) \sim c t^{-\pi/2\alpha}$$

The survivor function of $\tau$ is a series containing terms of the form

$$e^{-r_0^2/4t} t^{-1/2} I_{(\nu_n \pm 1)/2}\left(r_0^2/4t\right)$$

with $\nu_n = n\pi/\alpha$. Multiplying terms by $t^{\pi/2\alpha}$ and ignoring the exponential terms (which tend to one as $t \to \infty$) we get a series with terms of the form

$$t^{(\pi-\alpha)/2\alpha} I_{(\nu_n \pm 1)/2}\left(r_0^2/4t\right)$$

Noting that

$$I_\nu\left(r_0^2/4t\right) \sim k_\nu t^{-\nu}$$

where $k_\nu$ is a positive constant depending on the order, we see that

$$t^{(\pi-\alpha)/2\alpha} I_{(\nu_n+1)/2}\left(r_0^2/4t\right) \sim k_{(\nu_n+1)/2} t^{-1-(n-1)\pi/2\alpha}$$

which tends to zero for all $n \geq 1$. Next we see that

$$t^{(\pi-\alpha)/2\alpha} I_{(\nu_n-1)/2}\left(r_0^2/4t\right) \sim k_{(\nu_n-1)/2} t^{-(n-1)\pi/2\alpha}$$

which is a constant for $n = 1$ and tends to zero for $n \geq 2$. All of this shows that, in the series for $t^{\pi/2\alpha} P\left(\tau > t\right)$, all terms except the first tend to zero as $t$ explodes, while the first term tends to a constant. That is

$$\lim_{t\to\infty} t^{\pi/2\alpha} P\left(\tau > t\right) = k$$

for some constant $k$ (which we can determine exactly), and therefore

$$P\left(\tau > t\right) \sim k t^{-\pi/2\alpha}$$

as $t \to \infty$.

Now in the situation where the barriers are different, say $b_1 < b_2$ we have

$$F_1(t) > F_2(t) \implies F_1^{-1}(u) < F_2^{-1}(u)$$

which implies

$$P\left(\tau_1 > F_1^{-1}(u),\ \tau_2 > F_2^{-1}(u)\right) \leq P\left(\tau_1 > F_1^{-1}(u),\ \tau_2 > F_1^{-1}(u)\right) = P\left(\tau > F_1^{-1}(u)\right)$$

And this means that

$$\begin{aligned}
\lambda_U &= \lim_{u \nearrow 1} P\left(\tau_2 > F_2^{-1}(u) | \tau_1 > F_1^{-1}(u)\right) \\
&= \lim_{u \nearrow 1} \frac{P\left(\tau_1 > F_1^{-1}(u),\ \tau_2 > F_2^{-1}(u)\right)}{P\left(\tau_1 > F_1^{-1}(u)\right)} \\
&\leq \lim_{u \nearrow 1} \frac{P\left(\tau > F_1^{-1}(u)\right)}{P\left(\tau_1 > F_1^{-1}(u)\right)} \\
&= 0
\end{aligned}$$

with the last equality following from our result when the barrier levels are equal. Since $\lambda_U \geq 0$ we conclude that there is no upper-tail dependence between the pair in any case.

We have had difficulty obtaining the analogous result in the case of lower tail dependence. However it has come to our attention through a private communication that Bo Shi, a graduate student at the University of Pittsburgh, has successfully established that the pair has no lower tail dependence.

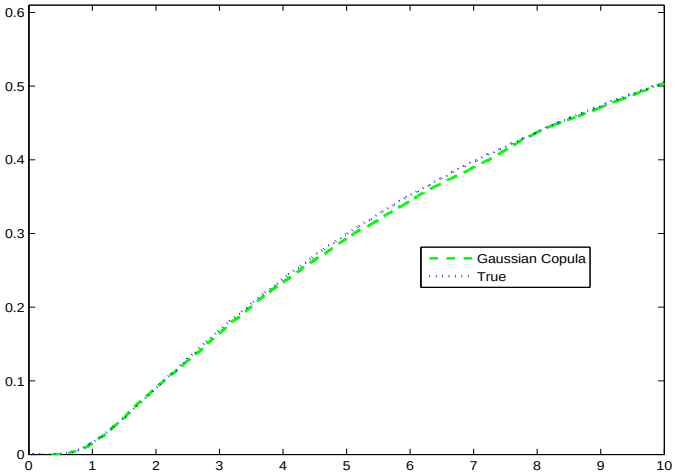
### 2.3.3 First-to-Default in Three Dimensions

We may also use this approximation to get further insight into the similarity between the copula of $(\tau_1, \ldots, \tau_N)$ and the Gaussian copula. To do so we will answer the following question - suppose that $Y_i$ has the same marginal distribution as $\tau_i$, but that the copula of $(Y_1, \ldots, Y_N)$ is Gaussian with correlation matrix $\hat{\Sigma}$. How similar are the distributions of $\tau = \min(\tau_1, \ldots, \tau_N)$ and $Y = \min(Y_1, \ldots, Y_N)$? Clearly the answer will depend on the Gaussian correlation matrix $\hat{\Sigma}$, and we choose to make the comparison in such a way as to match the bivariate survival probabilities over a fixed time horizon $T$. That is, we choose $\hat{\Sigma}_{ij}$ such that

$$P(\tau_i > T,\ \tau_j > T) = P(Y_i > T,\ Y_j > T)$$

Using $N = 3$ with correlations and barrier levels of (.5,.11,.09) and (5,8,2.4) for the Brownian motion, we obtain "matched" correlations of (.4873,.1011,.0813) which we use in the Gaussian copula. Figure 2.4 plots Bhansali and Wise's [16] approximation to the distribution of $\min(\tau_1, \tau_2, \tau_3)$ against a simulated empirical distribution of $(Y_1, Y_2, Y_3)$ where the $Y_i$ have the same marginal distributions as the $\tau_i$, but a Gaussian copula with the matched correlation parameters.

Figure 2.4: True vs. Gaussian Copula




As the two distributions are virtually indistinguishable this indicates that the first default time in a "correlated Brownian drivers" model is quite similar to the first default time in a Gaussian copula model. As noted in the introduction to this chapter the first default time is the fundamental object in many credit risk applications, indicating that one might expect reasonably similar results from the two models.

# Chapter 3

# Stochastic Barriers and Time-Changed Brownian Motion

The original impetus for the material presented in this chapter was to justify an approximation scheme used in Chapter 4. As we delved deeper into the subject we realized that the results presented here are of sufficient generality and independent interest to warrant their own chapter.

The problem we are interested in here is as follows. For fixed $T > 0$ suppose that $f : [0, T] \mapsto \mathbb{R}$ is a deterministic function with $f(0) > 0$. We define $\psi(f)$ as the probability that a standard Brownian motion breaches the (upper) barrier $f$ at some point over the interval $[0, T]$. Our interest lies in the random variable created by replacing the deterministic $f$ with a stochastic process $A$. That is, we are interested in the random variable $L := \psi(A)$, where $A = \{A_t : 0 \leq t \leq T\}$ is seen as a random element of an appropriate function space. In applications one may be compelled to compute expectations of the form $E[h(L)]$ for various deterministic functions $h$. In general the distribution of $L$ is extraordinarily difficult, if not impossible, to determine and one is forced to rely on simulation methods. For reasons discussed in Section 3.1, exact simulation is generally not possible. As such this chapter is dedicated to approximate simulation methods, their justification and an assessment of their accuracy.

In Section 3.4 we generalize our methods and results to the case of a time-changed Brownian motion. To this end let $g : [0, T] \mapsto [0, \infty)$ be a strictly increasing function with $g(0) = 0$. We define $\Psi(f, g)$ as the probability that a Brownian motion, subjected to the time-change $g$, breaches the barrier $f$ at any point over the interval $[0, T]$. Section 3.4 extends our methods and results to approximate simulation of $\Psi(A, B)$, where $B$ is a stochastic time change.

## 3.1 Motivation and Related Literature

Suppose that $W$ is a standard Brownian motion and that $f$ is a deterministic function. The first passage time is defined as

$$\tau_f = \inf\{t \geq 0 : \ W(t) \geq f(t)\}$$

In general closed form expressions for crossing probabilities $P(\tau_f \leq T)$ are only available in very specific cases, for example linear barriers or barriers of the form given by Daniels [37] and later generalized by Di Nardo et al. [39]. However the approximation of such probabilities is a very well-developed problem in the literature. These efforts were pioneered by Durbin [46] and Park and Schuurmann [95], who focused on characterizing the cumulative distribution function of $\tau_f$ as the solution to a Volterra equation of the first kind[1]. This approach has the distinct advantage of providing rapid approximations (see Section 3.3.2), however it does not appear possible to obtain bounds on the resulting approximation error. A more recent approach, pioneered by Pötzelberger and Wang, approximates $P(\tau_f \leq T)$ with $P(\tau_{f_n} \leq T)$, where $f_n$ is a piecewise linear approximation to $f$. Section 3.3.1 discusses two methods for computing such probabilities. The advantage of this approach is that it provides explicit and computable bounds on the approximation error. As such we will focus primarily on this approach, as these bounds are necessary in order to prove our convergence results in Section 3.2.

Suppose now that we are given a stochastic process $A$, independent of the standard Brownian motion $W$, and inquire about the first passage time

$$\tau_A = \inf\{t \geq 0 : \ W(t) \geq A(t)\} \tag{3.1}$$

There has been *much* less attention devoted to this problem in the literature. When $A$ is such that $X = W - A$ is a Markov process (with $A(0) > 0$), one may appeal to the integral equation derived by Peskir and Shiryaev [96]

$$P(X_t \geq 0) = \int_0^t P(X_t \geq 0 \,|X_s = 0)\, dG(s) \tag{3.2}$$

where $G$ is the cumulative distribution function of $\tau_A$. Provided the kernel here is non-singular, numerical approximation of (3.2) is straightforward. When $A$ is the sum of a deterministic function and a process with stationary and independent

[1]Durbin [46] investigates techniques for numerical solution of a singular equation which was originally derived by Fortet [56] Park and Schuurmann [95] derive a non-singular equation which admits simple numerical approximation.

increments, Beekman and Park [13] investigate the approximation of probabilities $P(\tau_A \leq T)$ using a similar approach based on integral equations. Unfortunately in both cases it does not appear possible to bound the resulting error.

For the general $A$ we see that only isolated analytic or semi-analytic results are available, and it is natural to inquire about simulation methods. Provided $\psi(A)$ is a version of the conditional crossing probability $P(\tau_A \leq T \,|\sigma(A_t : \, 0 \leq t \leq T))$, we note that $P(\tau_A \leq T) = E[\psi(A)]$. In an ideal world one would be able to carry out the following program for simulating realizations of $L = \psi(A)$, and hence approximating the desired first-passage probability $E[L]$

- Simulate a realization of $A$, say $a = \{a_t : \, 0 \leq t \leq T\}$
- Evaluate $\psi(a)$

Unfortunately there are two problems here for the general $A$. First, it may not be possible to simulate a full path, as would be the case when $A$ is the integral of a diffusion process. Hence in the first step we may have to content ourselves with simulating a linear approximation to $A$, say $A_n$. The second problem arises from the fact that even if we can simulate a full path of $A$, for example if $A$ were the integral of a shot-noise process[2], closed-form expressions for the functional $\psi(a)$ are unavailable for all but the simplest $a$, for example linear or of the form given in Daniels [37]. Fortunately semi-analytic expressions are available in the case that $a$ is piecewise linear and continuous. Hence in the case where we *can* simulate a full path of $A$, we might approximate $\psi(a)$ with $\psi(a_n)$, where $a_n$ is a piecewise linear approximation to $a$. But this simply amounts to simulating one observation of the random variable $\psi(A_n)$. Thus we see that in general, regardless of the form of $A$, both problems in the "idealized algorithm" can be solved by considering the following approximate (and perhaps more importantly, feasible) algorithm

- Simulate a full path of $A_n$, say $a_n = \{a_n(t) : \, 0 \leq t \leq T\}$. Here $A_n$ is simply a piecewise linear process which approximates $A$ in a certain sense.
- Evaluate $\psi(a_n)$

In using this algorithm we will be able to simulate observations of $L_n = \psi(A_n)$ exactly (up to a small numerical or Monte Carlo integration error) and approximate $E[L]$ with $E[L_n]$. Such an approximation will be justified provided $L_n$ converges

[2]This processes are discussed in more detail in Section 3.3.3

weakly to $L$, since $L$ is bounded. In Section 3.2 we prove almost sure convergence under very mild conditions on $A$, and in Section 3.3.3 we show how one may obtain an estimate of the approximation error $|E[L_n - L]|$ via simulation.

Further motivation for this problem is as follows. In Chapter 4 we propose a model for credit risk where the random variable $L = \Psi(A, B)$ represents the proportion of defaults in an asymptotically large portfolio. In applications of the model one might want to estimate crucial quantities such as the probability that portfolio losses exceed some critical upper level, or compute expectations such as $E\left[(L-K)^+\right]$ for constant $K$. These latter expectations are the necessary ingredients for pricing collateralized debt obligations on the portfolio. In both cases analytic results are difficult at best, impossible at worst. Thus weak convergence of $L_n = \Psi(A_n, B_n)$ to $L = \Psi(A, B)$ makes model implementation possible.

Before proceeding we recall the definition of a random element (see Billingsley [19]), which will be used in Sections 3.2 and 3.4. Suppose that $(S, d)$ is a metric space with Borel $\sigma$-algebra $\mathcal{S}$. Given a probability space $(\Omega, \mathcal{F}, Q)$, a mapping $X : \Omega \to S$ is called a random element of $S$ if it is measurable between $(\Omega, \mathcal{F})$ and $(S, \mathcal{S})$.

## 3.2 Main Results

Throughout this section we will confine ourselves to a fixed time horizon $[0, T]$, $W$ will denote a standard Brownian motion and the set $\mathcal{H}[0, T]$ will consist of all absolutely continuous functions $f$ over $[0, T]$, whose derivatives are square-integrable

$$\int_0^T [f'(t)]^2 \, dt < \infty$$

This space is often referred to as the Cameron-Martin space, and in what follows we endow it with the uniform metric

$$d(f, g) = \sup\{|f(t) - g(t)| : \ 0 \le t \le T\}$$

For $f \in \mathcal{H}[0, T]$ we define the first passage time of $W$ to the barrier $f$ as

$$\tau_f = \inf\{t \ge 0 : \ W(t) \ge f(t)\} \tag{3.3}$$

and note that $\tau_f = 0$ if $f(0) \le 0$. Our first result of this section states that first passage probabilities $P(\tau_f \le T)$ are continuous in the barrier $f$.

**Theorem 3.2.1.** *The functional* $\psi : \mathcal{H}[0,T] \mapsto [0,1]$ *defined by*

$$\psi(f) = P(\tau_f \leq T) \tag{3.4}$$

*is continuous.*

*Proof.* Our first step is to show that for $f \in \mathcal{H}[0,T]$ the random variable $M_f$ defined as

$$M_f = \max\{W(t) - f(t) : \ 0 \leq t \leq T\} \tag{3.5}$$

is a continuous random variable. Since $M_f = M_{f-f(0)} - f(0)$ we may assume without loss of generality that $f(0) = 0$. Let $B \subset \mathbb{R}$ be a Borel set with Lebesgue measure zero. Using Girsanov's Theorem we see that

$$P(M_f \in B) = E[\Lambda I(M_0 \in B)]$$

where

$$\Lambda = \exp\left(-\int_0^T f'(t)dW_t - \frac{1}{2}\int_0^T [f'(t)]^2 dt\right)$$

Since the maximum of a standard Brownian motion is continuous, it follows that $P(M_f \in B) = 0$, hence $M_f$ is continuous.

Now let $f \in \mathcal{H}[0,T]$ and suppose that $f_n$ is a sequence with $\epsilon_n = d(f, f_n) \to 0$. It is easy to see that

$$|M_f - M_{f_n}| \leq \epsilon_n \qquad \text{a.s.}$$

Using this fact and the continuity of $M_f$ and $M_{f_n}$ we obtain the inequality, valid for any real $m$

$$P(M_f \geq m + \epsilon_n) \leq P(M_{f_n} \geq m) \leq P(M_f \geq m - \epsilon_n) \tag{3.6}$$

Since $M_f$ is continuous and $\psi(f) = P(M_f \geq 0)$, it follows that $\psi(f_n) \to \psi(f)$, as required. □

Theorem 3.2.1 reveals that if $f_n$ is a sequence of approximating functions which converge uniformly to $f$, then $\psi(f_n) \to \psi(f)$. Moreover, setting $\epsilon_n = d(f, f_n)$, switching $f$ and $f_n$ in (3.6) and using the fact that

$$\psi(f_n + \epsilon_n) \leq \psi(f_n) \leq \psi(f_n - \epsilon_n)$$

one readily obtains the following bound on the approximation error in terms of the distance $\epsilon_n = d(f, f_n)$

$$|\psi(f) - \psi(f_n)| \leq \max(\psi(f_n - \epsilon_n) - \psi(f_n), \psi(f_n) - \psi(f_n + \epsilon_n)) \tag{3.7}$$

An approximation scheme only becomes operational (i.e. practical) when $\psi(f_n)$ and $\psi(f_n \pm \epsilon_n)$ can be computed, as such the only "sensible" schemes are those in which the approximating functions are piecewise linear. There are myriad ways of choosing a piecewise linear approximating sequence $f_n$, of which we now discuss two. We take as given an array of node times

$$\{t_{i,n} : \ n \geq 1, \ 0 \leq i \leq n, \ 0 = t_{0,n} < t_{1,n} < \ldots < t_{n,n} = T\} \tag{3.8}$$

with the property that

$$\Delta_n = \max\{t_{i,n} - t_{i-1,n} : \ 1 \leq i \leq n\} \to 0 \quad \text{as } n \to \infty \tag{3.9}$$

We call a "type one" approximation the sequence $f_n$, where $f_n$ is linear over the interval $[t_{i-1,n}, t_{i,n}]$ with $f_n(t_{i,n}) = f(t_{i,n})$. For such an approximation it is straightforward to show that

$$d(f, f_n) \leq 2\Delta_n \sup\{|f'(t)| : \ 0 \leq t \leq T\} \tag{3.10}$$

hence such an approximating sequence is justified provided $f$ has a bounded derivative.

For a stochastic process $A_t$, defined on a probability space $(\Omega, \mathcal{F}, P)$, we define the type one approximation $A_n$ as the process which is linear over $[t_{i-1,n}, t_{i,n}]$ with $A_n(t_{i,n}) = A(t_{i,n})$. Thus $A_n$ is a piecewise linear process whose law is determined by that of the random vector $(A(0), A(t_{1,n}), \ldots, A(T))$. Before proceeding with our next main result, which deals with weak convergence of $\psi(A_n)$ to $\psi(A)$, there are some technical issues which we feel are prudent to discuss. If every sample of $A$ lies in $\mathcal{H}[0,T]$, then the mapping $\omega \mapsto A(\cdot, \omega)$ is measurable[3] and we may view $A$ as a random element of $\mathcal{H}[0,T]$. Moreover for each $\omega \in \Omega$ there exists a function $A'(\cdot, \omega) \in \mathcal{L}^2[0,T]$ with the property that $A(t, \omega) = \int_0^t A'(s, \omega)\, ds$. There are some delicate issues involved with such a construction. To begin such an $A'$ is clearly not unique, so we call any fixed choice a *determination* of $A'$. In addition it is not clear that in general, quantities such as $\omega \mapsto \sup_{0 \leq t \leq T} |A'(\omega, t)|$ will be measurable. As such we include this as a condition in the following corollary to Theorem 3.2.1

**Corollary 3.2.2.** *Let $A$ be a random element of $\mathcal{H}[0,T]$, and suppose there exists a determination of $A'$ such that*

$$\omega \mapsto \sup_{0 \leq t \leq T} |A'(\omega, t)|$$

[3]This follows from the facts that $\mathcal{H}[0,T]$ is separable and that a continuous function is determined by its values at rational times.

*is measurable with*

$$P\left(\sup_{0\leq t\leq T}|A'(t)|<\infty\right)=1$$

*Then* $\psi(A_n)\to\psi(A)$ *almost surely, where* $A_n$ *is the type one approximation to* $A$*.*

*Proof.* By (3.10) we have

$$d(A,A_n)\leq 2\Delta_n\sup\{|A'(t)|:\ 0\leq t\leq T\}$$

and since the bounding random variable is almost surely finite, we have that $d(A,A_n)\to 0$ almost surely as $n\to\infty$. Since $\psi$ is continuous it follows that $\psi(A_n)\to\psi(A)$ almost surely. □

The conditions of Corollary 3.2.2 are in fact quite general. For example they are satisfied in the situation $A_t=A_0+\int_0^t M_s ds$, where $M_t$ is a process with càdlàg sample paths. A situation where the conditions do not apply is also easily imagined, for example the case where $A_t=Xt^{3/4}$, where $X$ is a random variable.

In many cases, for example when $A$ is the integral of a diffusion process, only simulation of $A'$ is possible. In such cases the type one approximation is not feasible, and it is therefore desirable to investigate linear approximation schemes which rely solely on knowledge of a function's derivative. Motivated by the trapezoidal rule we define the "type two" approximating sequence $f_n$ via

$$f_n(t)=f_n(t_{i-1,n})+\frac{f'(t_{i-1,n})+f'(t_{i,n})}{2}(t-t_{i-1,n})\qquad t\in[t_{i-1,n},t_{i,n}]\tag{3.11}$$

with the convention that $f_n(0)=f(0)$. It is straightforward to show that for such an approximation we have

$$d(f,f_n)\leq T\cdot w(f',\Delta_n)\tag{3.12}$$

where $w$ denotes the modulus of continuity

$$w(f,\delta)=\sup\{|f(t)-f(s)|:\ \ 0\leq s,t\leq T,\ \ |s-t|\leq\delta\}$$

If $f'$ is continuous, then $w(f,\delta)\to 0$ as $\delta\to 0$, therefore the type two approximation is appropriate provided $f$ has a continuous derivative. Note that in this case *both* the type one and two approximations are appropriate. For a stochastic process $A$ with square-integrable derivative $A'$ we define the type two linear approximation as that process for which $A_n(0)=A(0)$ and $A_n$ is linear over $[t_{i-1,n},t_{i,n}]$ with slope $\left(A'(t_{i-1,n})+A'(t_{i,n})\right)/2$. Thus $A_n$ is a piecewise linear (and continuous) process whose law is determined by that of the random vector $(A(0),A'(0),A'(t_{1,n}),A'(t_{2,n}),\ldots,A'(T))$. As a second corollary to Theorem 3.2.1 we have

**Corollary 3.2.3.** *Let $A$ be a random element of $\mathcal{H}[0,T]$, and suppose that there exists a determination of $A'$ such that the event*

$$\{\omega : \ A'(\cdot,\omega) \text{ is continuous}\}$$

*is measurable and almost sure. Then $\psi(A_n) \to \psi(A)$ almost surely, where $A_n$ is the type two approximation to $A$.*

*Proof.* By (3.12) we have

$$d(A, A_n) \leq Tw(A', \Delta_n)$$

Since $A'$ is almost surely continuous we have that $d(A, A_n) \to 0$ almost surely as $n \to \infty$. Continuity of $\psi$ yields the desired results. □

## 3.3 Computational Methods and Examples

In this section we investigate two methods for computing first passage probabilities for standard Brownian motion to piecewise linear barriers. Both methods are constructed by combining the Markov property of Brownian motion with known closed-form results for linear barriers. The first method is numerical integration, which is essentially exact, but can become computationally expensive quite rapidly. The second method, Monte Carlo simulation, is significantly faster but introduces statistical error. We compare both methods when used to approximate first passage probabilities to the square-root barrier $f(t) = \sqrt{1+t}$. We find that using node spacings as large as 0.25 can produce approximations which are exact to within several significant digits. For completeness we also compare the efficiency of the Monte Carlo method to the integral equation approach, finding the latter to be superior in terms of computational efficiency.

Section 3.3.3 discusses approximation of $E[h(\psi(A))]$ via simulation. In particular we show how one may assess, also via simulation, the error inherent in approximating such quantities with $E[h(\psi(A_n))]$. As an illustrative example we investigate the case where $A$ is an integrated shot-noise process and find the approximation error to be quite reasonable with node spacings as large as 0.25. Moreover one can get a remarkably accurate estimate of this error with a very small sample size.

### 3.3.1 Piecewise Linear Barriers

In this section we look at two methods for approximating crossing probabilities in the case where the (deterministic) barrier $f$ is piecewise linear. Throughout this section $W$ will denote a Brownian motion and $P^x$ will denote the measure under which $W$ begins at $x$ - that is $P^x(W_0 = x) = 1$. When $x = 0$ we will remove the superscript, that is $P = P^0$. Recall that the first passage time is defined as

$$\tau = \inf\{t \leq 0 : W_t \geq f(t)\} \tag{3.13}$$

We begin with the simplest case where $f(t) = b+mt$ is "purely" linear, for which the following results are well-known (see Karatzas and Shreve [80] for example)

- Under $P^x$ the distribution function of $\tau$ is defined as

  $$F(x, b, m, t) = P^x(\tau \leq t)$$

  and the domain may be taken to be $\mathbb{R}^4$. For $t < 0$ this function is clearly zero, while for $x \geq b$ it is degenerate in the sense that $F(x, b, m, t) = 1$ for $t \geq 0$. In the non-degenerate case where $x < b$ we have, for $t \geq 0$

  $$F(x, b, m, t) = \Phi\left(-\frac{(b-x)+mt}{\sqrt{t}}\right) + e^{-2(b-x)m}\Phi\left(\frac{mt-(b-x)}{\sqrt{t}}\right) \tag{3.14}$$

  where $\Phi$ is the cumulative distribution function of a standard normal random variable.

- The sub-density $\psi(w, x, b, m, t)$ of the absorbed process $W_{t\wedge\tau}$ will have domain $\mathbb{R}^4 \times [0, \infty)$ and is defined via

  $$P^x(W_t \in B,\ \tau > t) = \int_B \psi(w, x, b, m, t)\, dw$$

  If $x \geq b$, then $\tau$ is almost surely zero under $P^x$, hence $\psi$ is zero for $x \geq b$. In the case where $x < b$, $\psi$ may be defined for $t = 0$ as $\delta(w - x)$, where $\delta$ is the Dirac delta function. Finally in the case where $x < b$ and $t > 0$ we have that

  $$\psi(w, x, b, m, t) = \frac{1}{\sqrt{2\pi t}}\left[\phi\left(\frac{w-x}{\sqrt{t}}\right) - e^{-2(b-x)m}\phi\left(\frac{w+x-2b}{\sqrt{t}}\right)\right] \tag{3.15}$$

  where $\phi$ is the probability density function of a standard normal random variable.

- The conditional crossing probability is defined as

$$p(x,y,b,m,t) = P^x(\tau \leq t \,|W_t = y)$$

The function $p$ has domain $\mathbb{R}^4 \times (0,\infty)$. For $x \geq b$ we have that $p$ is equal to one, while for $x < b$ it is given by

$$p(x,y,b,m,t) = \begin{cases} \exp\left(-\frac{2(b-x)(b+mt-y)}{t}\right) & y < b+mt \\ 1 & y \geq b+mt \end{cases} \tag{3.16}$$

Another result which does not appear to be as well known concerns the cumulative distribution function of the first passage time of a Brownian bridge to a linear barrier (note that such a random variable may be defective in the sense that it takes on the value $\infty$ with positive probability), namely probabilities of the form

$$q(x,y,b,m,t,T) = P^x(\tau \leq t \,|W_T = y) \tag{3.17}$$

The domain of this function may be taken as the product of $\mathbb{R}^4$ with the set $\{(t,T) : T > 0,\ t \leq T\}$. For $t < 0$ we have that $q = 0$ for all values of the other parameters. When $x \geq b$ we have that $q = 0$ for $t < 0$ and $q = 1$ for $t \geq 0$. An explicit form for $q$ in non-degenerate cases is provided in the following

**Proposition 3.3.1.** *For $T > 0$, $0 < t < T$ and $x < b$ we have that $q(x,y,m,b,t,T)$ defined by (3.17) is equal to*

$$d_1(x,y,b,m,t,T) + \exp\left(-\frac{2(b-x)(b+mT-y)}{T}\right) d_2(x,y,b,m,t,T) \tag{3.18}$$

*where*

$$\begin{aligned} d_1(x,y,b,m,t,T) &= \Phi\left(-\frac{\frac{t}{T}(b+mT-y) + (b-x)\left(1-\frac{t}{T}\right)}{\sqrt{t\left(1-\frac{t}{T}\right)}}\right) \\ d_2(x,y,b,m,t,T) &= \Phi\left(\frac{\frac{t}{T}(b+mT-y) - (b-x)\left(1-\frac{t}{T}\right)}{\sqrt{t\left(1-\frac{t}{T}\right)}}\right) \end{aligned}$$

*Proof.* To begin we note that $P^x(\tau \leq t \,|W_t) = p(x,W_t,b,m,t)$. Now, since $A := \{\tau \leq t\} \in \mathcal{F}_t^W$ we have that $W_T - W_t$ is independent of $\sigma(W_t, I_A)$, where $I_A$ denotes the indicator of the event $A$. Therefore

$$\begin{aligned} P^x(\tau \leq t \,|W_T) &= E^x[I_A \,|W_T] \\ &= E^x[E^x[I_A \,|W_T - W_t, W_t] \,|W_T] \\ &= E^x[E^x[I_A \,|W_t] \,|W_T] \\ &= E^x[p(x,W_t,b,m,t) \,|W_T] \end{aligned}$$

Thus in order to evaluate the desired quantity we simply need to compute the integral

$$\int_{\mathbb{R}} p\left(x, w, b, m, t\right) P^{x}\left(W_{t} \in dw \,|W_{T}=y\right) \tag{3.19}$$

Recalling (3.16) we see that this integral may be split into two pieces, one of which is

$$\int_{b+mt}^{\infty} P^{x}\left(W_{t} \in dw \,|W_{T}=y\right)=P^{x}\left(W_{t} \geq b+mt \,|W_{T}=y\right)=d_{1}\left(x, y, b, m, t, T\right)$$

where we have used the fact that, conditional upon $W_T = y$, the distribution of $W_t$ is Gaussian with mean $x + \frac{t}{T}(y - x)$ and variance $t\left(1 - \frac{t}{T}\right)$. The second piece of (3.19) is given by

$$\int_{-\infty}^{b+mt} e^{-2(b-x)(b+mt-w)/t} P^{x}\left(W_{t} \in dw \,|W_{T}=y\right)$$

It is easy to verify that if $X$ is Gaussian with mean $\mu$ and variance $\sigma^2$, then

$$E\left[e^{aX} I\left(X \leq c\right)\right]=\exp\left(a\left(\mu+\frac{a\sigma^{2}}{2}\right)\right) \Phi\left(\frac{c-\left(\mu+a\sigma^{2}\right)}{\sigma}\right)$$

Using this result and the known conditional distribution of $W_t$, given $W_T = y$ we obtain that this integral is indeed given by

$$\exp\left(-\frac{2\left(b-x\right)\left(b+mT-y\right)}{T}\right) d_{2}\left(x, y, b, m, t, T\right)$$

as required. □

Our goal now is to show how these results, combined with the Markov property of Brownian motion, can be used to compute crossing probabilities to piecewise linear barriers. For the remainder of this section we fix $T > 0$ and suppose that $f$ is piecewise linear and continuous over $[0, T]$ with node times $0 = t_0 < t_1 < \ldots < t_{N-1} < t_N = T$. Denote the values of the function at each node by

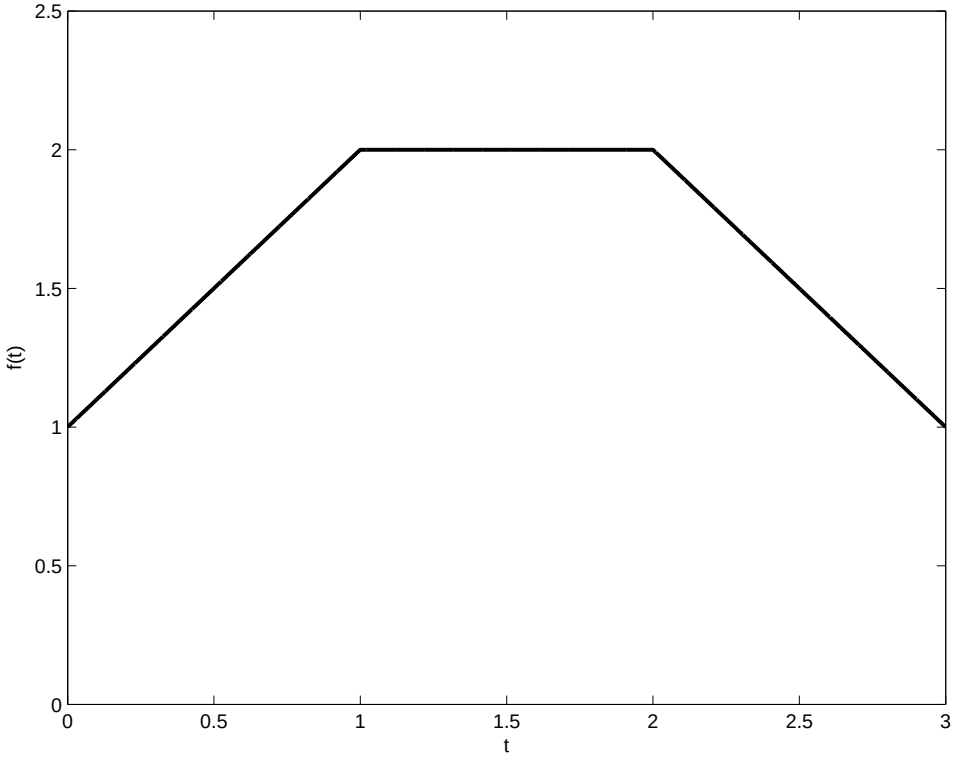
$$f_{i}=f(t_{i}) \qquad i=0,1, \ldots, N \tag{3.20}$$

and the slope of $f$ over the interval $[t_{i-1}, t_i]$ as

$$m_{i}=\frac{f_{i}-f_{i-1}}{t_{i}-t_{i-1}} \qquad i=1,2, \ldots, N \tag{3.21}$$

We will illustrate each method in the case $T = N = 3$, $f_0 = 1$, $(m_1, m_2, m_3) = (1, 0, -1)$ with equally-spaced nodes. This barrier is pictured in Figure 3.1 below.

Figure 3.1: Sample Barrier




**Numerical Integration**

For given $s < t$ define $A_{s,t}$ to be the event that $W$ remains below $f$ over $[s,t]$

$$A_{s,t} = \{W_u < f(u) : \ u \in [s,t]\} \tag{3.22}$$

By the Markov property of Brownian motion we have for $t_i < t \leq t_{i+1}$

$$\begin{aligned} P^x\left(A_{t_i,t} \,|W_{t_i}\right) &= 1 - F\left(W_{t_i}, f_i, m_{i+1}, t - t_i\right) \\ P^x\left(W_t \in dw, \ A_{t_i,t} \,|W_{t_i}\right) &= \psi\left(w, W_{t_i}, f_i, m_{i+1}, t - t_i\right) dw \end{aligned}$$

where $F$ and $\psi$ are given by (3.14) and (3.15), respectively. Now define, for each $i \geq 1$, the function

$$\psi^i(w) = P\left(W_{t_i} \in dw, \ \tau > t_i\right) = P\left(W_{t_i} \in dw, \ A_{0,t_i}\right)$$

and note that $\psi^1(w)$ is available in closed form

$$\psi^1(w) = \psi\left(w, 0, f_0, m_1, t_1\right)$$

with $\psi$ given by (3.15). Due to the Markov property of Brownian motion we obtain the following recursive relationship

$$\begin{aligned}
\psi^{i+1}(w) &= P\left(W_{t_{i+1}} \in dw,\ A_{0,t_{i+1}}\right) \\
&= P\left(W_{t_{i+1}} \in dw,\ A_{0,t_i},\ A_{t_i,t_{i+1}}\right) \\
&= E\left[P\left(W_{t_{i+1}} \in dw,\ A_{0,t_i},\ A_{t_i,t_{i+1}} \,|W_{t_i}\right)\right] \\
&= E\left[P\left(W_{t_{i+1}} \in dw,\ A_{t_i,t_{i+1}} \,|W_{t_i}\right) P\left(A_{0,t_i} \,|W_{t_i}\right)\right] \\
&= \int_{-\infty}^{\infty} \psi\left(w, x, f_i, m_{i+1}, t_{i+1} - t_i\right) P\left(A_{0,t_i} \,|W_{t_i} \in dx\right) P\left(W_{t_i} \in dx\right) \\
&= \int_{-\infty}^{f_i} \psi\left(w, x, f_i, m_{i+1}, t_{i+1} - t_i\right) \psi^i\left(x\right) dx
\end{aligned}$$

Starting with the known function $\psi^1$ we may recursively compute all the $\psi^i$ numerically, and we have found that the simple trapezoidal rule works quite well. Armed with the $\psi^i$ we may now compute first passage probabilities. To see this, using the Markov property of Brownian motion once again we obtain for $t_i < t \leq t_{i+1}$

$$\begin{aligned}
P\left(\tau \in (t_i, t]\right) &= P\left(A^C_{t_i,t},\ A_{0,t_i}\right) \\
&= E\left[P\left(A^C_{t_i,t} \,|W_{t_i}\right) P\left(A_{0,t_i} \,|W_{t_i}\right)\right] \\
&= \int_{-\infty}^{\infty} F\left(x, f_i, m_{i+1}, t - t_i\right) P\left(A_{0,t_i} \,|W_{t_i} \in dx\right) P\left(W_{t_i} \in dx\right) \\
&= \int_{-\infty}^{f_i} F\left(x, f_i, m_{i+1}, t - t_i\right) \psi^i(x) dx
\end{aligned}$$

And since $P\left(\tau \leq t_1\right) = F\left(0, f_0, m_1, t_1\right)$ is available in closed form we may use the following relationship for $i \geq 1$ and $t_i < t \leq t_{i+1}$

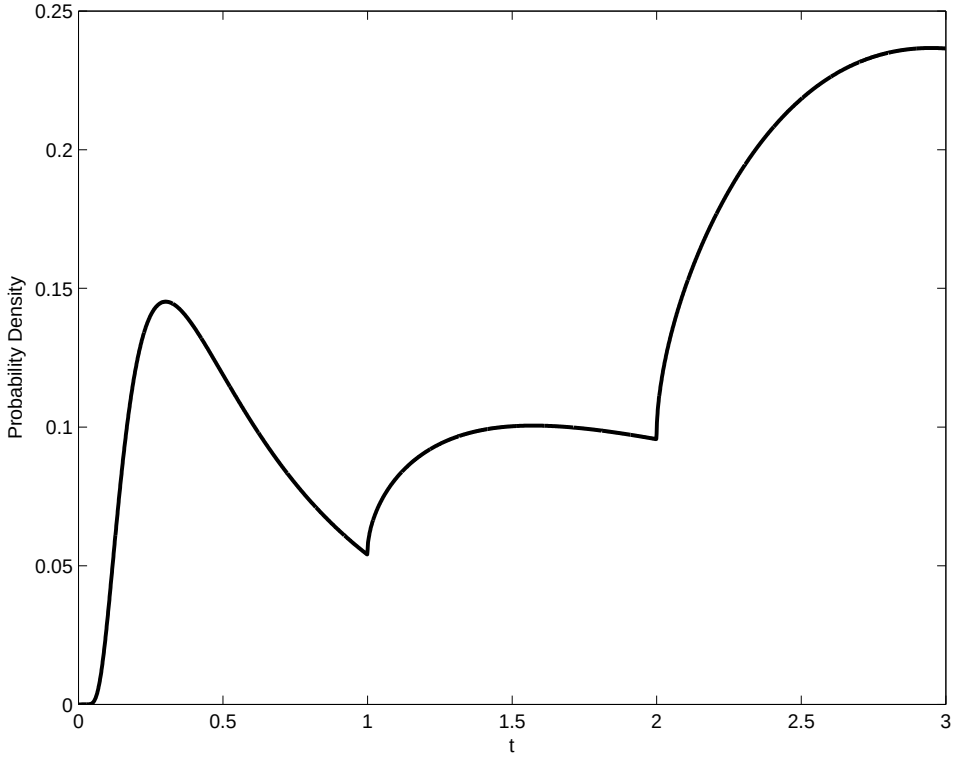
$$P\left(\tau \leq t\right) = P\left(\tau \leq t_i\right) + P\left(\tau \in (t_i, t]\right)$$

We implemented this method to compute the cumulative distribution function of $\tau$ for the barrier pictured in Figure 3.1, computing $F_j = P\left(\tau \leq s_j\right)$ for 3,000 equally-spaced points over the interval $[0, 3]$. We found the shape of the (approximated) density function to be quite interesting, and it is pictured in Figure 3.2. The density was approximated via $\left(F_{j+1} - F_j\right) / \left(s_{j+1} - s_j\right)$.

### Monte Carlo

The method described in the previous section can become quite slow, especially as the number of nodes increases. A much more rapid approach is to use simulation

Figure 3.2: Density Function - Piecewise Linear Example




methods. Without loss of generality assume we wish to approximate $P(\tau \leq T)$, and for $1 \leq i \leq N$ define the random variable

$$Q_i = P\left(A_{t_{i-1},t_i} \middle| W_{t_{i-1}}, W_{t_i}\right) = 1 - p\left(W_{t_{i-1}}, W_{t_i}, f_{i-1}, m_i, t_i - t_{i-1}\right) \tag{3.23}$$

By the Markov property of Brownian motion we have that

$$\begin{aligned} P(\tau \leq T) &= 1 - P(\tau > T) \\ &= 1 - P\left(\bigcap_{i=1}^{N} A_{t_{i-1},t_i}\right) \\ &= 1 - E\left[P\left(\bigcap_{i=1}^{N} A_{t_{i-1},t_i} \middle| W_{t_1}, \ldots, W_T\right)\right] \\ &= 1 - E\left[\prod_{i=1}^{N} P\left(A_{t_{i-1},t_i} \middle| W_{t_{i-1}}, W_{t_i}\right)\right] \\ &= 1 - E\left[\prod_{i=1}^{N} Q_i\right] \end{aligned}$$

An unbiased estimator for $P(\tau \leq T)$ is easily simulated by first simulating a "skeleton" path $W_{t_1}, \ldots, W_T$ and then computing the realized value of $\prod_{i=1}^{N} Q_i$. We note

that in order to determine the value of this product, it is not always necessary to simulate $W$ at each and every node time. Indeed if it is found that $W_{t_1} \geq f_1$, then $Q_1 = 0$ and hence $\prod_{i=1}^{N} Q_i = 0$. We see, therefore, that we may terminate the algorithm as soon as a simulated value is found to exceed the barrier. As a result, implementing the algorithm recursively can produce modest efficiency gains. Also note that as an immediate by-product of this simulation scheme one can compute for no extra cost the variables $1 - \prod_{j=1}^{i} Q_j$ for $1 \leq i \leq N$, thus obtaining unbiased estimators for $P(\tau \leq t_i)$ for each node time.

Consider now the case where $N = 2$ and suppose we are interested in computing $P_j = P(\tau \leq s_j)$ for $n$ values of $s_j \in [t_1, t_2]$. Without loss of generality assume that $s_1 = t_1$, $s_n = t_2$ and set $s_0 = 0$. In analogy with the method described in the previous paragraph we could set $Q_i = 1 - p\left(W_{s_{i-1}}, W_{s_i}, f_1, m_2, s_i - t_1\right)$ and use the estimator

$$P_j = 1 - \prod_{i=1}^{j} Q_i$$

which is unbiased for $P(\tau \leq s_j)$. Note that this requires simulating the Brownian path at $n$ time points, which may become expensive as $n$ gets large. As an alternative we may note that for $s \in (t_1, t_2)$ we have

$$P(A_{t_1,s} \,|W_{t_1}, W_{t_2}) = 1 - q(W_{t_1}, W_{t_2}, f_1, m_2, s - t_1, t_2 - t_1)$$

where $q$ is given by (3.18). Hence in order to construct an unbiased estimator we may set

$$\begin{aligned} Q_1 &= 1 - p(0, W_{t_1}, f_0, m_1, t_1) \\ Q_n &= 1 - p(W_{t_1}, W_{t_2}, f_1, m_2, t_2 - t_1) \\ Q_i &= 1 - q(W_{t_1}, W_{t_2}, f_1, m_2, s_i - t_1, t_2 - t_1) \qquad 2 \leq i \leq n-1 \end{aligned}$$

An unbiased estimator for $P(\tau \leq s_j)$ is then given by

$$P_j = 1 - \prod_{i=1}^{j} Q_i$$

and only requires simulating the Brownian motion at two time-points, as opposed to $n$. We have found that the main advantage of this method is that it tends to produce "smoother" estimates of the distribution function. This is illustrated in Figure 3.3, which plots the estimated distribution functions for the barrier in Figure 3.1, using both the latter "semi-analytic" method and the former "crude" method. We have found this to be the only major difference between the two methods, as neither method appears to be significantly more efficient or accurate.

Figure 3.3: MC Distribution Functions - Piecewise Linear Example. Each panel illustrates the estimated distribution function using a different Monte Carlo method.

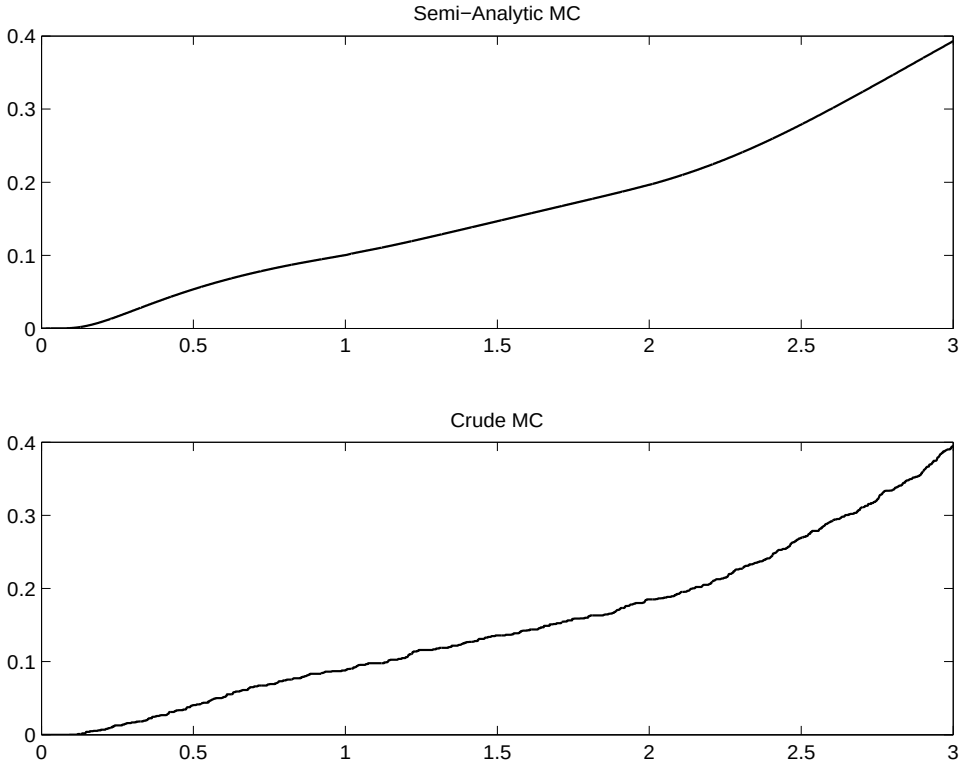


Table 3.1: Numerical Approximation

| $N$ | $\Delta$ | Time | $\psi(f_N)$ | Error Bound | $\psi(g_N)$ | Error Bound |
|---|---|---|---|---|---|---|
| 1 | 5 | .0382 | .4613 | .0616 | .4216 | .1361 |
| 2 | 2.5 | .2498 | .4388 | .0258 | .4201 | .0391 |
| 5 | 1 | .9251 | .4273 | .0067 | .4229 | .0069 |
| 10 | 0.5 | 3.1548 | .4250 | .0021 | .4238 | .0018 |
| 20 | 0.25 | 6.6806 | .4245 | .0006 | .4242 | .0005 |
| 40 | 0.125 | 13.3618 | .4243 | .0002 | .4242 | .0002 |

### 3.3.2 Square-Root Boundary

In this section we apply the methods of Section 3.3.1 to the problem of approximating $\psi(f) = P(\tau_f \leq T)$ for the boundary $f(t) = \sqrt{1+t}$ and $T = 5$. We use $N$ equally spaced nodes so that $t_i = i\Delta$ for $\Delta = T/N$. In order to bound the error using (3.7) we must be able to compute the distance between $f$ and its linear approximations. For the type-one approximation $f_N$ defined by $f_N(t_i) = f(t_i)$, concavity of $f$ allows us to determine that the maximum distance between $f$ and $f_N$ over the interval $[t_{i-1}, t_i]$ occurs at $t_i^*$ which satisfies

$$2\sqrt{1+t_i^*} = \frac{t_i - t_{i-1}}{\sqrt{1+t_i} - \sqrt{1+t_{i-1}}}$$

Since $f_N \leq f$ over the entire interval $[0, T]$, for the type one approximation we have that $\epsilon_N = d(f_N, f)$ is given by the maximum of $f(t_i^*) - f_N(t_i^*)$. For the type two approximation $g_N$ defined by $g_N(0) = 1$ and $g_N'(t) = (f'(t_{i-1}) + f'(t_i))/2$ for $t \in (t_{i-1}, t_i)$ we approximate $\epsilon_N = d(g_N, f)$ numerically.

Table 3.1 displays the results when using numerical integration. The first column lists the number of nodes used in the approximation, while the second reports the associated distance between nodes. The third column reports the computational time required, in seconds, with computations being conducted using Matlab 7 on a PC with a Pentium 4 processor. The next two columns display the value of the type one linear approximation along with its error bound

$$\max\left(\psi(f_N - \epsilon_N) - \psi(f_N), \psi(f_N) - \psi(f_N + \epsilon_N)\right) \tag{3.24}$$

The final two columns reported the analogous quantities for the type two approximation. We see that for both approximations the error becomes acceptable quite rapidly, being one-fifth of one percent for node spacings as large as 0.5. As a final

note we see that the computational time required is roughly linear in the number of nodes used, requiring approximately 0.3309 seconds per node.

We also compared a Monte Carlo estimator, using only the type one approximation. With $P = 1 - \prod_{i=1}^{N} Q_i$ and

$$Q_i = 1 - p\left(W_{t_{i-1}}, W_{t_i}, f_N(t_{i-1}), \frac{f(t_i) - f(t_{i-1})}{t_i - t_{i-1}}, t_i - t_{i-1}\right) \tag{3.25}$$

we simulate 10,000 skeleton paths for $W$, compute $P_j$ for each path and estimate the desired probability via $\bar{P}$, the average of the simulated $P_j$. Note that if we replace $f_N(t_{i-1})$ in (3.25) with $f_N(t_{i-1}) + \epsilon_N$, we may obtain an unbiased estimator, say $P^-$, of $\psi(f_N + \epsilon_N)$. Similarly, replacing $f_N(t_{i-1})$ with $f_N(t_{i-1}) - \epsilon_N$ we obtain an unbiased estimator, say $P^+$, of $\psi(f_N - \epsilon_N)$. Now, since (3.24) is equal to

$$\max\left(E\left[P^+ - P\right], E\left[P - P^-\right]\right) \leq E\left[\max\left(P^+ - P, P - P^-\right)\right] \tag{3.26}$$

we may obtain an estimate (3.24) by averaging the simulated values of

$$\max\left(P_j^+ - P_j, P_j - P_j^-\right)$$

Table 3.2 reports the results. The first column reports the number of nodes used, while the second and third columns compare the computational times required for Monte Carlo and numerical integration, in seconds. The fourth column reproduces the type one approximations from Table 3.1, while the fifth column reports $\bar{P}$, the average of the simulated $P_j$, with the standard error of $\bar{P}$ reported in parentheses. The sixth column reproduces the upper bound on the approximation error from Table 3.1, while the final column reports the Monte Carlo estimates of this quantity, with standard errors in parentheses. The most striking feature is that the Monte Carlo algorithm is *much* faster, indeed almost 45 times faster when 40 nodes are used. In addition Monte Carlo appears quite accurate, using 40 nodes the estimated probability lies within three significant digits of the numerical value. Indeed it appears that benefit of increased efficiency is well worth the cost of moderate statistical error. As a final note, it appears that we may estimate the approximation error (i.e. the error due to using a linear approximation) given by (3.24) with remarkable accuracy.

As noted in Section 3.1, piecewise linear approximation is only one of two techniques well-established techniques for approximating first passage probabilities. As an alternative we may consider the approach proposed by Park and Schuurmann [95], where the authors show that if $G$ denotes the cumulative distribution function

Table 3.2: Monte Carlo Approximation

| $N$ | Time | | Approximation | | Error Bound | |
|---|---|---|---|---|---|---|
| | Numerical | MC | Numerical | MC | Numerical | MC |
| 1 | .0382 | .0088 | .4613 | .4583 | .0616 | .0640 |
| | | | | (.0030) | | (.0002682) |
| 2 | .2498 | .0141 | .4388 | .4357 | .0258 | .0268 |
| | | | | (.0037) | | (.0001996) |
| 5 | .9251 | .0353 | .4273 | .4256 | .0067 | .0069 |
| | | | | (.0043) | | (.0000900) |
| 10 | 3.1548 | .0735 | .4250 | .4227 | .0021 | .0021 |
| | | | | (.0045) | | (.0000380) |
| 20 | 6.6806 | .1516 | .4245 | .4231 | .0006 | .0006 |
| | | | | (.0047) | | (.0000136) |
| 40 | 13.3618 | .3060 | .4243 | .4241 | .0002 | .0002 |
| | | | | (.0048) | | (.0000046) |

of the first passage time, then

$$\Phi\left(-\frac{f(t)}{\sqrt{t}}\right) = \int_0^t \Phi\left(-\frac{f(t)-f(s)}{\sqrt{t-s}}\right) dG(s) \tag{3.27}$$

Equation (3.27) is a Volterra equation of the first kind with kernel $K(s,t) = \Phi\left(-\frac{f(t)-f(s)}{\sqrt{t-s}}\right)$. When the barrier $f$ is differentiable this kernel is non-singular, and since $\lim_{s \nearrow t} K(s,t) = \frac{1}{2}$ we may define $K(t,t) = \frac{1}{2}$. A rapid approximation to the solution of (3.27) is then possible by noting

$$\Phi\left(-\frac{f(t_i)}{\sqrt{t_i}}\right) = \int_0^{t_i} K\left(s,t_i\right) dG(s) \approx \sum_{j=1}^{i} K\left(t_j,t_i\right)\left[G\left(t_j\right) - G\left(t_{j-1}\right)\right] \tag{3.28}$$

and this system of equations is easily solved in a recursive manner. Table 3.3 presents the results from implementing this procedure, and compares the computational time required (in seconds) with that of the Monte Carlo approach. We see that the integral equation approach provides an incredibly efficient alternative to Monte Carlo. When efficiency is of utmost importance, then, it appears that this may be the most preferable alternative. However it does not appear possible to obtain error bounds on this approximation, hence when an assessment of approximation error is of utmost importance one might prefer to use Monte Carlo.

Table 3.3: Integral Equation Approximation

| $N$ | $\Delta$ | Approximation | Time | MC Time |
|---|---|---|---|---|
| 1 | 5 | .2733 | .0003 | .0088 |
| 2 | 2.5 | .3409 | .0004 | .0141 |
| 5 | 1 | .3904 | .0021 | .0353 |
| 10 | 0.5 | .4073 | .0025 | .0735 |
| 20 | 0.25 | .4156 | .0042 | .1516 |
| 40 | 0.125 | .4199 | .0118 | .3060 |

### 3.3.3 Stochastic Barriers

Suppose that $A$ is a random element of $\mathcal{H}[0,T]$, and let $A_n$ denote a piecewise linear approximation to $A$. In light of the development in Section 3.3.1 we are now able to simulate $\psi(A_n)$ exactly, provided of course that we are able to simulate $A_n$ exactly. Moreover by Corollaries 3.2.2 and 3.2.3 we may approximate $E[L]$ with $E[L_n]$ under very mild conditions on $A$. Here $L = \psi(A)$ and $L_n = \psi(A_n)$. These observations suggest simulating variates $L_n$ and using their average $\bar{L}_n$ as a "point estimate" of $E[L]$. Note that the subscript $n$ refers to the number of nodes in the approximation, and *not* the number of simulations used. The estimation error in such a scheme may then be decomposed into two sources as follows

$$E[L] - \bar{L}_n = \underbrace{E[L] - E[L_n]}_{\text{Approximation Error}} + \underbrace{E[L_n] - \bar{L}_n}_{\text{Statistical Error}} \tag{3.29}$$

The second term here is "simulation noise," and may be "controlled" by either using a large number of simulated paths or an appropriate variance reduction technique. The first term is more dangerous, as it cannot be dealt with simply by increasing the number of simulations (though we do know it tends to zero as $n \to \infty$). In order to *confidently* approximate $E[L]$ with $E[L_n]$, it is necessary to assess the approximation error $|E[L] - E[L_n]|$. In this section we demonstrate how one may assess this error via simulation.

To begin let $A_n$ denote a linear approximation to $A$, and define the upper and lower approximations to $A$ as

$$A_n \pm \epsilon_n = \{A_n(t) \pm \epsilon_n : \ 0 \le t \le T\}$$

where the random variable $\epsilon_n$ is defined as $\epsilon_n = d(A, A_n)$. In addition we define

$$L_n^- = \psi(A_n + \epsilon_n) \qquad\qquad L_n^+ = \psi(A_n - \epsilon_n)$$

In light of (3.7) we obtain

$$|L - L_n| \leq \max\left(L_n^+ - L_n, L_n - L_n^-\right)$$

leading to the following upper bound on the approximation error

$$|E[L] - E[L_n]| \leq E[|L - L_n|] \leq E\left[\max\left(L_n^+ - L_n, L_n - L_n^-\right)\right] \tag{3.30}$$

The crucial fact here is that the random variable $U := \max(L_n^+ - L_n, L_n - L_n^-)$ can be simulated, thus we may approximate its mean via Monte Carlo, thereby obtaining an estimate of the approximation error. In the more general case where we are interested in approximating $E[h(L)]$ where $h$ is increasing and continuous we obtain the analogous bound on the approximation error

$$|E[h(L)] - E[h(L_n)]| \leq E\left[\max\left(h\left(L_n^+\right) - h(L_n), h(L_n) - h\left(L_n^-\right)\right)\right]$$

and the random variable $U_h := \max(h(L_n^+) - h(L_n), h(L_n) - h(L_n^-))$ may also be simulated, providing an estimate of the approximation error.

We apply this method when the barrier is given by the integral of the stationary shot-noise process, that is

$$A_t = b + \int_0^t M_s ds$$

where $b > 0$ is a constant and $M_t$ is given by

$$M_t = M_0 e^{-\alpha t} + \sum_{i=1}^{N_t} e^{-\alpha(t-T_i)} X_i$$

Here $\alpha > 0$ is a constant, $N_t$ is a homogeneous Poisson process with intensity $\lambda$ and event times $T_1, T_2, \ldots$ and the $X_i$ are i.i.d. random variables, independent of the Poisson process. We assume the jumps have the symmetric Laplace distribution with mean zero and scale parameter $\beta$, and that $M_0$ has characteristic function $\phi(u) = (1 + \beta^2 u^2)^{-\lambda/2\alpha}$. It can be shown that in this case $M_t$ is stationary. In addition, we note that $M_0$ is easily simulated as the difference between two independent gamma variates.

As the derivative of $A$ is bounded, but discontinuous, we must use the type one approximation. Our goal is to assess the approximation error for estimating both $E[L]$ and $E[h(L)]$, where $h(x) = (x - K)^+$ for a constant $K$. Relevant quantities are simulated according to the following recipe

- Simulate a realization of $A$, say $a$. Note that this only requires simulating a finite number of variables, namely the number of events, the event times and the jump sizes. Having simulated these variables the value $a(t)$ can be computed for any $t$.

Table 3.4: Approximation Error for the Shot-Noise Example

| $N$ | $\Delta$ | $\bar{L}_n$ | $\bar{U}$ | $\overline{h(L_n)}$ | $\bar{U}_h$ |
|---|---|---|---|---|---|
| 1 | 5 | .2997 | .0955 | .0352 | .0434 |
| | | (.0079) | (.0031) | (.0037) | (.0031) |
| 2 | 2.5 | .2982 | .0480 | .0394 | .0206 |
| | | (.0083) | (.0014) | (.0041) | (.0015) |
| 5 | 1 | .3099 | .0198 | .0445 | .0077 |
| | | (.0085) | (.0006) | (.0042) | (.0006) |
| 10 | 0.5 | .3028 | .0098 | .0368 | .0036 |
| | | (.0080) | (.0002) | (.0037) | (.0003) |
| 20 | 0.25 | .3119 | .0049 | .0459 | .0018 |
| | | (.0086) | (.0001) | (.0043) | (.0001) |

- Compute the associated realization of $\epsilon_n = d(A, A_n)$, say $\varepsilon_n = d(a, a_n)$. This is accomplished by computing the values $a(t)$ and $a_n(t)$ at $T \cdot 10^{-3}$ equally-spaced points, and set $\varepsilon_n$ equal to the maximum of the distances between these values.
- Compute $\ell_n = \psi(a_n)$, $\ell_n^+ = \psi(a_n - \varepsilon_n)$ and $\ell_n^- = \psi(a_n + \varepsilon_n)$. We compute these quantities using numerical integration.
- Set $u = \max(\ell_n^+ - \ell_n, \ell_n - \ell_n^-)$
- Set $u_h = \max(h(\ell_n^+) - h(\ell_n), h(\ell_n) - h(\ell_n^-))$

Table 3.4 presents the results of the experiment. The relevant parameters here are $b = 2.5$, $\lambda = 2.3$, $\alpha = 0.5$, $\beta = 0.2$, $T = 5$ and $K = 0.4$. These results are based on 500 simulated paths for $A$.

The first column indicates the number of nodes used in the approximation, while the second column indicates the associated node spacing. The third column reports the average $\bar{L}_n$, which may be seen as an estimate of $E[L]$. Standard errors are reported in parentheses. The fourth column reports the average $\bar{U}$, which may be seen as an estimate of the approximation error. The most striking feature is that the approximation error can be estimated quite efficiently, as indicated by the very small standard errors. The last two columns report the analogous quantities for estimating $E\left[(L - 0.4)^+\right]$. Here we see that the approximation error is quite reasonable, even for node spacings as large as 0.25. Moreover the estimate of this error is remarkably accurate.

## 3.4 Introducing a Time Change

In this section we generalize the results of Section 3.2 by introducing a stochastic time change to the Brownian motion. We continue to confine ourselves to the fixed time horizon $[0,T]$ and $W$ continues denote a standard Brownian motion. Let $\mathcal{C}[0,T]$ denote the set of continuous functions on this interval, and $\mathcal{C}^+[0,T]$ denote those $g \in \mathcal{C}[0,T]$ which are strictly increasing with $g(0)=0$. Both function spaces are endowed with the uniform metric $d$. Our first result concerns the random variable

$$M_{f,g} = \max\left\{W(g(t)) - f(t) :\ 0 \le t \le T\right\} \tag{3.31}$$

and is as follows

**Proposition 3.4.1.** *If $f_n \to f$ uniformly in $\mathcal{C}[0,T]$ and $g_n \to g$ uniformly in $\mathcal{C}^+[0,T]$, then $M_{f_n,g_n}$ converges weakly to $M_{f,g}$.*

*Proof.* Define the processes $Y_n = W(g_n) - f_n$ and $Y = W(g) - f$, noting that $Y$ and $Y_n$ are random elements of $(\mathcal{C}[0,T], d)$. Since $g_n(T) \to g(T)$ and all functions concerned are continuous, we have $K = \sup(g(T), g_1(T), \ldots) < \infty$. Now

$$d(Y, Y_n) \le d(f, f_n) + w(W, d(g, g_n))$$

where

$$w(W,\delta) = \sup\left\{|W_t - W_s| :\ 0 \le s,t \le K\ \ |s-t| \le \delta\right\}$$

Since $W$ is uniformly continuous on $[0,K]$ with probability one, we have that $w(W,\delta) \to 0$ almost surely as $\delta \to 0$. Therefore $d(Y,Y_n) \to 0$ almost surely as $n \to \infty$, hence $Y_n$ converges weakly to $Y$. Finally, since $\lambda : \mathcal{C}[0,T] \to [0,\infty)$ defined by $\lambda(f) = \max(|f(t)| :\ 0 \le t \le T)$ is continuous, it follows that $M_{f_n,g_n} = \lambda(Y_n)$ converges weakly to $M_{f,g} = \lambda(Y)$. □

Our ultimate goal here is to demonstrate that first-passage probabilities

$$P(\tau_{f,g} \le T)$$

are continuous in the barrier/time-change pair, where

$$\tau_{f,g} = \inf\left\{t \ge 0 :\ W(g(t)) \ge f(t)\right\} \tag{3.32}$$

In order to demonstrate this continuity we may restrict $(f,g)$ to an appropriate space, over which the random variables are $M_{f,g}$ are continuous. To this end we let $\mathcal{AC}[0,T]$ denote the set of absolutely continuous functions on $[0,T]$, and introduce

$\mathcal{AC}^+[0,T]$ as those $g \in \mathcal{AC}[0,T]$ with the property that $g(0) = 0$ and $g'$ is bounded away from zero, that is $\inf\{g'(t) : 0 \leq t \leq T\} > 0$ (note that this implies $g$ is strictly increasing). We are now ready for our main result of this section, and in what follows we endow the space $\mathcal{H}[0,T] \times \mathcal{AC}^+[0,T]$ with the metric

$$\rho\left((f_1,g_1),(f_2,g_2)\right) = \sqrt{d(f_1,f_2) + d(g_1,g_2)}$$

where $d$ is the uniform metric.

**Theorem 3.4.2.** *The functional* $\Psi : \mathcal{H}[0,T] \times \mathcal{AC}^+[0,T] \to [0,1]$ *defined by*

$$\Psi(f,g) = P(\tau_{f,g} \leq T) \tag{3.33}$$

*is continuous.*

*Proof.* Our first step is to show that if $(f,g) \in \mathcal{H}[0,T] \times \mathcal{AC}^+[0,T]$, then $f \circ g^{-1} \in \mathcal{H}[0,g(T)]$. Since $g' > 0$ almost everywhere, it follows that $g^{-1}$ is absolutely continuous (see [107]). Absolute continuity of $f$ and $g^{-1}$, combined with the fact that $g^{-1}$ is strictly increasing, imply that the composition $f \circ g^{-1}$ is also absolutely continuous. Moreover $\left(f \circ g^{-1}\right)'(s) = f'\left(g^{-1}(s)\right)/g'\left(g^{-1}(s)\right)$, and we see that

$$\int_0^{g(T)} \left[\frac{f'\left(g^{-1}(s)\right)}{g'\left(g^{-1}(s)\right)}\right]^2 ds = \int_0^T \frac{\left[f'(t)\right]^2}{g'(t)} dt < \infty$$

since $g'$ is bounded away from zero. Therefore $f \circ g^{-1} \in \mathcal{H}[0,g(T)]$. Now, since

$$M_{f,g} = \max\left\{W(s) - f\left(g^{-1}(s)\right) : 0 \leq s \leq g(T)\right\}$$

it follows that $M_{f,g}$ is a continuous random variable (recall the proof of Theorem 3.2.1).

Now if $(f_n,g_n) \to (f,g)$, Proposition 3.4.1 ensures that $M_{f_n,g_n}$ converges weakly to $M_{f,g}$, and since $M_{f,g}$ is continuous we have

$$\Psi(f_n,g_n) = P(M_{f_n,g_n} \geq 0) \to P(M_{f,g} \geq 0) = \Psi(f,g)$$

as required. □

The introduction of a time change poses no new computational challenges. To see this suppose we are interested in approximating $\Psi(f,g)$ and let $(f_n,g_n)$ be a sequence of appropriate (either type one or two) piecewise linear approximating functions. Assuming that $f_n, g_n$ have the same node times $0 = t_0 < t_1 < \ldots < t_n = T$ we may note that the composition $h_n = f_n \circ g_n^{-1}$ is also piecewise linear

(over $[0, g_n(T)]$) with node times $0 = g_n(0) < g_n(t_1) < \ldots < g_n(T)$, and the slope over each interval is given by $(f_n(t_i) - f_n(t_{i-1}))/(g_n(t_i) - g_n(t_{i-1}))$. Since $\Psi(f_n, g_n) = \psi(h_n)$, where the domain of $\psi$ here is $\mathcal{H}[0, g_n(T)]$, we may use the methods described in Section 3.3.1 to compute this quantity. Also note that when $g_n(T) = g(T)$ we may use (3.7) to bound the approximation error as

$$|\Psi(f, g) - \Psi(f_n, g_n)| \leq \max(\psi(h_n - \epsilon_n) - \psi(h_n), \psi(h_n) - \psi(h_n + \epsilon_n))$$

where

$$\epsilon_n = d\left(f \circ g^{-1}, f_n \circ g_n^{-1}\right) = \sup_{0 \leq s \leq g(T)} \left|f\left(g^{-1}(s)\right) - f_n\left(g_n^{-1}(s)\right)\right|$$

In addition when one does not require an estimate of the approximation error, we may appeal to the following Volterra equation of the first kind (see Peskir and Shiryaev [96])

$$\Phi\left(-\frac{f(t)}{\sqrt{g(t)}}\right) = \int_0^t \Phi\left(-\frac{f(t) - f(s)}{\sqrt{g(t) - g(s)}}\right) dH(s)$$

where $H$ denotes the cumulative distribution function of $\tau_{f,g}$. When $f$ and $g$ are differentiable the kernel here is non-singular, as such an approximation of $H(T)$ may obtained using a system of equations analogous to (3.28).

We are now prepared to derive the following analogue of Corollaries 3.2.2 and 3.2.3

**Corollary 3.4.3.** *Let $A$ be a random element of $\mathcal{H}[0, T]$ and let $B$, defined on the same probability space, be a random element of $\mathcal{AC}^+[0, T]$. Then*

*(i) If there exist determinations of $A'$ and $B'$ which are almost surely finite, then $\Psi(A_n, B_n) \to \Psi(A, B)$ almost surely, where $A_n$ and $B_n$ are the type one approximations to $A$ and $B$, respectively.*

*(ii) If there exist determinations of $A'$ and $B'$ which are almost surely continuous, then $\Psi(A_n, B_n) \to \Psi(A, B)$ almost surely, where $A_n$ and $B_n$ are the type two approximations to $A$ and $B$, respectively.*

*Proof.* In each case the distance between $(A, B)$ and $(A_n, B_n)$ converges to zero almost surely. Continuity of $\Psi$ then ensures the desired result. □

We conclude this section with a similar experiment to that performed in Section 3.3.3, only replacing $\psi(A)$ with $\Psi(A, B)$. To this end we let $A$ be as defined there

Table 3.5: Approximation Error for the Time-Changed Shot-Noise Example

| $N$ | $\Delta$ | $\bar{L}_n$ | $\bar{U}$ | $\overline{h(L_n)}$ | $\bar{U}_h$ |
|---|---|---|---|---|---|
| 1 | 5 | .2117 | .0975 | .0318 | .0394 |
| | | (.0093) | (.0039) | (.0041) | (.0036) |
| 2 | 2.5 | .2069 | .0517 | .0306 | .0205 |
| | | (.0093) | (.0022) | (.0040) | (.0022) |
| 5 | 1 | .2033 | .0199 | .0296 | .0070 |
| | | (.0092) | (.0008) | (.0041) | (.0008) |
| 10 | 0.5 | .2212 | .0112 | .0318 | .0042 |
| | | (.0091) | (.0004) | (.0037) | (.0004) |
| 20 | 0.25 | .2110 | .0053 | .0280 | .0019 |
| | | (.0090) | (.0002) | (.0037) | (.0002) |

and let $B_t = \int_0^t V_s ds$, where $V_t$ is a shot-noise process with exponential jumps, and we assume the two processes are independent. Table 3.4 presents the results of this experiment, namely assessing the approximation error $E[L]$ and $E[h(L)]$, where $L = \Psi(A, B)$. The results are quite similar, namely the approximation error appears quite reasonable, and can be estimated very accurately.

# Chapter 4

# Alternatives to the Black-Cox Model

In this chapter we propose a multivariate first-passage framework for credit risk. In order to motivate the framework let us briefly re-examine the multivariate Black-Cox model discussed in Section 1.1.4. Despite its significant intuitive appeal the model is simply not capable of describing market data for multiname credit derivatives. Recall that in this model a firm defaults upon first passage of its "credit quality" process, denoted $X^i$, to zero. Credit quality is defined as the log-ratio of the firm's asset value to its default threshold and is given by

$$X_t^i = x_0^i + \mu_i t + \sigma_i W_t^i \tag{4.1}$$

Here $\mu_i, \sigma_i, x_0^i$ are constant parameters while the $W^i$ are correlated Brownian motion. The parameters $\mu_i$ and $\sigma_i$ represent *trend* and *volatility* in credit qualities, respectively, while systematic risk appears under the guise of correlated "noise" about this trend.

Our contention is that the "location" of systematic risk here is the model's fatal flaw. It is well-documented that quantities such as returns, volatilities and correlations between financial assets are significantly influenced by the general state of the economy. In recognition of this fact, our model in effect removes systematic risk from the driving Brownian motion in (4.1), and places it in trend and volatility. We allow for both stochastic trend and volatility in credit qualities, with dependence introduced by letting these quantities to be driven by systematic factors common to all obligors.

In order to test the model's abilities we calibrate several versions to market data for CDX index tranches. We use two sets of quotes, the first consisting of pre-crisis

quotes taken from November 2006 and the second consisting of more recent quotes taken from March 2008. For the 2006 data, which includes quotes for super-senior tranches, we obtain very encouraging results when the model is calibrated across both tranches and maturities simultaneously. In addition the calibrated parameters predict CDS spreads, which are not used in the calibration, that match up almost perfectly with market quotes. For the 2008 data, which does not include super-senior or CDS quotes, we also obtain very encouraging results. As an example the model discussed in Section 4.2.3 contains ten parameters, and produces an average relative pricing error of only 3% when fitted to fifteen total spreads (each of five tranches at three different maturities). We are particularly pleased with these results, as the industry-standard model is not capable of describing this "distressed" data. For the 15-30% tranches, there are simply no parameter values in the Gaussian copula model which are capable of producing spreads as large as those observed in early 2008. Our model is able to simultaneously price these tranches at five, seven and ten-year maturities with a maximum error of only 0.8 basis points. This is accomplished while maintaining a very reasonable fit to both equity and mezzanine tranches. Our hope is that these pleasant results encourage market participants and regulators to seriously consider the framework proposed here.

## 4.1 General Framework

As in the widely popular factor models, we model both systematic and idiosyncratic risk explicitly in our framework. Systematic risk is modeled via the pair of processes $(M, V)$, which are not assumed to be independent of one another. We often refer to this pair as the "systematic factors." In order to facilitate discussion we loosely interpret these factors as being representative of the economic environment in which obligors operate, however we do not make an explicit link between the factors and any specific macroeconomic covariates. Idiosyncratic risk is introduced via a sequence of independent Brownian motions $W^1, W^2, \ldots$, and we assume that $(M, V)$ is independent of this sequence. Together these elements combine to drive the "credit quality" of obligor $i$ as follows

$$dX_t^i = \mu_i(M_t)\,dt + \sigma_i(V_t)\,dW_t^i \qquad X_0^i = x_0^i \tag{4.2}$$

Here $x_0^i > 0$ is a constant, while $\mu_i, \sigma_i$ are deterministic functions. In order that $X^i$ be well-defined we require $\mu_i(M_t)$ and $\sigma_i^2(V_t)$ to have integrable sample paths. To

this end we assume that $\mu_i$ and $\sigma_i$ are continuous functions, and that all sample paths of the systematic factors are càdlàg. Note also that while we do not allow the initial value of credit quality to be stochastic, this added generality could be handled quite easily. We decree that obligor $i$ defaults upon first passage of $X^i$ to zero, that is

$$\tau_i = \inf\left\{t \geq 0 : \ X_t^i \leq 0\right\} \tag{4.3}$$

A heuristic interpretation of this framework is as follows. Conditional upon the realized values of $M_t, V_t$, the incremental change $X_{t+h}^i - X_t^i$ is approximately Gaussian with mean $h\mu_i(M_t)$ and variance $h\sigma_i^2(V_t)$. Moreover incremental changes of distinct obligors are approximately independent. Thus we may think of the systematic factors as "setting the tone" for a day's operations, and once this tone has been set the fortunes of individual obligors are independent. By allowing the systematic factors to be time-varying and stochastic, we allow the economic environment in which obligors operate to evolve dynamically through time. This is not allowed in the Black-Cox model. One may also view our framework as providing a very natural extension of one-period factor models, incorporating the time dynamics which are sorely missed in those models. Moreover this framework circumvents many of the difficulties involved in extending these models to a discrete-time setting, as discussed by Andersen [5].

It is worth noting that several existing models fit within this general framework. For example if trend and volatility were functions of a finite-state Markov chain we would recover the model used by Kuen et al. [84] to price credit default swaps in a single-name setting. If drift were constant and volatility were expressed as a function of one or more Ornstein-Uhlenbeck processes, we would recover the models used by Fouque et al. to study the effects of stochastic volatility on bond yields [57] and portfolio loss distributions [58].

It is worthwhile to note that, in general, credit qualities are not Markov processes in this framework. For example if volatility were constant (and non-stochastic) and $\mu_i(M_t)$ were an Ornstein-Uhlenbeck process, it is easily verified that $X^i$ is not Markov. In this case $X^i$ is a Gaussian process whose covariance function is available in closed form. Moreover Feller [53] provides a necessary and sufficient condition, based on the covariance function, for a Gaussian process to be Markov. It is straightforward to verify that the condition is not satisfied for this example.

Additional insights into this framework may be gained by noting that credit qualities are semi-martingales with finite variation part given by

$$A_t^i = x_0^i + \int_0^t \mu_i\left(M_s\right) ds \tag{4.4}$$

and quadratic variation

$$B_t^i = \int_0^t \sigma_i^2 \left(V_s\right) ds \tag{4.5}$$

In modeling the derivatives of these quantities we implicitly restrict ourselves to the class of *continuous* semi-martingales. Thus, despite the fact that the systematic factors may be allowed to possess jumps, credit qualities will always be continuous. One might suspect that in a purely continuous framework such as this, severe "market crashes" are not possible. However we will see in Section 4.2 that this is not true, as scenarios where a large percentage of a portfolio (say 50%) defaults in a short time horizon (say six months) are entirely possible.

The predictability of default times is often of theoretical interest in dynamic models of credit risk. This property depends on the information available to investors. If investors here are endowed with the filtration generated by $X^i$, then $\tau_i$ is predictable since this process is continuous. As such default does not come as a surprise, as immediately prior to the event investors are aware that default is imminent. In Section 4.2.2 we will see that if investors are also privy to the realized paths of the systematic factors, this may lead to situations where the default event is in a sense "super-predictable," as certain realizations of these factors may allow investors to predict the time of default with near-certainty as much as five years in advance. As such we decree that the systematic factors are unobservable, in the sense that they are not both adapted to the filtration available to investors.

Let us now turn to the problem of computing marginal default probabilities $P\left(\tau_i \leq t\right)$. To this end we note that the law of $X^i$ may be expressed in terms of time-changed Brownian motion with stochastic drift

$$X_t^i \stackrel{\mathcal{L}}{=} A_t^i + W^i\left(B_t^i\right) \stackrel{\mathcal{L}}{=} A_t^i - W^i\left(B_t^i\right) \tag{4.6}$$

where $\stackrel{\mathcal{L}}{=}$ denotes equality in law. Hence the distribution of $\tau_i$ is identical to that of $\tilde{\tau}_i$, the first passage time of the process $W^i\left(B^i\right)$ to the stochastic barrier $A^i$. Thus

$$P\left(\tau_i \leq t\right) = E\left[P\left(\tilde{\tau}_i \leq t \,\middle|\, A^i, B^i\right)\right] = E\left[\Psi_t\left(A^i, B^i\right)\right]$$

where $\Psi_t$ is given by (3.33).[1] Approximation of such quantities via simulation was the subject of Chapter 3, and we see that the methods discussed there may be brought to bear on the present problem. Moreover these simulation methods are

[1]Note the slight change of notation here. For deterministic functions $f$ and $g$ defined on $[0,T]$ and $\tau_{f,g}$ defined by (3.32), we define $\Psi_t\left(f,g\right) = P\left(\tau_{f,g} \leq t\right)$ for each $t \in [0,T]$.

valid under very minimal conditions on the systematic factors, the only requirements are either continuous or bounded sample paths and the ability to simulate linear versions of $A^i$ and $B^i$.

Seen in the light of (4.6), our framework bears some resemblance to the multivariate first-passage models presented by Luciano and Schoutens [89], as well as Hurd [75]. These models may be described as pure time-change models, where $A^i$ is proportional to $B^i$, leading to credit qualities which are time changed Brownian motion. In [89] the time-change is a Gamma subordinator, leading to Lévy credit qualities, while in [75] $B^i$ could take more general forms such as the integral of a CIR process.

The most important feature of this framework from a computational point of view is that of *conditional independence*. Conditional upon the realized paths of the systematic factors, credit qualities are independent diffusion processes, in fact they are conditionally Gauss-Markov. The next section exploits this structure to represent the proportion of defaults on an asymptotically large portfolio as a path functional of the systematic factors. This will have significant consequences in terms of dimension reduction when implementing the model in a large-portfolio setting, as well as provide useful insights into the nature and behaviour of portfolio losses in the model.

### 4.1.1 Large Portfolio Approximation

In this section we investigate the large $N$ asymptotics of the proportion of defaults

$$D_t^N = \frac{1}{N} \sum_{i=1}^{N} I\left(\tau_i \leq t\right) \tag{4.7}$$

Our main result, whose proof is provided following a brief discussion, is the following

**Proposition 4.1.1.** *Let $\mathcal{H}_t = \sigma\left(M_s, V_s : \ 0 \leq s \leq t\right)$ denote the filtration generated by the systematic factors. Then for each $t$ we have*

$$\lim_{N \to \infty} \left[ D_t^N - E\left[ D_t^N \left| \mathcal{H}_t \right. \right] \right] = 0$$

*almost surely.*

Intuitively we see from this result that one can predict the proportion of defaults in a large portfolio based solely on the information provided by the systematic factors. More formally, when it exists we call

$$D_t = \lim_{N \to \infty} D_t^N \tag{4.8}$$

the *asymptotic proportion of defaults*, and note that a necessary and sufficient condition for $D_t$ to be well-defined is that the conditional expectations $E\left[D_t^N \,|\mathcal{H}_t\right]$ have an almost sure limit. In this case we clearly have

$$D_t = \lim_{N\to\infty} E\left[D_t^N \,|\mathcal{H}_t\right]$$

so that $D_t$ is necessarily $\mathcal{H}_t$-measurable. Thus the sequence of idiosyncratic, or firm-specific, risk processes $W^1, W^2, \ldots$ are in fact irrelevant with respect to the realized value of $D_t$. This lends formal justification to heuristic claims such as "in a large portfolio all risk is systematic" or "in a large portfolio idiosyncratic risk can be diversified away."

Let us now suppose that credit qualities are homogeneous in the sense that $\mu_i = \mu$, $\sigma_i = \sigma$ and $x_0^i = x_0$, so that

$$dX_t^i = \mu\left(M_t\right) dt + \sigma\left(V_t\right) dW_t^i \qquad X_0^i = x_0$$

In this situation the laws of the $X^i$ are identical, and we clearly have that conditional default probabilities are equal across obligors

$$P\left(\tau_i \le t \,|\mathcal{H}_t\right) = P\left(\tau_j \le t \,|\mathcal{H}_t\right) \qquad \forall\, i,j$$

It is easy to explicitly identify the asymptotic proportion of defaults here, since

$$\begin{aligned} D_t &= \lim_{N\to\infty} E\left[D_t^N \,|\mathcal{H}_t\right] \\ &= \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} P\left(\tau_i \le t \,|\mathcal{H}_t\right) \\ &= P\left(\tau_1 \le t \,|\mathcal{H}_t\right) \end{aligned}$$

Hence in the homogeneous case, the asymptotic proportion of defaults is given by the conditional default probability of an arbitrary obligor. In light of our discussion in the previous section we see that $D_t = \Psi_t\left(A, B\right)$, providing an explicit representation of the proportion of defaults in terms of the systematic factors.

As another example consider a portfolio consisting of a finite number of groups, with credit qualities being homogeneous within groups. We may think of obligors within each group as belonging to the same industry, or possessing the same credit rating. Indexing the groups by $k = 1, \ldots, K$, within-group homogeneity is tantamount to the assumption that if obligor $i$ belongs to group $k$, then

$$dX_t^i = \mu_k\left(M_t\right) + \sigma_k\left(V_t\right) dW_t^i \qquad X_0^i = x_0^k$$

Thus if obligors $i$ and $j$ both belong to group $k$, then their conditional default probabilities are identical

$$P\left(\tau_i \leq t \,|\mathcal{H}_t\right) = P\left(\tau_j \leq t \,|\mathcal{H}_t\right) = \Psi_t\left(A^k, B^k\right)$$

where $A_t^k = \int_0^t \mu_k\left(M_s\right) ds$ and $B_t^k = \int_0^t \sigma_k^2\left(V_s\right) ds$. Thus $\Psi_t\left(A^k, B^k\right)$ represents the conditional default probability of an arbitrary obligor belonging to group $k$. Now let $w_{k,N}$ denote the proportion of obligors 1 through $N$ which belong to group $k$. Provided that the proportion of obligors in each group is asymptotically stable, that is $w_k := \lim_{N\to\infty} w_{k,N}$ is well-defined for each $k$, we have

$$\begin{aligned} D_t &= \lim_{N\to\infty} E\left[D_t^N \,|\mathcal{H}_t\right] \\ &= \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} P\left(\tau_i \leq t \,|\mathcal{H}_t\right) \\ &= \lim_{N\to\infty} \sum_{k=1}^{K} w_{k,N} \Psi_t\left(A^k, B^k\right) \\ &= \sum_{k=1}^{K} w_k \Psi_t\left(A^k, B^k\right) \end{aligned}$$

and we see that in such a portfolio, the asymptotic proportion of defaults is simply a weighted average of conditional default probabilities.

### Proof of Proposition 4.1.1

Turning to the proof of Proposition 4.1.1 we begin by noting that $D_t^N$ can be described as the proportion of successes in a sequence of conditionally independent Bernoulli trials. To make this notion precise let us suppose that $Z_1, Z_2, \ldots$ are a sequence of Bernoulli variables[2] defined on some probability space $(\Omega, \mathcal{F}, P)$. Furthermore suppose these variables are conditionally independent, given some sub-$\sigma$-algebra $\mathcal{G} \subset \mathcal{F}$. In fact, Theorem 4.1.2 below remains true under the apparently weaker assumption that the $Z_i$ are conditionally uncorrelated, that is

$$E\left[Z_i Z_j \,|\mathcal{G}\right] = E\left[Z_i \,|\mathcal{G}\right] E\left[Z_j \,|\mathcal{G}\right] \qquad \forall i \neq j \tag{4.9}$$

We note that a conditionally independent sequence automatically satisfies (4.9). The proportion of "successes" in $n$ trials is then given by

$$\frac{1}{n} \sum_{i=1}^{n} Z_i \tag{4.10}$$

[2]By a Bernoulli variable we mean a random variable taking on only the values zero and one.

Conditional independence alone is not sufficient for (4.10) to converge almost surely. To see this take any two deterministic sequences of zeros and ones, say $\{x_i\}_{i=1}^{\infty}$ and $\{y_i\}_{i=1}^{\infty}$, for which $\frac{1}{n}\sum_{i=1}^{n} x_i$ and $\frac{1}{n}\sum_{i=1}^{n} y_i$ fail to converge.[3] Letting $U$ be a Bernoulli variable, and setting $Z_i = x_i$ if $U = 1$ and $Z_i = y_i$ otherwise, we obtain a sequence of conditionally independent variables with the property that with probability one, (4.10) fails to converge. Theorem 4.1.2 shows that a necessary and sufficient condition for (4.10) to converge is that $\frac{1}{n}\sum_{i=1}^{n} E\left[Z_i \,|\mathcal{G}\right]$ converges. In addition this result shows that when it exists, the limiting proportion of successes is necessarily $\mathcal{G}$-measurable.

**Theorem 4.1.2.** *Suppose that $Z_1, Z_2, \ldots$ are conditionally independent Bernoulli variables. Then with probability one we have*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \left(Z_i - E\left[Z_i|\mathcal{G}\right]\right) = 0 \tag{4.11}$$

*Proof.* To begin we denote $\tilde{Z}_i = E\left[Z_i \,|\mathcal{G}\right]$. A simple application of the tower property, combined with $\mathcal{G}$-measurability of $\tilde{Z}_j$ and (4.9) yields, for any $i \neq j$

$$E\left[Z_i\tilde{Z}_j\right] = E\left[\tilde{Z}_i\tilde{Z}_j\right] = E\left[Z_iZ_j\right]$$

Next we define

$$Y_i = \frac{Z_i - \tilde{Z}_i}{i}$$

and note that the $Y_i$ form an orthogonal sequence, that is $E\left[Y_iY_j\right] = 0$ whenever $i \neq j$, with second moments bounded by

$$E\left[Y_i^2\right] \leq \frac{1}{i^2}$$

Using the integral and comparison tests for series it follows that

$$\sum_{i=1}^{\infty} \log^2(i) E\left[Y_i^2\right] < \infty$$

and we may now invoke the Rademacher-Menchoff Fundamental Convergence Theorem for orthogonal random variables (see [108], Theorem 2.3.2), which guarantees that the partial sums $S_n = \sum_{i=1}^{n} Y_i$ converge almost surely to some random variable $S$. Applying Kronecker's Lemma we obtain the desired result. □

[3] A concrete example is furnished by considering sequences such as

$$0, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{2}{5}, \frac{3}{6}, \frac{3}{7}, \ldots, \frac{3}{12}, \frac{4}{13}, \ldots, \frac{9}{18}, \ldots$$

which oscillates between $\frac{1}{2}$ and $\frac{1}{4}$

Proposition 4.1.1 now follows easily by setting $Z_i = I(\tau_i \leq t)$ and $\mathcal{G} = \mathcal{H}_t$. We wish to stress the fact that Theorem 4.1.2 may be applied to *any* model in which default times are conditionally independent. This includes the widely popular factor models, as well as many intensity-based models such as those based on Cox processes.

### 4.1.2 Simulation of Portfolio Losses

Consider a portfolio consisting of $N$ obligors with nominal exposures $E_i$ and recovery rates $R_i$, which we assume are deterministic. The crucial object in applications is the percentage loss on the portfolio

$$L_t^N = \sum_{i=1}^{N} w_i (1 - R_i) I(\tau_i \leq t) \tag{4.12}$$

where $w_i = E_i / \sum_{i=1}^{N} E_i$. In this section we discuss the problem of simulating trajectories for portfolio losses over a fixed time interval $[0, T]$. In particular we are interested in simulating the process $L^N$ at discrete times $0 < t_1 < t_2 < \ldots < t_n = T$. A critical observation here is that we need not simulate default times exactly, we need only simulate the interval of default for each name

$$I_i = \min_{1 \leq j \leq n} \{j : \tau_i \leq t_j\} \tag{4.13}$$

with the convention that $\min\{\emptyset\} = +\infty$.

In the case where credit qualities are homogeneous we have that conditional on the realized paths of the systematic factors, the distribution of $I_i$ is determined solely by the values $D_{t_1}, \ldots, D_{t_n}$, where $D_{t_i} = P(\tau_1 \leq t_i | \mathcal{H}_{t_i})$.[4] Moreover, conditional on these realized paths the $I_i$ are independent random variables. Therefore in order to simulate portfolio losses we may carry out the following program

- Simulate $D_{t_1}, \ldots, D_{t_n}$ (or indeed any random vector having the correct joint distribution)
- For each obligor, independently generate a random number $U_i$ and set

$$I_i = \min_{1 \leq j \leq n} \left\{j : U_i \leq D_{t_j}\right\}$$

[4]Recall that $D_t = \Psi_t(A, B)$. We assume for the moment that these variables may be simulated exactly.

Having simulated the interval of default for each name, it is a simple matter to reconstruct the path of portfolio losses. In the more general case where credit qualities are not homogeneous, this program would simply need to be repeated for each group of homogeneous obligors. For the model presented in Section 4.2.2 we simulated losses over a ten-year period at 3-month intervals for a homogeneous portfolio with 125 names. Using Matlab 7 on a PC with a Pentium 4 processor, this required approximately 2.8 seconds for 10,000 trajectories.

It is in fact not always necessary to generate the interval of default for each name. When the portfolio is equally-weighted and recovery rates are constant across obligors, the percentage loss is simply proportional to the proportion of defaults, that is $L_t^N = (1-R)D_t^N$. In order to simulate portfolio losses here, one need not keep track of *who* defaults, as each default impacts the portfolio in exactly the same way. One might imagine that efficiency gains are possible by devising a scheme which only keeps track of the number of defaults in each interval, and not their identity. Assuming credit qualities are homogeneous one such scheme may be constructed as follows. Conditional upon the realized paths of the systematic factors, the law $\left(D_{t_1}^N, \ldots, D_{t_n}^N\right)$ is determined by the values $(D_{t_1}, \ldots, D_{t_n})$. Given the value of $D_{t_{i-1}}^N$, there are $N\left(1 - D_{t_{i-1}}^N\right)$ obligors which have not defaulted, each of whom defaults in the interval $(t_{i-1}, t_i]$ with probability $\left(D_{t_i} - D_{t_{i-1}}\right) / \left(1 - D_{t_{i-1}}\right)$. Therefore in order to simulate portfolio losses we may carry out the following program

- Simulate $D_{t_1}, \ldots, D_{t_n}$
- For each $i$ generate a binomial variate, $Z_i$, with $N\left(1 - D_{t_{i-1}}\right)$ trials and success probability $\left(D_{t_i} - D_{t_{i-1}}\right) / \left(1 - D_{t_{i-1}}\right)$.
- Set $D_{t_{i+1}}^N = D_{t_i}^N + \frac{Z_i}{N}$

In the more general case where credit qualities are not homogeneous, this program would simply need to be repeated for each group of homogeneous obligors. Implementing this scheme in the scenario described in the previous paragraph, we find that simulation of 10,000 trajectories required approximately 6.5 seconds, roughly 6 seconds of which was spent generating the binomial variates. We conclude that, somewhat surprisingly, the more general algorithm described in the previous paragraph, which keeps track of the identity of defaults, is in fact much more efficient.

As a final note we point out that simulating the *asymptotic* proportion of defaults, $D_t$, is often significantly faster than simulating the exact number of defaults

$D_t^N$. Indeed simulating the asymptotic proportion of defaults in the scenario described in each of the previous two paragraphs requires only 0.35 seconds for 10,000 trajectories, nearly one-tenth the time required to simulate the exact number of defaults.

## 4.2 Calibration Results and Discussion

In this section we calibrate several specifications of our framework to market data for index tranche quotes. Each of the models investigated here is specified up to a finite number of parameters, say $\theta \in \mathbb{R}^d$. In each case we assume that credit qualities are homogeneous, recovery rates are 40% for each name and the risk-free term structure of interest rates is flat at 5%. In addition we calibrate the asymptotic version of the model, that is we make the simplifying assumption that the underlying portfolio is infinitely large.

Having specified the parameters of a particular model, tranche spreads are computed via simulation. The general procedure for Monte Carlo valuation of tranche spreads is discussed in Appendix A, and this procedure requires the simulation of portfolio loss trajectories. In light of the assumptions made here, this requires simulation of

$$L_t = (1 - R) D_t = (1 - R) \Psi_t (A, B)$$

In situations where exact simulation is not possible we simulate approximate losses via $\Psi_t (A^n, B^n)$, as discussed in Chapter 3. As the building blocks for tranche spreads are expectations of the form $E\left[(L_t - K)^+\right]$ for various constants $K$, Corollary 3.4.3 justifies such an approximation.

Given market quotes for each of $I$ tranches at $J$ distinct maturities, let $s_{i,j}^{mkt}$ denote the market spread for tranche $i$ at maturity $j$. In addition, for a particular model and parameter choice $\theta$, let $s_{i,j}^{mod}(\theta)$ denote the model-implied spread for the associated tranche and maturity. The units of $s_{i,j}^{mod}(\theta)$ will always be taken to be basis points. As we will typically have nearly twice as many spreads as parameters, a perfect fit to all tranches is not a realistic goal. As such we define a discrepancy function $d(\theta)$ and attempt to minimize the function numerically. Two choices we investigated were the mean average error

$$d_1(\theta) = \frac{1}{IJ} \sum_{i=1}^{I} \sum_{j=1}^{J} \left| s_{i,j}^{mkt} - s_{i,j}^{mod}(\theta) \right|$$

and the mean relative error

$$d_2(\theta) = \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{\left|s_{i,j}^{mkt} - s_{i,j}^{mod}(\theta)\right|}{s_{i,j}^{mkt}}$$

We have found that the choice of objective function can significantly impact the minimization results. Using the mean average error tends to provide a nearly perfect fit to the equity and mezzanine tranches at the expense of the senior tranches. Conversely using the mean relative error tends to provide a very good fit to the senior and super-senior tranches at the expense of the equity and mezzanine. This should not be surprising, as equity spreads are typically in the thousands of basis points, while senior tranches typically trade in the neighbourhood of ten basis points (at least this was true before the credit crunch). Note that, despite the fact that we sometimes report equity spreads as percentage points, all spreads are computed in terms of basis points. This ensures the units are uniform across spreads in the calibration procedure.

For either choice of objective function, the surface $d_i(\theta)$ appears to be highly irregular, with certain parameters being highly influential in certain regions while highly insignificant in others. As a consequence we have found that the "greedy" Nelder-Mead algorithm tends to get trapped in local minima, and "optimal" results are significantly influenced by starting values. As a result we used a stochastic minimization algorithm known as simulated annealing. For an excellent introduction to the general theory of simulated annealing, the reader is referred to Aarts and van Larrhoven [1]. For a detailed exposition of the specific algorithm implemented here the reader is referred to Corana et al. [34]. Our implementation followed the pseudo-code given there virtually word-for-word.

The basic idea behind this algorithm is as follows. Begin with an initial point $\theta^{(0)}$, and select a candidate point $\theta^*$ by randomly perturbing one co-ordinate of the initial point. For example $\theta^*$ might be selected by setting $\theta_k^* = \theta_k^{(0)}$ for $2 \leq k \leq d$ and $\theta_1^* = \theta_1^{(0)} + \epsilon$, where $\epsilon$ is uniformly distributed on the interval $[-c, c]$ for some constant $c > 0$. If $\theta^*$ produces a lower value of the objective function, that is $d(\theta^*) < d\left(\theta^{(0)}\right)$, we accept $\theta^*$ as the next point and set $\theta^{(1)} = \theta^*$. If $\theta^*$ produces a higher value for the objective function we accept it as the next point with probability $Q = \exp\left(\left(d\left(\theta^{(0)}\right) - d(\theta^*)\right)/T\right)$, where the constant $T > 0$ is often referred to as the "temperature." That is, we generate a random number $U$ and set $\theta^{(1)} = \theta^*$ if $U \leq Q$ and $\theta^{(1)} = \theta^{(0)}$ otherwise. The fact that we are allowed to accept points which yield higher values of the objective function allows us to escape from local minima. The algorithm proceeds by cycling through each parameter in order, and

Corana et al. [34] recommend reducing the temperature every 200 cycles, that is after each parameter has been perturbed 200 times. In addition they recommend adjusting $c$ (a different value of which is used for each parameter) every ten cycles.

We have found simulated annealing to be vastly superior to Nelder-Mead. To begin it tends to produce much lower values for the objective function (and hence much better fits to tranche spreads), in many cases up to one-third the value found by Nelder-Mead. Perhaps more importantly, we have found simulated annealing to be quite robust with respect to starting values, and this gives us much more confidence in our calibrated parameters. The downside in using simulated annealing is that it requires significantly more function evaluations. We have found that 20-30 temperatures are typically required for the algorithm to converge. Recalling that each temperature requires 200 function evaluations, this implies that convergence typically requires between four and six thousand function evaluations per parameter. Thus in a model with eight parameters we typically need to value tranches spreads using somewhere between 32,000 and 48,000 different parameter sets. In light of this fact we hope the reader can empathize with our decision to use the asymptotic approximation.

### 4.2.1 A Simple One-Factor Model

In this section we investigate what is perhaps the simplest specification of our general framework, which we term the "random drift" model. In this specification credit qualities are simply Brownian motion with random drift

$$dX_t^i = Mdt + dW_t^i \qquad X_0^i = x_0$$

where $M$ is a random variable. Conditional upon the realized value of $M$ credit qualities are independent Brownian motion with constant drift, as such conditional default probabilities are available in closed form. Recalling that the asymptotic proportion of defaults is equal to the conditional default probability of an arbitrary obligor we obtain

$$D_t = \Phi\left(-\frac{x_0 + Mt}{\sqrt{t}}\right) + e^{-2x_0 M}\Phi\left(\frac{Mt - x_0}{\sqrt{t}}\right) \tag{4.14}$$

It is easy to see that the proportion of defaults is monotone in $M$, with smaller values of this factor producing a larger proportion of defaults. In addition, we note that exact simulation of portfolio losses is possible here.

The distribution of $M$ is arbitrary here, and in order to investigate the importance of heavy tails for the model's ability to price senior tranches accurately, we have attempted to calibrate the model under two different distributional assumptions. In the first case we take $M$ to be Gaussian with mean $\mu$ and variance $\sigma^2$, while in the second case we take $M$ to be a symmetric Laplace variable. The Laplace distribution is discussed in detail in Appendix C, at this point we note that when parametrized by its mean ($\mu$) and variance ($\sigma^2$), the symmetric Laplace density is given by

$$f(m) = \frac{1}{\sqrt{2\sigma^2}} \exp\left(-\frac{\sqrt{2}\,|m-\mu|}{\sigma}\right) \qquad m \in \mathbb{R}$$

The Laplace density thus has much heavier tails than the Gaussian. There are thus three parameters in this model - the mean and variance of the market factor, as well as the initial value of credit quality.

Table 4.1 presents the results of calibrating the model parameters to market data for iTraxx tranches. The data was obtained from Ferrarese [54] and consists of market quotes for five-year tranches as recorded on April 13, 2006. In [54] the author calibrated six different factor models of the form discussed in Section 1.1.2. We reproduce the results found in [54] for the equicorrelated Gaussian copula model, as well as a model with normal inverse Gaussian factors and stochastic correlation (SC NIG in the table). The former is in fact the standard industry model, while the latter provided the best fit to the data among the six models.

Comparing the two random drift models we see that, as would be expected, the heavier-tailed Laplace distribution provides a much better fit to the more senior tranches. We note also that the average five-year CDS spread, which was not included in the calibration, is priced quite well relative to the market. In both cases calibrated parameters are comparable, although presumably owing to its heavier tails the Laplace specification can "get away" with a smaller mean and standard deviation.

Comparing the random drift and factor models, we see that both random drift specifications vastly outperform the industry standard. In addition we see that the random drift model with Laplace drift provides a better overall fit to the data than the best factor model, with a total pricing error of 2.1 basis points as compared to 12.7 basis points.

Figure 4.1 plots sample trajectories for portfolio losses in the random drift model. Using the calibrated parameters we compute the $25^{th}$, $50^{th}$ and $75^{th}$ quantiles of $M$. That is, we determine the unique value of $m$ for which $P(M \leq m) = 0.25$

Table 4.1: Calibration to iTraxx Data. All spreads expressed in basis points and quoted as running premia. Average pricing error used for objective function.

| DJ iTraxx 5Y | | | | | | |
|---|---|---|---|---|---|---|
| Tranche | 0-3% | 3-6% | 6-9% | 9-12% | 12-22% | CDS |
| Market | 1,226 | 63 | 18 | 9 | 4 | 31.5 |
| Random Drift Models | | | | | | |
| Gaussian Drift | 1,226 | 63.1 | 9.5 | 3.2 | 0.5 | 27.7 |
| Laplace Drift | 1,226 | 63 | 18.6 | 7.9 | 3.6 | 29.1 |
| Factor Models | | | | | | |
| Normal Copula | 1,226.7 | 117.5 | 23.4 | 4.7 | 0.4 | 31.5 |
| SC NIG | 1234.1 | 62.1 | 20.6 | 9.5 | 3.4 | 31.5 |

| Parameters | | | |
|---|---|---|---|
| Model | $x_0$ | $\mu$ | $\sigma$ |
| Gaussian | 1.1678 | 1.7966 | .3517 |
| Laplace | 1.4156 | 1.4393 | .2587 |

Table 4.2: Implied Spreads for Longer Maturity

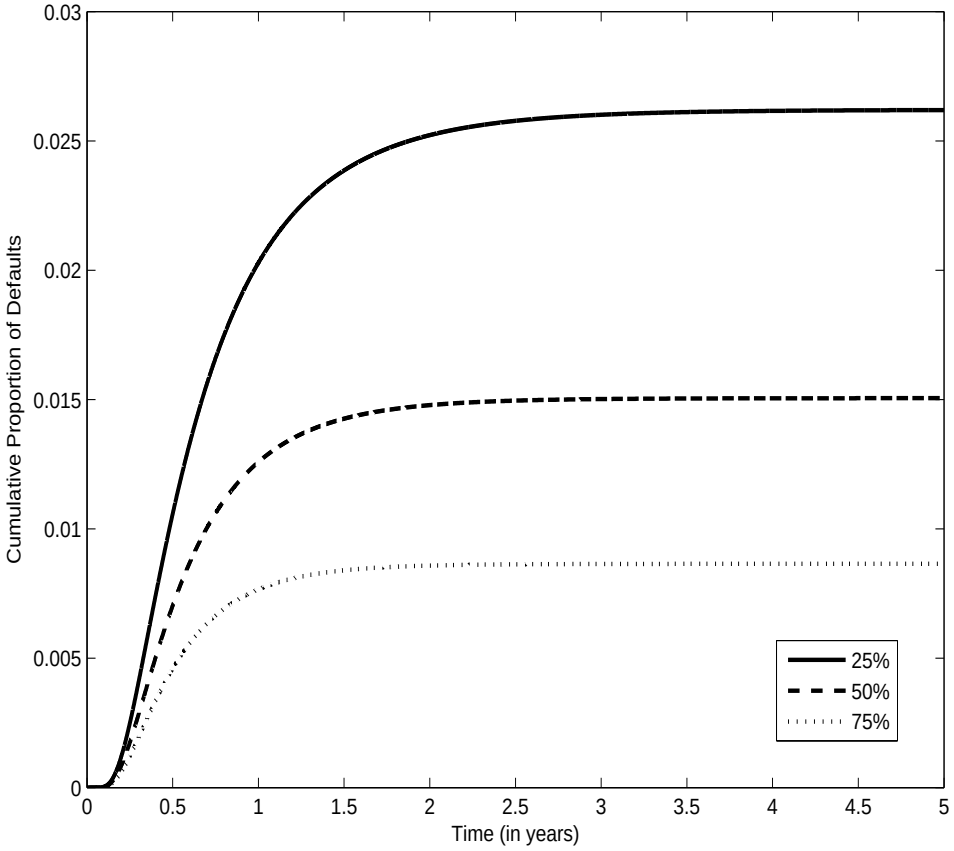
| | 0-3% | 3-6% | 6-9% | 9-12% | 12-22% | CDS |
|---|---|---|---|---|---|---|
| 5Y | 1,226 | 63 | 18.6 | 7.9 | 3.6 | 29.1 |
| 7Y | 950 | 48.8 | 15.3 | 6.5 | 3.1 | 22.1 |
| 10Y | 731 | 37.1 | 12.0 | 5.1 | 2.5 | 16.7 |

and plot the corresponding path for $D_t$, repeating the exercise for 0.5 and 0.75. Inspection of the figure reveals that virtually all defaults occur within the first year or two, and those obligors who survive this period essentially live forever.

This "now or never" phenomenon (obligors either default very soon, or never at all) has significant implications for pricing longer-maturity tranches. Using the calibrated parameters for the five-year spreads and the Laplace model, we calculated the implied fair spreads for longer-dated contracts. Table 4.2 presents the results, and in all cases we see that spreads decrease with maturity, a phenomenon which is most emphatically *not* observed in the market.

Our conclusions from this calibration exercise are as follows. To begin the simple random drift model is quite capable of calibrating to single-maturity quotes. In-

Figure 4.1: Sample Loss Paths in the Random Drift Model



deed this model outperformed both the industry standard Gaussian copula model, as well as the richer factor model with heavy-tailed factors and stochastic correlation. We have also noted that heavy tails are crucial for accurate pricing of more senior tranches. Finally we have pointed out the dangers in calibrating to only one particular maturity, which can lead to nonsensical spreads for longer-dated tranches.

### 4.2.2 Linear Models

In this section we investigate a slightly more general specification of our framework by introducing stochastic volatility in credit qualities. In this case we obtain

$$dX_t^i = Mdt + \sqrt{V}dW_t^i \qquad X_0^i = x_0$$

where $M$ and $V$ are (not necessarily independent) random variables. Conditional upon the systematic factors, credit qualities are simply independent Brownian motion with drift. Hence conditional default probabilities are simply first-passage probabilities for such processes to the level zero, and are available in closed form

$$D_t = \Phi\left(-\frac{x_0 + Mt}{\sqrt{Vt}}\right) + e^{-2x_0M/V}\Phi\left(\frac{Mt - x_0}{\sqrt{Vt}}\right) \tag{4.15}$$

As with the random drift model, we note that exact simulation of portfolio losses is possible (and quite easy) here.

In order to model the joint distribution of $(M, V)$ we employ the copula approach. Owing to the success of the Laplace distribution in the previous section, we calibrate this model under the assumption that the marginal distributions of $M$ and $V$ are Laplace and log-Laplace, respectively, and tie these margins together with a Gaussian copula. The properties of the Laplace and log-Laplace distributions are reviewed in Appendix C. There are eight parameters in this model - three for each of the marginal densities, the copula parameter $\rho$, and the initial level of credit quality $x_0$.

Table 4.3 presents our preliminary calibration results. The data was obtained from DiGraziano and Rogers [41] and consists of market quotes for CDX tranches at each of three maturities, as of November 1, 2006. In [41] the authors calibrate an intensity model consisting of 32 parameters, and obtain a more or less perfect fit. In order to truly test the model's abilities we have included the super-senior tranches (30-100%) in the calibration. Also note that we do not include default swap spreads in the calibration.

Table 4.3: Calibration to CDX Data - Linear Model. Equity spreads expressed in percentage points and quoted as upfront fees with 500 basis point running premium. All other spreads expressed in basis points and quoted as running premia. Relative pricing error used as objective function.

| | 5Y | | | | | | |
|---|---|---|---|---|---|---|---|
| | 0-3% | 3-7% | 7-10% | 10-15% | 15-30% | 30-100% | CDS |
| Market | 24.38 | 90 | 19 | 7 | 3.5 | 1.73 | 35 |
| Model | 24.43 | 90.2 | 17.5 | 7 | 2.5 | 0.38 | 34.8 |
| | 7Y | | | | | | |
| | 0-3% | 3-7% | 7-10% | 10-15% | 15-30% | 30-100% | CDS |
| Market | 40.44 | 209 | 46 | 20 | 5.75 | 3.12 | 45 |
| Model | 40.61 | 250.5 | 45 | 20 | 9.3 | 2 | 47.3 |
| | 10Y | | | | | | |
| | 0-3% | 3-7% | 7-10% | 10-15% | 15-30% | 30-100% | CDS |
| Market | 51.25 | 471 | 112 | 53 | 14 | 4 | 57 |
| Model | 49.1 | 471.1 | 112 | 44 | 19.8 | 4 | 57.5 |

| Marginal Parameters | | | |
|---|---|---|---|
| | $\alpha$ | $\beta_1$ | $\beta_2$ |
| $M$ | .0835 | .0514 | .0706 |
| $V$ | -1.4958 | .2809 | .6399 |

| Other Parameters | | |
|---|---|---|
| | $x_0$ | $\rho$ |
| | 1.8371 | .8908 |

The results presented here are quite encouraging. We obtain a uniformly good fit across both maturities and tranches, with only two egregious pricing errors in the five-year super-senior and seven-year mezzanine spreads. Of particular note is the fact that swap spreads, which were not included in the calibration, are priced quite accurately.

Having calibrated the model, we asked ourselves the question "what does it take for the super-senior tranches to experience losses here?" In order to gain insights into the answer to this question we performed the following experiment. Using the calibrated parameters we simulated 10,000 realizations of $(M, V)$, which can be thought of as 10,000 possible scenarios for the economic conditions in which obligors operate. Recall that this is a very loose interpretation of the systematic factors, however we find that it facilitates discussion quite well. Among these 10,000 simulated scenarios the five-year super-senior tranche experienced losses only thirteen times. Our first observation when examining these scenarios was that they are all characterized by abnormally low values of both $M$ and $V$. Letting $v_i$ denote the realized value of $V$ on the $i^{th}$ such scenario, it turns out that the maximum value of $P(V \leq v_i)$ is 3.15%, while the average of these values is 0.44%. The analogous quantities for $P(M \leq m_i)$ are 0.25% and 0.09%, respectively.

When we originally came to this discovery we must admit that we were slightly disappointed. Our naive intuition was that, since low volatility is "good" and high volatility is "bad," the predictions of this model were somehow "nonsensical." However the nature of market crashes in this model has a very sensible explanation which coincides nicely with empirical observations. For the moment let us characterize economic downturns as those scenarios in which $M$ is abnormally low. A heuristic justification of this interpretation will be given momentarily. With this interpretation, and in light of the strong positive correlation between $M$ and $V$, we see that these downturns also tend to be accompanied by abnormally small values of $V$. Now, $X_t^i = x_0 + Mt + \sqrt{V}W_t^i$ and we see that when $V$ is close to zero the idiosyncratic component $W_t^i$ has very little impact on credit quality. Thus we reach the conclusion that *in a severe recession all risk is systematic.* It is tempting to conclude that this coincides with the empirically observed phenomenon that correlations between many financial assets tend to be significantly higher in recessionary periods. It is important to remember, however, that conditional upon the systematic factors credit qualities are totally independent in our framework, hence there is zero conditional correlation between different obligors. Nonetheless in these crash scenarios their behaviour is virtually identical, namely processes characterized by a linear trend with very small fluctuations about that trend. Thus we might be

tempted to conclude that there is significant "apparent correlation" between credit qualities in a severe recession, and interpreting this phrase very loosely.

A heuristic justification for interpreting downturns as being characterized by abnormally small values of $M$ is as follows. To begin, note that by the law of the iterated logarithm for Brownian motion (see Karatzas and Shreve [80]) we have that

$$\sqrt{V}W_t^i = O\left(\sqrt{2t\log\log t}\right) = o(t)$$

almost surely as $t \to \infty$. Hence

$$\frac{X_t^i}{t} = M + o(t)$$

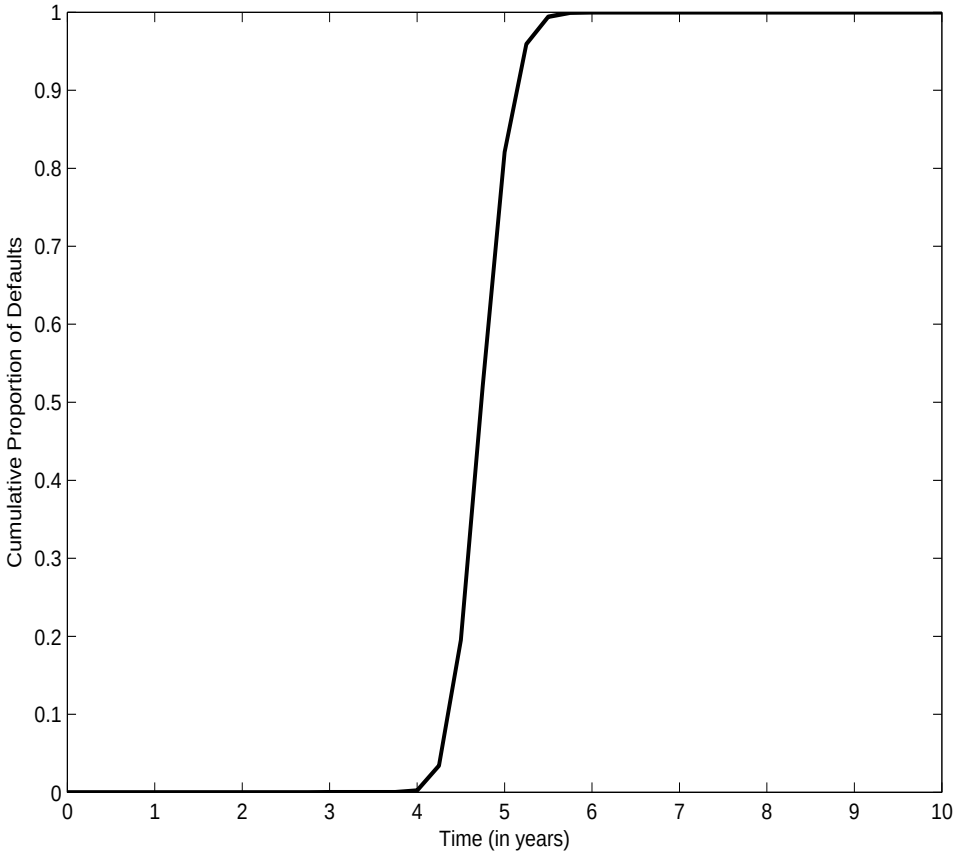
and we may indeed interpret $M$ as the dominant factor with respect to the long-term behaviour of credit qualities. This in turn lends credence to our characterization of recessionary periods begin characterized by abnormally low values of this factor.

Figure 4.2 plots the realized trajectory of portfolio losses in one of our "crash" scenarios, corresponding to simulated values of $M = -0.3873$ and $V = .0025$. We note that $P\left(M \leq -0.3875\right) = 0.07\%$ and $P\left(V \leq .0025\right) = 0.06\%$. As the idiosyncratic component of credit qualities is negligible here, we see that in this scenario credit qualities behave much like the deterministic and linear process $1.8371 - .3875t$, which strikes zero with certainty at time $t^* = 4.7434$. From the figure we see that indeed, virtually all obligors default within a very short window of this critical time. These features are in fact characteristic of each of our "crash scenarios," namely 100% of the portfolio defaulting within a very short window of the critical time $-x_0/M$. If only the realized trajectory of credit quality was available to market participants, the default event would be predictable in the traditional sense, meaning investors would know that default is imminent immediately prior to the occurence of said event. However if the realized values of the systematic factors were available to market participants, the default event of an individual obligor would become what we call "super-predictable," meaning that the default time can be predicted with near-certainty approximately four years in advance. As alluded to in Section 4.1, in order to avoid this "super-predictability" we assume that the systematic factors are unobserved by market participants.

We now proceed with a more formal investigation of this phenomenon of "low-volatility market crashes." Let us begin by defining the function

$$h(m, v, x_0, t) = \Phi\left(-\frac{x_0 + mt}{\sqrt{vt}}\right) + e^{-2x_0 m/v}\Phi\left(\frac{mt - x_0}{\sqrt{vt}}\right) \tag{4.16}$$

Figure 4.2: Market Crashes in the Linear Model



and note that $h$ is simply the probability that a linear Brownian motion $x_0 + mt + \sqrt{v}W_t$ strikes zero by time $t$. Our original presumption, admittedly driven largely by financial lore, was that low volatility is "good" and high volatility is "bad." As such we expected the function $h$ to be increasing in $v$. Indeed for all values of $(m, x_0, t)$ it can be shown that

$$\lim_{v\to\infty} h(m, v, b, t) = 1$$

hence such processes immediately hit zero as volatility increases without bound. In the context of our financial model this means that all obligors default instantly. The behaviour of $h$ when $v$ is close to zero is in fact ambiguous, as it depends on the values of the other arguments. For values of $(m, x_0, t)$ such that $x_0 + mt > 0$ we have

$$\frac{\partial h}{\partial v} > 0 \qquad \lim_{v\to 0} h(m, v, x_0, t) = 0$$

indicating that our naive presumption was in fact correct in this case. However when $(m, x_0, t)$ are fixed such that $x_0 + mt < 0$ we have

$$\lim_{v\to 0} h(m, v, x_0, t) = 1$$

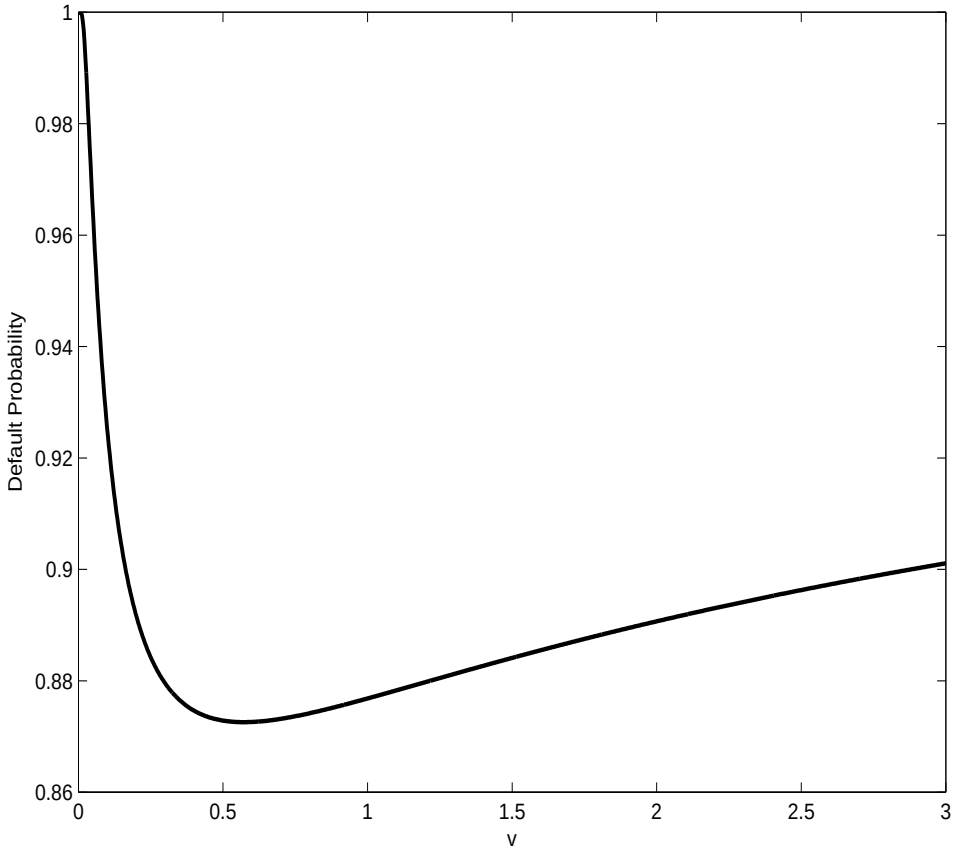
Thus *as volatility disappears obligors default with certainty.* This fact may seem counter-intuitive at first, yet it has a simple explanation. When $v = 0$ credit quality is the deterministic and linear process $x_0 + mt$. Thus all obligors default at exactly the same moment in time, namely $t^* = -\frac{x_0}{m}$. When $v$ is small but non-zero the presence of stochastic behaviour serves to make avoiding default possible (though not terribly likely), and default probabilities initially decrease as stochastic behaviour is introduced.

Figure 4.3 plots the function $h(-0.4, v, 0.6, 3)$, which represents three-year default probabilities as a function of volatility when $x_0 = 0.6$ and $m = -0.4$. We see that the introduction of stochastic behaviour initially serves to lower default probabilities, but that this effect eventually "wears off," as default probabilities again begin to increase with volatility. As a final note we point out $\partial h/\partial m < 0$ for all values of $(m, v, x_0, t)$, indicating that default probabilities are in fact monotone in the factor $M$, and there is no ambiguity here.

It is also worthwhile to investigate the behaviour of conditional default probabilities in terms of the dimensionless quantities $A = x_0/\sqrt{V}$ and $B = M/\sqrt{V}$. It is straightforward to see that the function $h$ defined by (4.16) is equal to

$$h(m, v, x_0, t) = H\left(\frac{x_0}{\sqrt{v}}, \frac{m}{\sqrt{v}}, t\right)$$

Figure 4.3: Effect of Volatility on Default Probabilities



where

$$H(a,b,t) = \Phi\left(-\frac{a+bt}{\sqrt{t}}\right) + e^{-2ab}\Phi\left(\frac{bt-a}{\sqrt{t}}\right) \tag{4.17}$$

is the probability that a standard Brownian motion strikes a linear barrier, with initial value $a$ and slope $b$, by time $t$. It is clear that $H$ is monotone increasing in both $a$ and $b$, since the barrier becomes more remote as these quantities are increased. Thus the effect of the *dimensionless* quantities on conditional default probabilities is unambiguous. It is interesting to note here that our previous discussion concerning the effect of $m$ and $v$ indicates that if $a \to \infty$ and $b \to -\infty$ while the ratio $a/b$ is held fixed at a negative value, then $H$ tends to one for any fixed $t > 0$. Thus in terms of crossing probabilities for a standard Brownian motion to a linear barrier, the effect of an infinitely negative slope dominates that of an infinitely remote initial position.

### Calibrating to Distressed Data

Recent times have seen significant turmoil in markets for credit derivatives. In early 2008 spreads for all tranches were significantly higher than historical levels, with senior tranches trading at spreads in excess of 115 basis points, more than thirty times their typical pre-crisis levels. We obtained a set of CDX quotes as of March 10, 2008 from Krekel [83], according to whom

> In February and March 2008 it was temporarily not possible to calibrate the standard Gaussian base correlation model to the complete set of CDX and iTraxx tranche quotes. The reason is that the Gaussian base correlation model was not able to generate enough probability for high portfolio losses, while preserving the calibration to mezzanine and equity tranches.

In light of this fact we felt it would be instructive to investigate the performance of our model under these harsher conditions. Table 4.4 presents the calibration results. We note that neither super-senior tranche nor default swap data was available. The results are again encouraging, with the overall relative pricing error at 5.3%. The model is able to capture the senior spreads reasonably well, while at the same time maintaining an acceptable fit to the equity and mezzanine.

Let us now compare the two sets of calibrated parameters. As expected the initial value of credit quality is much lower for the distressed data, indicating that the market feels obligors are in a much more precarious position now than they

Table 4.4: Calibration to Distressed CDX Data - Linear Model. Equity spreads expressed in percentage points and quoted as upfront fees with 500 basis point running premium. All other spreads expressed in basis points and quoted as running premia. Relative pricing error used as objective function.

| | 5Y | | | | |
|---|---|---|---|---|---|
| | 0-3% | 3-7% | 7-10% | 10-15% | 15-30% |
| Market | 67.38 | 727 | 403 | 204 | 115 |
| Model | 65.90 | 733 | 355 | 219 | 100 |
| | 7Y | | | | |
| | 0-3% | 3-7% | 7-10% | 10-15% | 15-30% |
| Market | 70.5 | 780 | 440 | 248 | 128.5 |
| Model | 70.79 | 859 | 417 | 265 | 128.1 |
| | 10Y | | | | |
| | 0-3% | 3-7% | 7-10% | 10-15% | 15-30% |
| Market | 73.5 | 895.5 | 509 | 282 | 139.5 |
| Model | 71.76 | 894.7 | 430 | 277 | 141.2 |

| Marginal Parameters | | | |
|---|---|---|---|
| | $\alpha$ | $\beta_1$ | $\beta_2$ |
| $M$ | .0831 | .01 | .0534 |
| $V$ | -3.2536 | .0271 | .1455 |

| Other Parameters | | |
|---|---|---|
| | $x_0$ | $\rho$ |
| | 0.5865 | .8217 |

were two years ago. In order to compare the marginal parameters for $M$ we plot the two calibrated densities in the top panel of Figure 4.4. As the location parameters are almost identical, the major difference between these densities is that the "distressed" density is skewed much more heavily to the left. This may be taken as a sign that the market feels major economic downturns are much more likely today than they were two years ago. Again this is not a surprise. The most striking difference between the two sets of parameters is revealed by inspecting the calibrated densities for $V$. These are plotted in the lower panels of Figure 4.4, and we encourage the reader to note the two different scales on which these are plotted. The distressed density is *much* more heavily concentrated about small values, and the difference is quite staggering - the essential support of $V$ in recent times $[0, 0.2]$, an interval which was assigned a probability of approximately 58% by the "normal" data. Recalling that, when $V$ is small, idiosyncratic risk has very little effect on credit qualities, we may interpret these findings as an indication that, at the present time, idiosyncratic risk is insignificant relative to systematic forces.

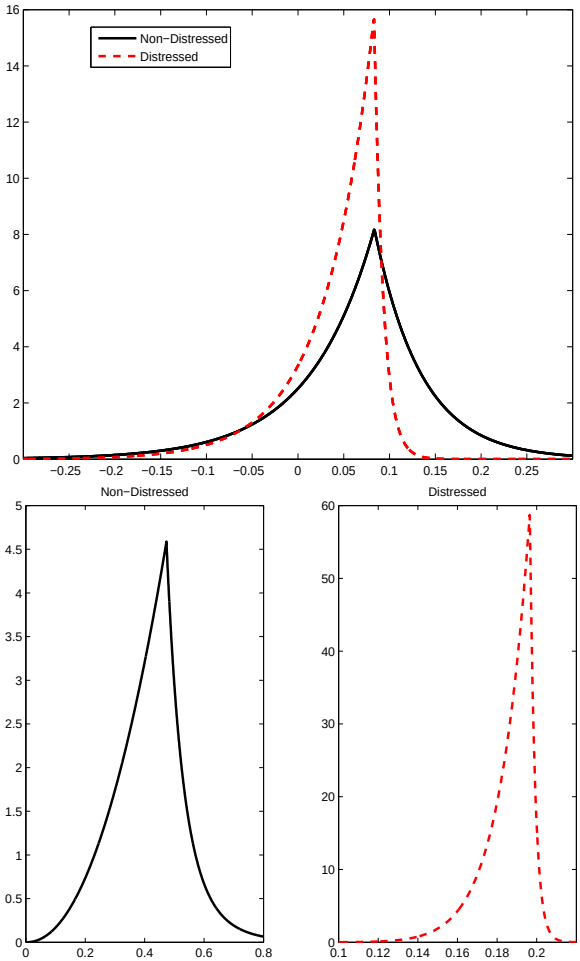
Based our findings here we feel that a re-interpretation of $V$ may be in order. It appears as if the data "wants" this factor to play a role akin to a stochastic correlation factor, reflecting the dependence between credit qualities. Moreover this dependence should be stronger in downturns, with $V$ downgrading the relative importance of idiosyncratic risk in such times. We believe that such a re-interpretation could serve to enhance the modeling process, and will have more to say on the subject in our concluding chapter.

### Theoretical Shortcomings

We are pleased with the fact that the model presented in this section calibrates well to market data, and predicts behaviour in severe downturns that is consistent with empirical fact. However no model is perfect, and we now discuss two theoretical shortcomings of our linear specification. To begin, at least under "normal conditions," super-senior tranches experience losses only when 100% of the portfolio defaults, and we observe an "all-or-nothing" phenomenon similar to the "now-or-never" phenomenon of the random drift model. We may classify these scenarios as severe economic downturns, and due to the lack of time dynamics in the systematic factors, there is essentially no recovering from a recession here. The systematic components are realized today, and the economic environment in which obligors operate remains frozen forever.

Before proceeding with the second shortcoming, we make the following obser-

Figure 4.4: Calibrated Densities



vation. Suppose that $a_i + b_i t$ are two linear functions with $a_i > 0$, and let $F_i$ denote the cumulative distribution function of the first passage time of a standard Brownian motion to the $i^{th}$ barrier. It is a fact that $F_1 = F_2$ if and only if $a_1 = a_2$ and $b_1 = b_2$. To see this let $f_i$ denote the associated densities, and note that $F_1 = F_2$ if and only if $f_1 = f_2$. Moreover if $f_1 = f_2$ then the ratio $\log(f_1(t)/f_2(t))$ must be identically zero, which is tantamount to the following polynomial being identically zero

$$\left(a_2^2 - a_1^2\right) + 2t\left[\log(a_1/a_2) - a_1 b_1 + a_2 b_2\right] + t^2\left(b_2^2 - b_1^2\right)$$

which is only possible if $a_1 = a_2$ and $b_1 = b_2$. In words we see that there is a one-to-one correspondence between linear functions and associated first-passage distributions.

Returning to our model we note that conditional on the realized values of the systematic factors the asymptotic proportion of defaults is simply the cumulative distribution function of the first passage time of a standard Brownian motion to a straight line with initial value $x_0/\sqrt{V}$ and slope $M/\sqrt{V}$. Thus if one observes the trajectory of $D_t$ over any interval $[0, T]$, one can "back out" the realized values of these quantities. In practice a simple iterative procedure to determine the initial value and slope often converges quite rapidly. Having determined the realized values of $x_0/\sqrt{V}$ and $M/\sqrt{V}$ our observer can now predict the future evolution of portfolio losses *with certainty*, which is clearly at odds with reality.

### 4.2.3 Dynamic Models

Despite the fact that the linear model calibrates quite well to market data, it suffers from significant shortcomings. In particular an observer of asymptotic losses over any time interval may predict future losses with certainty. This problem may be traced to the fact that the "market environment" in which firms operate is ostensibly determined "today," and is not permitted to evolve over time. We have seen that this implies that there is in effect no recovering from economic downturns, with the result that super-senior tranches may only experience losses in "doomsday" scenarios where 100% of a portfolio defaults in a finite time.

Presumably these problems may be "solved" by introducing time dynamics to the systematic factors. To this end, in this section we investigate the homogeneous model

$$dX_t^i = M_t dt + \sqrt{V_t} dW_t^i \qquad X_0^i = x_0$$

where $M_t$ and $V_t$ are now mean-reverting diffusion processes

$$\begin{aligned} dM_t &= \theta_M (\mu_M - M_t)\, dt + \sigma_M (M_t)\, dZ_t^1 \\ dV_t &= \theta_V (\mu_V - V_t)\, dt + \sigma_V (V_t)\, dZ_t^2 \end{aligned}$$

Here $Z^1, Z^2$ are correlated standard Brownian motion, independent of the sequence $W^1, W^2, \ldots$. Owing to the success of the Laplace and log-Laplace distributions in the previous section, we use the results of Bibby et al. [17] to choose the functions $\sigma_M, \sigma_V$ in such a way as to ensure $M$ and $V$ are ergodic, with Laplace and log-Laplace invariant densities, respectively. Details of these constructions are contained in Appendix C, and explicit expressions for $\sigma_M$ and $\sigma_V$ are given by (C.6) and (C.7), respectively.

In this dynamic model we have that for fixed $T > 0$, conditional default probabilities, and hence asymptotic portfolio losses, are given by

$$D_T = P(\tau_i \leq T \,|\mathcal{H}_T) = \Psi_T(A, B)$$

where $\Psi$ is defined and studied in Section 3.4 and

$$A = \left\{ x_0 + \int_0^t M_s ds \ : 0 \leq t \leq T \right\} \qquad B = \left\{ \int_0^t V_s ds \ : 0 \leq t \leq T \right\}$$

Exact simulation of $D_T$ is not possible, hence we must rely on the methods discussed in Chapter 3. The first step in simulating portfolio losses is the simulation of paths for $M_t$ and $V_t$. To this end we use a simple Euler scheme with a time-step of $h = 10^{-2}$. That is, we generate the systematic factors recursively via

$$\begin{aligned} M_{t_{i+1}} &= M_{t_i} + h\theta_M (\mu_M - M_{t_i}) + \sqrt{h\sigma_M (M_{t_i})} Y_i^1 \\ V_{t_{i+1}} &= V_{t_i} + h\theta_V (\mu_V - V_{t_i}) + \sqrt{h\sigma_V (V_{t_i})} Y_i^2 \end{aligned}$$

where $t_i = ih$ and $(Y_i^1, Y_i^2)$ are correlated standard normal variables, the correlation being set equal to the correlation between $Z_t^1$ and $Z_t^2$. As a final note, we point out that due to the correlation between $Z^1$ and $Z^2$, the bivariate stationary distribution of $(M_t, V_t)$ is non-trivial (note that the marginal stationary distributions are available). In order to produce a pair $(M_0, V_0)$ which is approximately distributed according to the correct bivariate distribution,[5] we "begin" these processes five years in advance, where they are set equal to their long-run means. That is, we set $M_{-5} = \mu_M$ and $V_{-5} = \mu_V$ and use our Euler scheme to simulate the values $(M_0, V_0)$.

[5] Assuming the initial values of the systematic factors are distributed according to their stationary distribution is consistent with the assumption that these factors are unobservable.

Having simulated paths for the systematic factors, the next step is to construct linear approximations to their integrals. Clearly we must use the type two approximation, and for reasons of computational efficiency we choose node spacings of $\Delta = 0.25$ and set

$$A_n(t_i) = x_0 + \Delta \sum_{k=1}^{i} M_{t_k} \qquad B_n(t_i) = \Delta \sum_{k=1}^{i} V_{t_k}$$

where $t_i = i\Delta$. Note that not all simulated values of $(M_t, V_t)$ are used in constructing the linear approximation. In order that those values which *are* used are accurately simulated, we are required to generate more than we need. Armed with the simulated linear approximations $A_n, B_n$ it is a simple matter to obtain the associated realization of the default proportion $D_T = \Psi_T(A_n, B_n)$. Again, for reasons of computational efficiency we use the Monte Carlo technique as opposed to numerical integration.

In order to verify that the linear approximation is justified in this case, we first note that $M_t$ has continuous sample paths. In addition it can be shown (see [17]) that the origin is not attainable for $V_t$ here, that is $P(V_t > 0 \quad \forall t \in [0, T]) = 1$. As the origin is unattainable, and as sample paths of $V$ are continuous, almost all paths of this process are bounded away from zero on any bounded time interval. As a result the conditions of Corollary 3.4.3 are satisfied.

There are ten parameters in this model. These include the initial value of credit quality $x_0$ as well as the correlation $\rho$ between $Z^1$ and $Z^2$. Each systematic factor also has four "marginal" parameters. These include the mean reversion rates $\theta_M$ and $\theta_V$, as well as the three parameters governing each invariant density. These parameters influence both the stationary distributions, on which their influence is straightforward, as well as the diffusion coefficients $\sigma_M$ and $\sigma_V$.

Table 4.5 presents the calibration results for both of our sets of data. In both cases we obtain a better overall fit than with the linear model. For the "non-distressed" 2006 data the average relative pricing error here is 15.14 basis points, as compared to 16.45 for the linear model. We note that despite the fact we obtain a better overall fit here, we obtain a slightly worse fit to the super-senior tranches. Turning to the "distressed" 2008 data we see that we obtain a better fit than with the linear model. The mean relative pricing error here is 3.36 basis points, as compared to 5.3 for the linear model. In addition the senior tranches, the bane of the industry-standard model, are priced to a total average error of only 1.2 basis points, with the maximum pricing error being 0.8 basis points for the seven-year tranche.

Table 4.5: Calibration to CDX Data - Diffusion Model. Equity spreads expressed in percentage points and quoted as upfront fees with 500 basis point running premium. All other spreads expressed in basis points and quoted as running premia. Relative pricing error used as objective function.

| Non-Distressed Data | | | | | | | |
|---|---|---|---|---|---|---|---|
| | 5Y | | | | | | |
| | 0-3% | 3-7% | 7-10% | 10-15% | 15-30% | 30-100% | CDS |
| Market | 24.38 | 90 | 19 | 7 | 3.5 | 1.73 | 35 |
| Model | 22.30 | 89.4 | 19.1 | 8 | 3.5 | 0.36 | 33.6 |
| | 7Y | | | | | | |
| | 0-3% | 3-7% | 7-10% | 10-15% | 15-30% | 30-100% | CDS |
| Market | 40.44 | 209 | 46 | 20 | 5.75 | 3.12 | 45 |
| Model | 40.90 | 235.7 | 46.5 | 18.9 | 6.21 | 1.39 | 46 |
| | 10Y | | | | | | |
| | 0-3% | 3-7% | 7-10% | 10-15% | 15-30% | 30-100% | CDS |
| Market | 51.25 | 471 | 112 | 53 | 14 | 4 | 57 |
| Model | 51.04 | 471 | 110.5 | 42 | 14 | 1.49 | 55.4 |

| Distressed Data | | | | | |
|---|---|---|---|---|---|
| | 5Y | | | | |
| | 0-3% | 3-7% | 7-10% | 10-15% | 15-30% |
| Market | 67.38 | 727 | 403 | 204 | 115 |
| Model | 64.71 | 727 | 376 | 223 | 115 |
| | 7Y | | | | |
| | 0-3% | 3-7% | 7-10% | 10-15% | 15-30% |
| Market | 70.5 | 780 | 440 | 248 | 128.5 |
| Model | 70.46 | 842 | 437 | 263 | 129.3 |
| | 10Y | | | | |
| | 0-3% | 3-7% | 7-10% | 10-15% | 15-30% |
| Market | 73.5 | 895.5 | 509 | 282 | 139.5 |
| Model | 71.89 | 899.6 | 452 | 282 | 139.1 |

Table 4.6: Calibrated Parameters - Diffusion Model

| Parameter | Non-Distressed | Distressed |
|---|---|---|
| $x_0$ | 1.1155 | 2.5010 |
| $\theta_M$ | 0.5657 | 0.4638 |
| $\theta_V$ | 0.1252 | 2.4119 |
| $\mu_M$ | 0.0274 | 0.1968 |
| $\mu_V$ | 0.0627 | 0.6327 |
| $\nu_M$ | 0.0693 | 0.3244 |
| $\nu_V$ | 0.0567 | 0.2552 |
| $\rho$ | 0.9810 | 0.9967 |

We turn now to a comparison of the calibrated parameters, which are presented in Table 4.6. The table lists calibrated mean reversion rates $\theta_M, \theta_V$, mean reversion levels $\mu_M = E[M_t]$,$\mu_V = E[V_t]$, as well as calibrated standard deviations $\nu_M = \sqrt{\text{Var}(M_t)}$ and $\nu_V = \sqrt{\text{Var}(V_t)}$. We found this to be more instructive than comparing the calibrated parameters of the invariant densities directly.

The results here are rather surprising at first glance. To begin we note that the initial value of credit quality is *larger* for the distressed data than it is for the non-distressed, which appears at first glance to be counter-intuitive. A similar phenomenon is found when comparing mean reversion levels, we see that for both factors, the long-run means $\mu_M = E[M_t]$ and $\mu_V = E[V_t]$ are almost ten times *larger* in 2008 than in 2006. This stands in stark contrast to the linear model, where both systematic factors had "distressed" distributions which were more heavily concentrated and skewed towards smaller values. Here the calibrated invariant densities are much more heavily concentrated towards larger values of the systematic factors. Making only these two comparisons it might seem surprising that the 2008 parameters actually produced much larger spreads. It appears that the answer to this "paradox" lies in the calibrated variances for the systematic factors. In both cases the systematic factors are roughly 4.5 times as volatile in 2008 as they were in 2006.

Default probabilities and the portfolio loss distribution in this model are the result of a complex interplay between the model parameters. In order to generate large enough spreads, the model must make the two systematic processes much more volatile. However in order to prevent spreads from become too large, the model must compensate by starting obligors farther from the origin, and raising

the long-run trends of the systematic factors.

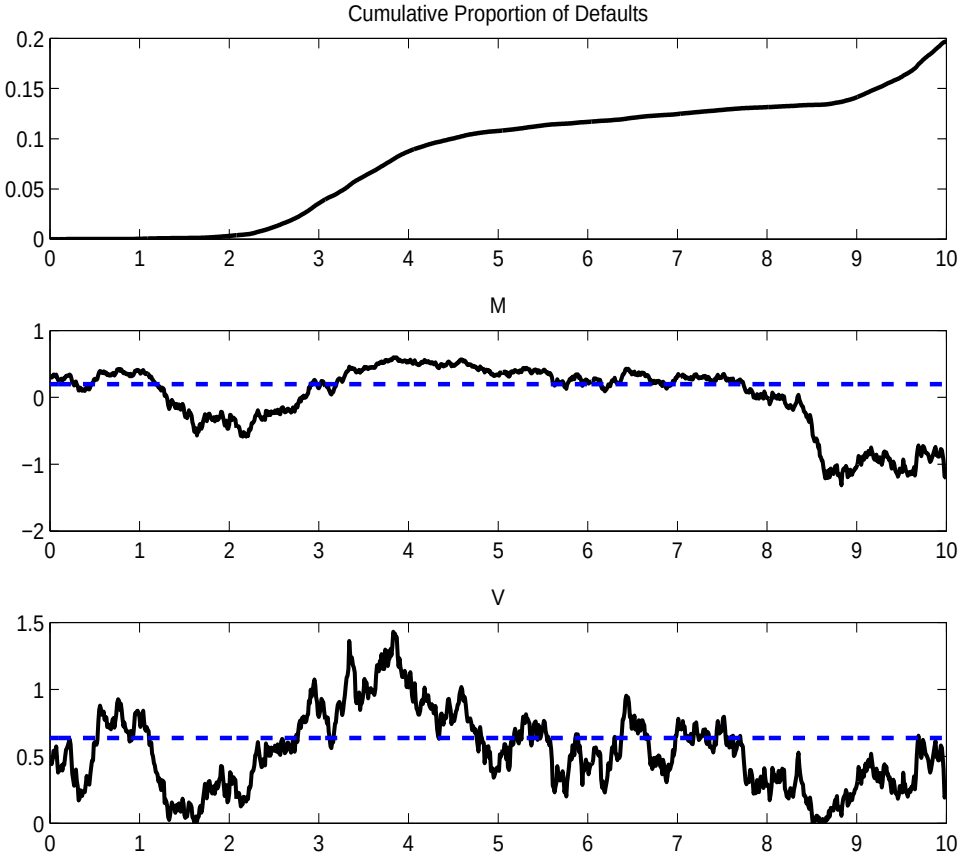
Figure 4.5 plots representative sample paths for the systematic factors, as well as the associated path for portfolio losses. Paths were simulated using an Euler time-step of $h = 10^{-2}$, and portfolio losses were approximated using node spacings equal to the Euler time-step, that is $\Delta = 10^{-2}$. This trajectory is typical in the sense that periods where defaults "cluster," as in the 2-4 year window here, are characterized by prolonged periods where $M$ is far below its long-run mean. Moreover given the extremely strong correlation between the Brownian motions driving the systematic factors, these periods also tend to be accompanied by periods where $V_t$ is near zero. Thus we reach the same conclusion as in the linear model, namely market crashes tend to be driven by the factor $M_t$, and idiosyncratic risk is dwarfed by systematic risk in these periods. It is also interesting to note that it takes time for a market crash to feed its way into portfolio losses. In the figure $M_t$ takes a dive after approximately one year, yet its effect is not felt by the portfolio until two years. This indicates that short-term downturns do not cause large portfolio losses, in order that 10% of the portfolio defaulted here $M$ had to experience a prolonged depression.

The tail end of Figure 4.5 also reveals a very pleasing property of our dynamic model. We note that the market crashed here between 2 and 4 years, yet recovered thereafter. In the linear model we saw that there was in effect no recovering from recession. Moreover after eight years there was another pronounced downswing, which led to another period of default clustering.

To summarize we feel that the phenomena illustrated by Figure 4.5 captures the essence of what we were trying to accomplish with a dynamic model. The almost cyclical nature of portfolio losses reflects the fact that we have produced a dynamic economic environment in which obligors operate. In addition suppose that one observes asymptotic losses over a fixed interval, say zero to five years. Even if one could somehow back out the associated realized paths for the systematic factors, one could not possibly predict the calamity which befell the economy after eight years (though one could in principle assess the *likelihood* of this event).

As a final note we point out that, given the high correlation between the systematic factors here, a one-factor model where $V_t = f(M_t)$ may be just as appropriate as our two-factor specification. In addition, given the somewhat counter-intuitive calibration results, perhaps the Laplace diffusion is not the most appropriate choice for $M_t$. Nonetheless we feel we have accomplished what we set out to do, namely develop a dynamic model which does a good job of describing market data. More

Figure 4.5: Illustrative Path - Dynamic Model



appropriate specifications will be the subject of future research.

## 4.3 Variance Reduction

In this section we take a deeper look at Monte Carlo estimators of fair tranche spreads in our model. In particular our goal is to enhance the efficiency (i.e. reduce the variability) in the "crude" estimators used in Section 4.2. To this end suppose that all parameters of the model have been specified, and that the risk-neutral probability measure is $P$. Further suppose that we are interested in estimating the fair spread on a particular tranche, say the super-senior tranche, with maturity $T$. Letting $L = \{L_t : 0 \leq t \leq T\}$ denote the trajectory of portfolio losses, the fair spread on this tranche may be expressed as

$$s = \frac{E\left[d\left(L\right)\right]}{E\left[p\left(L\right)\right]} \tag{4.18}$$

where $d(L)$ is the present value of the default leg for the tranche, and $p(L)$ is the present value of the protection leg. Note that in order to keep the discussion general we have not specified whether we are working with the asymptotic version of the model, and also note that our tranche experiences losses if and only if $d(L) > 0$.

In Section 4.2 we employed the following "crude" scheme for estimating fair tranche spreads

- Simulate independent portfolio loss paths $L_i$.
- Compute the associated present values of the default and protection legs, $d\left(L_i\right)$ and $p\left(L_i\right)$
- Estimate $s$ via

$$\hat{s} = \frac{\frac{1}{n}\sum_{i=1}^{n} d\left(L_i\right)}{\frac{1}{n}\sum_{i=1}^{n} p\left(L_i\right)} \tag{4.19}$$

As tranche losses tend to be rare for all but the equity and mezzanine tranches, we might view estimation of the numerator in (4.18) as the most troublesome component of such a scheme. That is, we might expect that most of the variability in (4.19) may be traced to the numerator of that quantity, namely the unbiased estimator for expected tranche losses

$$\frac{1}{n}\sum_{i=1}^{n} d\left(L_i\right) \tag{4.20}$$

Indeed in Section 4.2.2 we saw that a sample of 10,000 simulated scenarios produced only 13 in which the super-senior tranche experienced losses. Thus 9,987 scenarios, or 99.87% of the simulations, are essentially uniformative with respect to the default leg of this tranche. Put another way, though the original Monte Carlo sample size was 10,000, the effective sample size for estimating expected tranche losses is only 13.

Importance sampling is a technique which can be incredibly useful in precisely such situations. To this end suppose that $\tilde{P}$ is a measure which is equivalent to $P$, and let the random variable $\Lambda = dP/d\tilde{P}$ denote the likelihood ratio (Radon-Nikodym derivative) between the two measures. Further suppose that tranche losses are much less rare under $\tilde{P}$, for example under this measure the probability our tranche experiences losses might be 60%, as opposed to approximately 0.13% under $P$. An unbiased estimator for expected tranche losses is given by

$$\frac{1}{n}\sum_{i=1}^{n}\Lambda_i d\left(L_i\right) \tag{4.21}$$

where $(\Lambda_i, L_i)$ are independent copies of the likelihood ratio and portfolio losses, simulated according to $\tilde{P}$ (note that $\Lambda_i$ is not necessarily independent of $L_i$). Simulating portfolio losses under $\tilde{P}$ has the effect of producing more information on the relevant portion of the loss distribution, while multiplying by $\Lambda$ weights each observation by the relative likelihood of actually observing such a loss trajectory in practice (i.e. under $P$). Note that (4.21) is unbiased since $E\left[d\left(L\right)\right] = \tilde{E}\left[\Lambda d\left(L\right)\right]$.

In order that such a scheme be practical, it is necessary that both $\Lambda$ and $L$ be easily simulated under $\tilde{P}$. In addition, in order that the scheme produces more efficient estimators it must be the case that

$$\widetilde{\mathrm{Var}}\left(\Lambda d\left(L\right)\right) < \mathrm{Var}\left(d\left(L\right)\right) \tag{4.22}$$

where $\widetilde{\mathrm{Var}}$ denotes variance under the measure $\tilde{P}$. Identification of measures $\tilde{P}$ with these two properties is by no means trivial.

In general there are no standard methods for selecting effective importance measures. Ideally we would like to develop an objective procedure for such an endeavour, which would prove effective across a wide variety of specifications of our framework. Asmussen et al. [7] consider the use of entropy-minimization as a tool for selecting effective importance distributions in the context of estimating tail probabilities of sums of heavy-tailed random variables, and report very positive results. Extending their approach to our problem, one would first define a

parametric family of candidate measures $\{P_\theta : \theta \in \Theta\}$, where $\Theta$ is some (possibly infinite-dimensional) parameter space and each $P_\theta$ is equivalent to $P$. Next, one selects the importance measure as that member which minimizes the cross-entropy, or Kullback-Leibler distance

$$h(\theta) = E^* \left[ \log \frac{dP^*}{dP_\theta} \right] \tag{4.23}$$

where the measure $P^*$ is defined by

$$P^*(A) = P(A \,|\, d(L) > 0)$$

In this way, one hopes to obtain an importance measure under which the distribution of portfolio losses "looks as close as possible" to its conditional distribution, where we condition upon our tranche experiencing losses. We now turn to the results of a preliminary investigation into the use of this paradigm in the context of a very simple specification of our basic framework.

### 4.3.1 Preliminary Results

The model used in our preliminary investigation is the homogeneous "pure time-change" model, where the credit quality of obligor $i$ is given by

$$X_t^i = x_0 + bVt + \sqrt{V} W_t^i$$

where $x_0 > 0$ and $b \in \mathbb{R}$ are constants and $V$ is a random variable. This is a model where relevant quantities can in fact be determined numerically, and we hope to use it to gain insights into the cross-entropy paradigm in more sophisticated models where Monte Carlo is the method of choice.

We consider $K$ tranches, labeled $k = 1, \ldots, K$, with attachment and detachment points $A_k, D_k$ and a common maturity of $T$ years. As our focus is on the situation where tranche losses are rare, we ignore the equity tranche, that is we assume $A_k > 0$ for each $k$. In addition we assume that tranches are indexed by seniority, so that $A_k < A_{k+1}$.

We assume the underlying portfolio is asymptotically large, and that the total notional is one dollar. Before proceeding we wish to remind the reader of the following fundamental facts with respect to the model and its use in pricing tranches of CDOs

- The proportion of defaults by time $t$ is

$$D_t = P\left(\tau_i \leq t \,|V\,\right) = \Phi\left(-\frac{x_0 + bVt}{\sqrt{Vt}}\right) + e^{-2x_0 b}\Phi\left(\frac{bVt - x_0}{\sqrt{Vt}}\right)$$

- The percentage loss by time $t$ is

$$L_t = (1 - R)D_t$$

  where $R$ is the deterministic (and constant across firms) recovery rate

- For tranche $k$, the cumulative loss experienced by time $t$ is

$$f_k\left(L_t\right) = \left(L_t - A_k\right)^+ - \left(L_t - D_k\right)^+$$

  while the outstanding tranche principal at time $t$ is

$$g_k\left(L_t\right) = \left(D_k - A_k\right) - f_k\left(L_t\right)$$

- The present value of the default leg is

$$d_k\left(V\right) = \sum_{i=1}^{N} e^{-rt_i}\left[f_k\left(L_{t_i}\right) - f_k\left(L_{t_{i-1}}\right)\right]$$

  where $N$ is the number of payment dates and $t_i = i\Delta$ with $\Delta = T/N$ and $T$ the maturity of the contract. Here $r$ is the constant and deterministic risk-free interest rate.

- Given a contracted spread $S_k$, the present value of the protection leg is given by $S_k p_k(V)$, where

$$p_k\left(V\right) = \Delta \sum_{i=1}^{N} e^{-rt_i} \frac{g_k\left(L_{t_i}\right) + g_k\left(L_{t_{i-1}}\right)}{2}$$

- The fair spread for tranche $k$ is given by

$$s_k = \frac{E\left[d_k(V)\right]}{E\left[p_k(V)\right]} \tag{4.24}$$

  Thus the fair spread simply equates the expected present values of the two cash flow "legs."

We assume that $V$ has the log-Laplace density (C.3), discussed in greater detail in Appendix C. Recall that this is a three-parameter family with parameters $(\alpha, \beta_1, \beta_2)$. Thus there are five parameters in the model, and using the following parameter choices

| $x_0$ | $b$ | $\alpha$ | $\beta_1$ | $\beta_2$ |
|---|---|---|---|---|
| 1.0058 | -1.6688 | -3.7099 | 0.1418 | 0.0569 |

we obtain the following relevant quantities for five-year tranches.

| Tranche | $E[d_k(V)]$ | $P(d_k(V) > 0)$ | Model Spread | Market Spread |
|---|---|---|---|---|
| 3-7% | .0016 | .1232 | 90.1 bp | 90 bp |
| 7-10% | .0003 | .0173 | 19.4 bp | 19 bp |
| 10-15% | .0001 | .0062 | 6.0 bp | 7 bp |
| 15-30% | .0001 | .0016 | 0.79 bp | 3.5 bp |

The model spreads are those reported by DiGraziano and Rogers [41] for five-year tranches (we exclude both the equity and super-senior tranches). All quantities have been computed numerically and are treated as exact.

In order to use our cross-entropy paradigm we must first select a parametric family of candidate densities for the systematic factor $V$. To this end let $g$ denote the actual density of $V$, that is the density (C.3) with parameters

$$(\alpha, \beta_1, \beta_2) = (-3.7009, 0.1418, 0.0569)$$

In addition for $\mu \in \mathbb{R}$ define $g_\mu$ as the density (C.3) with parameters

$$(\mu, 0.1418, 0.0569)$$

Thus our candidates remain in the log-Laplace family, with only the "location-type" parameter being adjusted.

For tranche $k$ we have that $d_k(V) > 0$ if and only if $V \geq v_k$, where $v_k$ solves

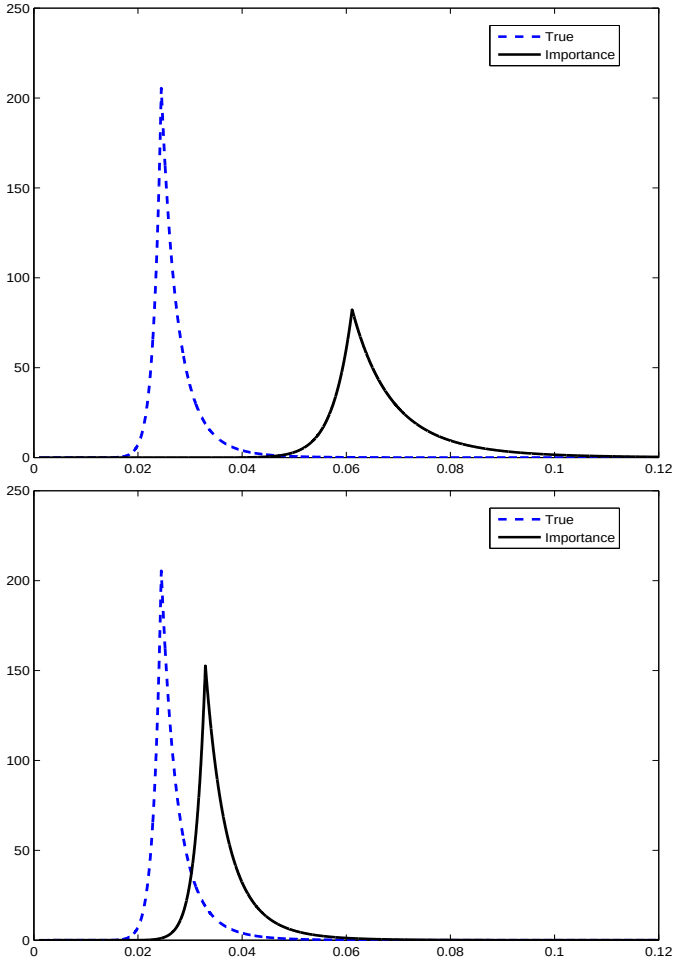
$$A_k = (1 - R)\left[\Phi\left(-\frac{x_0 + bv_k t}{\sqrt{v_k t}}\right) + e^{-2x_0 b}\Phi\left(\frac{bv_k t - x_0}{\sqrt{v_k t}}\right)\right] \tag{4.25}$$

Note that if $A_k \geq (1 - R)e^{-2x_0 b}$ then (4.25) has no solution and tranche $k$ never experiences losses. We call $v_k$ the "critical level" for tranche $k$. Letting $g^{(k)}$ denote the conditional density of $V$, given $V \geq v_k$, our goal is to minimize the Kullback-Leibler distance

$$e_k(\mu) = \int_{v_k}^{\infty} \log\left(\frac{g^{(k)}(v)}{g_\mu(v)}\right) g^{(k)}(v)dv \tag{4.26}$$

It is straightforward to derive a closed-form expression for (4.26), which is easily minimized numerically. We call that value of $\mu$ which minimizes (4.26) the "optimal importance" parameter for tranche $k$, and denote it by $\mu_k$. Moreover the density $g_{\mu_k}$ will be called the "optimal importance density" for tranche $k$.

Figure 4.6: Actual and Importance Densities




The top panel of Figure 4.6 plots the true density and the optimal importance density for the most senior tranche, namely the 15-30%. We see that we've both fattened the tails and shifted the density with this one parameter. The bottom panel of Figure 4.6 plots both densities when the 3-7% tranche is targeted. We see that there is a similar, but less pronounced effect. This figure indicates that we need to move the importance density "farther" from the true density as the seniority of the tranche under consideration increases.

Under each importance density, the associated tranche has a probability of experiencing losses of approximately 88%. Thus simulating $V$ according to the density $g_{\mu_k}$ will produce a sample in which tranche $k$ experiences losses in roughly 90% of simulated scenarios. Our paradigm has therefore proved successful according to our first criteria, namely that it provides more information on the relevant portion of the loss distribution. Our second, and more important criteria, is that the

paradigm enhances the efficiency of Monte Carlo estimators of expected tranche losses. That is, letting $\Lambda_k(v) = g(v)/g_{\mu_k}(v)$ denote the likelihood ratio between the actual and importance densities, we would like the standard deviation of the importance sampling estimator

$$\frac{1}{n}\sum_{i=1}^{n}\Lambda_k\left(V_i\right)d_k\left(V_i\right)$$

to be small relative to the standard deviation of the crude estimator

$$\frac{1}{n}\sum_{i=1}^{n}d_k\left(V_i\right)$$

Recall that in the former estimator the $V_i$ are simulated from the importance density $g_{\mu_k}$, while in the latter they are simulated from the actual density $g$. The following table illustrates the potential efficiency gains by reporting estimated values of

$$\sqrt{\frac{\mathrm{Var}\left(d_k\left(V\right)\right)}{\mathrm{Var}_k\left(\Lambda_k\left(V\right)d_k\left(V\right)\right)}}$$

where Var denotes variance under the actual density and $\mathrm{Var}_k$ denotes variance under the importance density $g_{\mu_k}$.

| | Tranche | | | |
|---|---|---|---|---|
| | 3-7% | 7-10% | 10-15% | 15 -30% |
| Gain | 4.3019 | 13.2825 | 16.6251 | 19.1957 |

The results here are quite encouraging. In terms of estimator standard deviations for the most senior tranche, simulating 1,000 observations under the optimal importance density is equivalent to simulating 20,000 observations under the actual density. In more sophisticated models where computing portfolio losses is a computational burden, such an improvement would pay massive dividends.

Suppose now that we are interested in the simultaneous estimation of expected losses for all tranches under consideration. Using a different set of simulated values for the systematic factor for each individual tranche would require four times as much computational effort as simulating one set, and using it to estimate the relevant quantity for all tranches. As such suppose that we choose our importance density by "targeting" tranche $k$, that is using the optimal importance density $g_{\mu_k}$. Further suppose we simulate values $V_i$ from this density and use them to estimate the expected losses on tranche $j$. That is we use the estimator

$$\frac{1}{n}\sum_{i=1}^{n}\Lambda_k\left(V_i\right)d_j\left(V_i\right)$$

Table 4.7: Efficiency Gains - Expected Tranche Loss

| | Targeted Tranche | | | |
|---|---|---|---|---|
| | 3-7% | 7-10% | 10-15% | 15 -30% |
| 3-7% | 4.3019 | 1.1549 | .4150 | .4330 |
| 7-10% | 2.7948 | 13.2825 | 7.5519 | 1.1459 |
| 10-15% | 2.1598 | 6.4476 | 16.6251 | 6.8861 |
| 15-30% | 1.3494 | 4.0609 | 7.1283 | 19.1957 |

where $V_i$ is simulated according to $g_{\mu_k}$. Table 4.7 illustrates the potential efficiency gains inherent in this endeavour. The $(j,k)$ element of the table reports estimated values for

$$\sqrt{\frac{\mathrm{Var}\left(d_j\left(V\right)\right)}{\mathrm{Var}_k\left(\Lambda_k\left(V\right)d_j\left(V\right)\right)}}$$

Inspection of the table reveals that there is no "globally optimal" tranche to target. In order to obtain maximum efficiency for estimating expected losses for tranche $k$, one must target that very tranche. In addition if the targeted tranche is "too far" from the tranche under consideration, importance sampling can actually lead to *less efficient* estimators. For example targeting the most senior tranche in order to estimate expected losses on the mezzanine leads to an importance sampling estimator which is nearly twice as variable as a crude estimator.

For completeness we also investigate the potential for efficiency gains when estimating the probability of tranche loss. The crude estimator would be

$$\frac{1}{n}\sum_{i=1}^{n} I\left(d_j\left(V_i\right)>0\right)$$

while the importance sampling estimator when tranche $k$ is targeted would be

$$\frac{1}{n}\sum_{i=1}^{n} \Lambda_k\left(V_i\right) I\left(d_j\left(V_i\right)>0\right)$$

The $(j,k)$ element of Table 4.8 reports estimated values for the ratio

$$\sqrt{\frac{\mathrm{Var}\left(I\left(d_j\left(V\right)>0\right)\right)}{\mathrm{Var}_k\left(\Lambda_k\left(V\right)I\left(d_j\left(V\right)>0\right)\right)}}$$

The results for estimating probability of tranche loss are broadly similar to those for estimating expected tranche losses, and we will not discuss them further.

Table 4.8: Efficiency Gains - Probability of Tranche Loss

| | Targeted Tranche | | | |
|---|---|---|---|---|
| | 3-7% | 7-10% | 10-15% | 15 -30% |
| 3-7% | 5.0796 | .2803 | .0395 | .7691 |
| 7-10% | 3.0697 | 14.6960 | 2.6154 | .4003 |
| 10-15% | 2.9979 | 9.3970 | 23.7802 | 2.4783 |
| 15-30% | 2.7906 | 7.7286 | 14.5967 | 48.6936 |

Recall that a crude Monte Carlo estimator of the fair spread $s_k$ would simulate $V$ according to its actual density $g$ and set

$$\hat{s}_k = \frac{\frac{1}{n}\sum_{i=1}^{n} d_k(V_i)}{\frac{1}{n}\sum_{i=1}^{n} p_k(V_i)} \tag{4.27}$$

In the introduction to this section we noted that, as the numerator here is the most troublesome component of CDO pricing, variance reduction techniques which improve estimation of this quantity might be expected to provide greater efficiency than the crude estimator. Unfortunately here we have found that this is in fact not the case. To this end we compared the standard deviation of (4.27) with that of the importance sampling estimator

$$\frac{\frac{1}{n}\sum_{i=1}^{n} \Lambda_k(V_i) d_k(V_i)}{\frac{1}{n}\sum_{i=1}^{n} \Lambda_k(V_i) p_k(V_i)} \tag{4.28}$$

We generated 1,000 realizations of the spread estimators (each estimate based on 10,000 realizations of $V$) and obtained the following results

| Estimated Standard Deviations of Spread Estimator | | | | |
|---|---|---|---|---|
| Method | Tranche | | | |
| | 0-3% | 3-7% | 7-10% | 10-15% |
| Crude | 3.2880 | 1.6809 | .9204 | .2755 |
| IS | 620.0723 | 348.1453 | 105.6633 | 13.6307 |

These somewhat surprising results can be traced to the behaviour of the likelihood ratios $\Lambda_k(v)$. Recalling Figure 4.6 we see that this ratio tends to be quite small when $v$ is large, and quite large when $v$ is small. Despite the fact that small values of $V$ are not likely under the importance density, a large sample *will* contain a few small values of the systematic factor. As $d_k(v)$ will be zero in these cases, abnormally large values of the likelihood ratio are offset by null tranche losses.

However this is *not* the case for the protection leg $p_k(v)$. In the rare event that the simulated value of $V$ is small, the associated likelihood ratio $\Lambda_k(V)$ is *not* offset by a null value of $p_k(V)$. Though likelihood ratio *is* bounded in all cases, the (sharp) upper bound is approximately ten million when the most senior tranche is targeted. Thus $\Lambda_k(V)$ will occasionally be extraordinarily large, and even one or two such values causes estimation of $E[p_k(V)]$ via

$$\frac{1}{n}\sum_{i=1}^{n}\Lambda_k(V_i)\,p_k(V_i)$$

to go horribly awry. Though the importance sampling estimator reduces variability in estimating the numerator, this benefit is offset by increased variability in the denominator. The net result is a massive increase in variability of the tranche spread estimator.

All is not lost, however, as we have found a way to use our importance sampling scheme to enhance the efficiency of spread estimators. We begin by noting that with $c_k = e^{\mu_k - \alpha}$, the distribution of $c_k V$ when $V$ has density $g$ is precisely $g_{\mu_k}$. Thus both $d_k(V)$ and $\Lambda_k(c_k V)\,d_k(c_k V)$ are unbiased estimators of $E[d_k(V)]$, when $V$ has density $g$. To see this note that

$$\begin{aligned} E[\Lambda_k(c_kV)\,d_k(c_kV)] &= E_k[\Lambda_k(V)\,d_k(V)] \\ &= E\left[\Lambda_k^{-1}(V)\,\Lambda_k(V)\,d_k(V)\right] \\ &= E[d_k(V)] \end{aligned}$$

This suggests the following estimator

$$\frac{\frac{1}{n}\sum_{i=1}^{n}\Lambda_k(c_kV_i)\,d_k(c_kV_i)}{\frac{1}{n}\sum_{i=1}^{n}p_k(V_i)}$$

where the $V_i$ are independent simulations of the systematic factor, drawn from the actual density $g$. The behaviour of the numerator of this estimator "mimics" the behaviour of the importance sampling estimator for $E[d_k(V)]$ which we saw perform so well earlier in the section. The following table reports the estimated efficiency gains from using this method.

| Estimated Efficiency Gains | | | |
|---|---|---|---|
| Tranche | | | |
| 3-7% | 7-10% | 10-15% | 15-30% |
| 3.9721 | 14.6391 | 22.9636 | 38.8251 |

We reach the encouraging conclusion that our cross-entropy importance sampling paradigm can significantly enhance the efficiency of Monte Carlo estimators for fair tranche spreads. As an alternative to using this "trick", we may consider the following more general approach. Let $\tilde{g}$ be any density with support $(0, \infty)$ and consider the importance sampling estimator

$$\frac{\frac{1}{n}\sum_{i=1}^{n} d_k(V_i)\,\tilde{\Lambda}(V_i)}{\frac{1}{n}\sum_{i=1}^{n} p_k(V_i)\,\tilde{\Lambda}(V_i)} \tag{4.29}$$

where $V_i$ is simulated from the density $\tilde{g}$ and $\tilde{\Lambda}(v) = g(v)/\tilde{g}(v)$. According to the development in Appendix D the variance of (4.29) is approximately minimized by setting

$$\tilde{g}_k(v) = c_k g(v) \left| \frac{d_k(v)}{E\left[d_k(V)\right]} - \frac{p_k(v)}{E\left[p_k(V)\right]} \right|$$

where

$$c_k = \frac{1}{E\left[\left|\frac{d_k(V)}{E[d_k(V)]} - \frac{p_k(V)}{E[p_k(V)]}\right|\right]}$$

The top panel of Figure 4.7 plots the density $\tilde{g}_k$ for the 15-30% tranche. We see that the variance-minimizing density has the distinct appearance of a mixture, the first component of which resembles the actual density $g$ and the second component of which assigns most of its mass to the critical region (i.e. the region where $d_k(V) > 0$) $[v_k, \infty)$. Motivated by this observation the bottom panel of Figure 4.7 plots the following crude approximation to the optimal density $\tilde{g}_k$ (for which we have no closed-form expression)

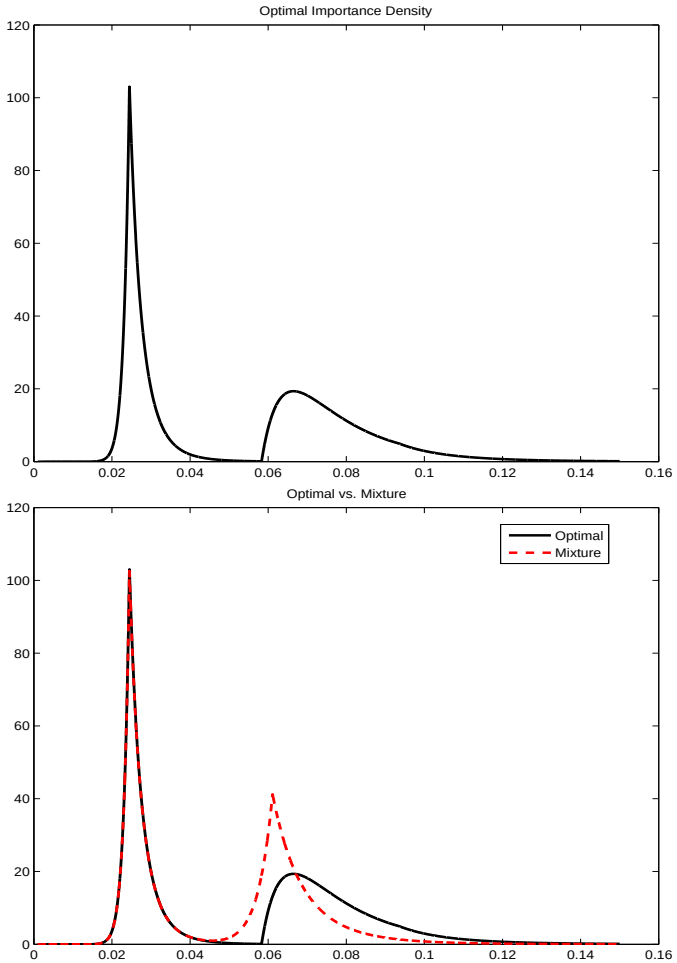
$$\hat{g}(v) = \frac{1}{2}\left(g(v) + g_{\mu_k}(v)\right)$$

where $g_{\mu_k}$ is the importance density given in the top panel of Figure 4.6. The advantage of this mixture approach is that the likelihood ratio is bounded above (by 2 in this case), eliminating the previously encountered problem of unstable likelihood ratios. Using this importance density we obtain an efficiency gain over a crude simulation of approximately 15. This preliminary exercise suggests that in order to minimize variance of the importance sampling estimator of fair tranche spreads, one use a mixture of the form

$$p \cdot g(v) + (1-p)\, g_1(v)$$

where $g$ is the actual density of $V$, and $g_1$ is a density which "targets" the event that our tranche experience losses. In using mixtures of this form one eliminates the problem of unstable likelihood ratios, while maintaining an emphasis on the

Figure 4.7: Optimal Importance Density



extreme tail of the tranche loss distribution. The optimal selection of the mixing parameter $p \in (0,1)$, as well as the rare-event density $g_1$, will be the subject of future research.

We also believe this technique, namely approximating the optimal importance density for tranche spread estimators via mixtures, should be applicable in more general models. In particular, for the linear model of Section 4.2.2, Figure 4.8 presents contour plots of the function

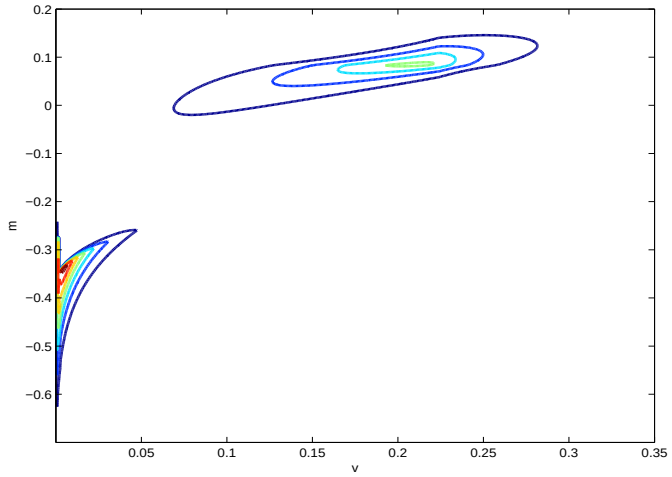
$$\tilde{f}(m,v) = f(m,v)\left|\frac{p_k(m,v)}{E[p_k(M,V)]} - \frac{d_k(m,v)}{E[d_k(M,V)]}\right|$$

where $f(m,v)$ is the calibrated joint density of $(M,V)$ (calibrated parameters are taken from the non-distressed data). These contours also resemble a mixture density, as there are two distinct "regions" here. The first set of contours closely resemble those of the actual density $f(m,v)$, while the second set of contours are concentrated in the region where both $M$ and $V$ are abnormally small. In Section 4.2.2 we saw that this is precisely the region where the super-senior tranche experiences losses. We have attempted two crude attempts at simulating from a density resembling that pictured in Figure 4.8, each of which simulates from a mixture of the form

$$p \cdot f(m,v) + (1-p) \cdot h(m,v)$$

where $p \in (0,1)$ and $h$ is a density designed to target super-senior tranche losses. Our first choice for $h$ was the uniform density on $[-.45,-.3] \times [0,.03]$, while our second choice was to simply re-center the bivariate density $f$ at the point $(m,v) = (-0.3357, 0.0049)$. In both cases, even after varying the parameter $p$, we were not able to obtain significant efficiency gains. However we believe this approach should still be fruitful, and plan to investigate more appropriate mixtures in the future.

Figure 4.8: Contour Plot - Optimal Importance Density



# Chapter 5

# Extensions and Future Work

In many ways this thesis culminated with Chapter 4. Our original motivation was a deeper understanding, and enhanced implementation, of the multivariate Black-Cox model. As we delved deeper into the subject we learned more about the similarity between this and the Gaussian copula, and began to realize that more appropriate models were required. This planted the seed[1] for the framework presented in Chapter 4. In turn, in an effort to implement this model appropriately, Chapter 3 was born.

In this concluding chapter we discuss future avenues for research with an emphasis on the material presented in Chapter 4. In order to implement the model we used the asymptotic approximation for portfolio losses. Indeed calibration was simply not feasible without using this approximation. We feel that a detailed analysis of this approximation and its effect on portfolio loss distributions and CDO spreads is in order.

We would also like to take a deeper look at the problem of inferring the paths of the systematic factors based on an observed trajectory of portfolio losses. In the dynamic model an observed trajectory is simply the first-passage distribution for a time-changed Brownian motion to a continuous barrier. The problem therefore becomes imputing, or reconstructing, the barrier and time change based on the first-passage distribution. Having done so one obtains the history of the systematic factors, and can then determine the conditional distribution of future losses. In the absence of a time change the first problem is well-studied in the literature - existence and uniqueness are available (see Chadam et al. [32]), as are computational methods for reconstructing the barrier based on a given first passage distribution (see [3] for

[1]The author wishes to express his utmost gratitude to Don L. McLeish for planting this seed.

example). Unfortunately the introduction of a time change appears to complicate matters significantly, as it becomes necessary to isolate the effects of both the barrier and time-change. Furthering this train of thought, we note that one never actually observes *asymptotic* losses. Our observable quantities are therefore a finite set of imperfectly observed first-passage probabilities, and we believe that results for empirical processes may be useful in this regard.

We believe there is also significant scope for enhancing the modeling process itself. The overwhelming conclusion of our calibration exercises was that, in order to correctly model "bad times," we had to completely remove idiosyncratic risk in this periods. We feel this warrants a re-interpretation of the "volatility" factor $V_t$ as something akin to a stochastic correlation factor, which weights the relative importance of systematic and idiosyncratic forces. To this end we might try modeling credit quality as

$$dX_t = \theta\left(M_t\right) M_t dt + \left(1 - \theta\left(M_t\right)\right) dW_t^i \tag{5.1}$$

where the function $\theta$ takes values in $[0,1]$ and explicitly weights the importance of each factor. Alternatively we might consider retaining $V_t$ as a volatility factor and trying something along the lines of

$$dX_t = \theta\left(M_t\right) M_t dt + \left(1 - \theta\left(M_t\right)\right) \sqrt{V_t} dW_t^i \tag{5.2}$$

thus allowing for both stochastic volatility and correlation. It would be interesting to see the effect of $V_t$ here when it is not required to play the stochastic correlation role. More generally we might model the idiosyncratic component as something other than a time-changed Brownian motion, for example

$$dX_t = \theta\left(M_t\right) M_t dt + \left(1 - \theta\left(M_t\right)\right) dY_t^i \tag{5.3}$$

where $Y^i$ is now a more general process. Note that this might significantly complicate evaluation of conditional default probabilities.

We are particularly pleased with the prospects for cross-entropy minimization as a tool for selecting importance measures in the rare event setting, and would like to extend the method to importance sampling for diffusion processes. This is an area which is very underdeveloped in the literature. To this end suppose that $X$ is a one-dimensional diffusion process

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

and suppose that we are interested in approximating an expectation of the form $E\left[\phi\left(X\right)\right]$, where $\phi : \mathcal{C}[0,T] \to \mathbb{R}$ is a path functional and the event $\{\phi\left(X\right) > 0\}$ is

rare. For example we may think of $\phi$ as an indicator of a rare event, the payoff of a deep out-of-the-money option, or, as in Section 4.3, as the loss on a senior CDO tranche. When $\phi$ is the indicator of a rare event, large deviations techniques might be employed (see Dembo and Zeitouni [38]), however this typically results in a very difficult variational problem.

In many situations this problem may be reduced to one involving standard Brownian motion as follows. Provided $\sigma$ is non-zero and has a continuous derivative we may define $Z_t = s(X_t)$, where

$$s(x) = \int_{x_0}^{x} \frac{1}{\sigma(z)} dz$$

where $x_0$ is a constant in the state space of $X$. Denoting the inverse of $s$ by $g$, Ito's Lemma shows

$$dZ_t = a(Z_t)\, dt + dW_t$$

where

$$a(z) = \frac{\mu(g(z))}{\sigma(g(z))} - \frac{1}{2}\sigma'(g(z))$$

Using the results of DiCesare and McLeish [40], provided

$$\exp\left(-\int_0^t a(Z_s)\, dW_s - \frac{1}{2}\int_0^t [a(Z_s)]^2\, ds\right)$$

is a martingale, it is possible to obtain unbiased estimators of $\phi(Z)$ by simulating $Z$ as a standard Brownian motion. Thus in many situations the general problem may be restated in terms of estimating expected values of path functionals of standard Brownian motion, to which we now turn.

Suppose now that $W_t$ is a standard Brownian motion over $[0, T]$. Recall that the first criteria for selecting an effective importance measures for estimating $E[\phi(W)]$ is that $W$ be easily simulated under this measure. As Brownian motion with deterministic drift is indeed easy to simulate, we define the candidate family as those defined by the Radon-Nikodym derivatives

$$\frac{dP_\theta}{dP} = \exp\left(\int_0^t \theta_s dW_s - \frac{1}{2}\int_0^t \theta_s^2 ds\right)$$

where $\theta \in \mathcal{L}^2[0, T]$. Note that we have

$$dW_t = \theta_t dt + dW_t^\theta$$

where $W_t^\theta = W_t - \int_0^t \theta_s ds$ is a standard Brownian motion under $P_\theta$. Our goal is to identify that $\theta$ which minimizes the Kullback-Leibler distance between $P_\theta$ and the

conditional measure $\tilde{P}$ defined by

$$\tilde{P}(A) = P\left(A \,|\phi\left(W\right) > 0\right)$$

Clearly $\tilde{P}$ is absolutely continuous with respect to $P$ (though the converse is not necessarily true), and the Radon-Nikodym derivative is easily seen to be

$$\frac{d\tilde{P}}{dP} = \frac{I\left(\phi\left(W\right) > 0\right)}{P\left(\phi\left(W\right) > 0\right)}$$

It is also clear that

$$\frac{d\tilde{P}}{dP} = \frac{1}{P\left(\phi\left(W\right) > 0\right)} \qquad \tilde{P} - \text{a.s.}$$

The quantity to be minimized is then

$$\begin{aligned}
\tilde{E}\left[\log \frac{d\tilde{P}}{dP_\theta}\right] &= \tilde{E}\left[\log \frac{d\tilde{P}}{dP} + \log \frac{dP}{dP_\theta}\right] \\
&= -\log\left(P\left(\phi\left(W\right) > 0\right)\right) + E\left[\frac{d\tilde{P}}{dP} \log \frac{dP}{dP_\theta}\right] \\
&= -\log\left(P\left(\phi\left(W\right) > 0\right)\right) + E\left[\log \frac{dP}{dP_\theta}\middle| \phi\left(W\right) > 0\right] \\
&= -\log\left(P\left(\phi\left(W\right) > 0\right)\right) - E\left[\int_0^t \theta_s dW_s \middle| \phi\left(W\right) > 0\right] + \frac{1}{2}\int_0^t \theta_s^2 ds
\end{aligned}$$

Proposition 5.0.1 identifies that function $\theta$ which minimizes this quantity.

**Proposition 5.0.1.** *Let $h_t = E\left[W_t \,|\phi\left(W\right) > 0\right]$. Provided $h$ is twice differentiable with square-integrable first derivative, cross-entropy is minimized by setting*

$$\theta_t = \frac{d}{dt} h_t =: \dot{h}_t$$

*Proof.* In order to minimize entropy we need only choose $\theta$ to minimize

$$-E\left[\int_0^t \theta_s dW_s \middle| \phi\left(W\right) > 0\right] + \frac{1}{2}\int_0^t \theta_s^2 ds \tag{5.4}$$

Since

$$\theta_t W_t = \int_0^t \theta_s dW_s + \int_0^t \dot{\theta}_s W_s ds$$

we obtain that

$$\begin{aligned}
-E\left[\int_0^t \theta_s dW_s \middle| \phi\left(W\right) > 0\right] &= \int_0^t \dot{\theta}_s h_s ds - \theta_t h_t \\
&= -\int_0^t \theta_s \dot{h}_s ds
\end{aligned}$$

Plugging into (5.4) we must choose the function $\theta$ to minimize

$$\frac{1}{2}\int_0^t \theta_s^2 ds - \int_0^t \theta_s \dot{h}_s ds = \frac{1}{2}\int_0^t \left[\theta_s - \dot{h}_s\right]^2 ds - \frac{1}{2}\int_0^t \dot{h}_s^2 ds$$

and clearly this is minimized by setting $\theta_t = \dot{h}_t$. □

This provides an objective criterion for choosing an importance measure, however the effectiveness of the result is not yet clear. The problem of optimal importance sampling for diffusion processes is very underdeveloped in the literature, however in a recent paper Guasoni and Robertson [66] consider optimal importance sampling as follows (their method is an extension of the method devised by Glasserman et al. [64] in the context of multivariate normal vectors). Letting $G(W) = \log(\phi(W))$ these authors base their optimality criteria based on "small-noise asymptotics," by defining

$$\alpha(\epsilon, \theta) = E\left[\exp\left(\epsilon^{-1}\left(2G\left(\sqrt{\epsilon}W\right) - \sqrt{\epsilon}\int_0^T \theta_t dW_t + \frac{1}{2}\int_0^T \theta_t^2 dt\right)\right)\right] \tag{5.5}$$

and setting

$$L(\theta) = \limsup_{\epsilon \to 0} \epsilon \log \alpha(\epsilon, \theta) \tag{5.6}$$

A candidate drift $\hat{\theta}$ is then deemed optimal if

$$\hat{\theta} = \min_{\theta} L(\theta) \tag{5.7}$$

Conditions under which an optimal drift exists are discussed in the paper, and when it exists the drift is expressed as the solution to a variational problem, which may be reduced to an Euler-Lagrange differential equation.

In order to investigate the relative performance of the cross-entropy method we apply it to the problem of pricing path-dependent options under the assumption the stock price follows

$$dS_t = S_t\left[rdt + \sigma dW_t\right]$$

The payoff to a geometric Asian option with strike price $K$ is given by

$$\left(\exp\left(\frac{1}{T}\int_0^T \log S_t dt\right) - K\right)^+$$

which can be re-written as a path functional of $W$ via

$$\phi(W) = \frac{K}{c}\left(\exp\left(a\int_0^T W_t dt\right) - c\right)^+$$

where $a = \sigma/T$ and $c = \frac{K}{S_0} e^{-T(r-\sigma^2/2)/2}$. The optimal drift according to the Guasoni/Robertson criteria is linear

$$\hat{\theta}_t = \hat{\alpha}\,(T-t)$$

where $\hat{\alpha}$ is the unique solution to

$$a\hat{\alpha}T^3 + 3\log\left(\frac{\hat{\alpha}-a}{c\hat{\alpha}}\right) = 0$$

In order to determine the entropy-minimizing drift we need only calculate

$$E\left[W_t \,\middle|\, \int_0^T W_t dt > \frac{\log c}{a}\right]$$

which is straightforward, and leads to a drift which is also linear

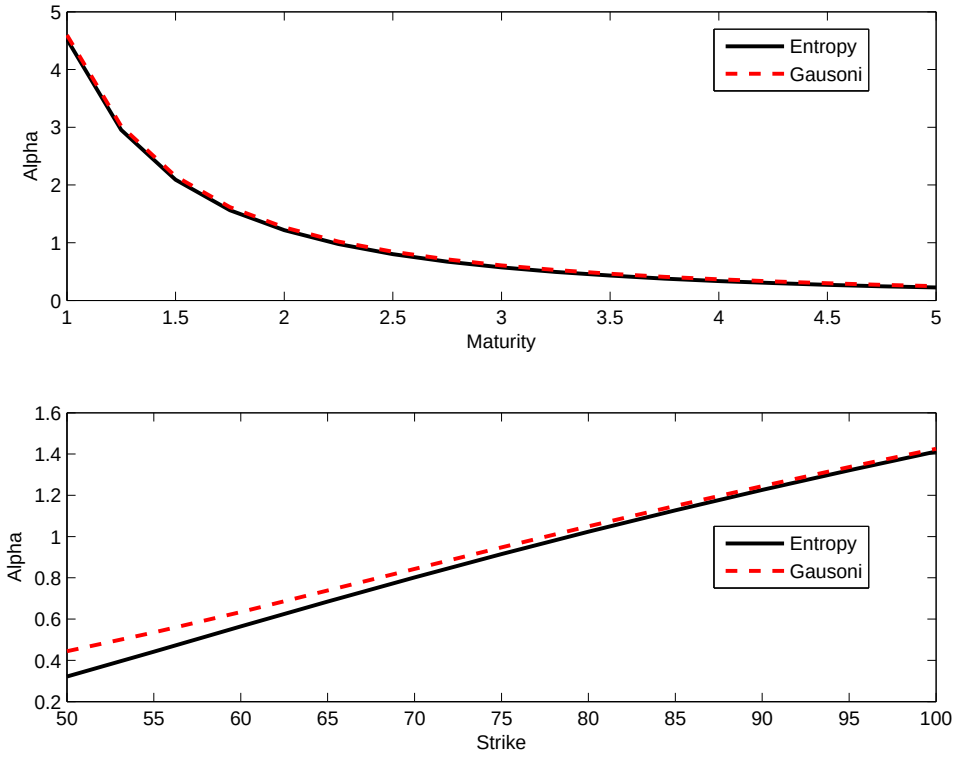
$$\tilde{\theta}_t = \tilde{\alpha}\,(T-t)$$

where

$$\tilde{\alpha} = \sqrt{\frac{3}{T^3}} \cdot \frac{\phi\left(-\tilde{K}\right)}{\Phi\left(-\tilde{K}\right)} \qquad \tilde{K} = \sqrt{\frac{3}{T^3}}\frac{\log c}{a}$$

Figure 5.1 plots the values of $\tilde{\alpha}$ and $\hat{\alpha}$ for various strike prices and maturities. In most cases these two parameters are indistinguishable. We are quite excited about the possibilities here, as this example illustrates that the entropy method performs quite well. Though an explicit determination of the optimal drift is not necessarily easy outside of special cases, the fact that it is expressed as an expectation leads us to believe that "preliminary simulations" may be useful in identifying drifts close to the optimal solution. In turn, this might provide a much more practical method than one involving difficult variational problems. As a final note we mention that applications of this method would not be limited to financial mathematics, indeed we believe it could be useful in many problems involving rare events of diffusion processes.

Figure 5.1: Guasoni's Parameter vs. Entropy Parameter



# APPENDICES

# Appendix A

# Liquid Credit Derivatives

Credit derivatives are traded in over-the-counter markets, and the specifics of each contract can vary widely between deals. Two of the most popular credit derivatives are credit default swaps (CDS) and CDS index tranches. These are among the most liquid due to their high degree of standardization.

For a good introduction to the world of credit derivatives, extending far beyond the basic products mentioned here, the reader is referred to Banks et al. [10]. Hull and White [73] provides an excellent introductions to single-name default swaps and their valuation. Finally, for an excellent description of CDS indices and index tranches the reader is referred to D'Amato and Gyntelberg [36]. Here we provide a brief discussion of default swaps and index tranches and present simple valuation formulae.

Roughly speaking a CDS is contract which provides insurance against losses suffered as a result of corporate defaults. There is an incredible variety among contracts which fall under this name, the simplest being a single-name CDS. The following description of such a contract is taken from D'Amato and Gyntelberg [36]

> A single-name CDS contract is an insurance contract covering the risk that a specified credit defaults. Following a defined credit event, the protection buyer receives a payment from the protection seller to compensate for credit losses. In return, the protection buyer pays a premium to the protection seller over the life of the contract.

The compensation payments and premium payments are typically referred to as the default leg and protection leg of the contract, respectively.

As an example the reference credit might be $1 million worth of General Motors bonds, and the contract might last for five years with quarterly payments. The protection payments are typically quoted as a "spread," which is agreed upon by the parties at the inception of the contract and stated as an annual simple interest rate. If the spread in our example were set at 6%, this would correspond to quarterly payments of $15,000. Should GM "survive" the five year life of the contract, the buyer would be required to make this payment at each of the 20 payment dates. Should GM default on its bonds at any time over the life of the contract, the seller would make a lump-sum payment to reimburse the buyer for the resulting losses. If the post-default value of GM's bonds were forty cents per dollar principal (corresponding to a recovery rate of 40%), then the seller would pay the buyer $600,000 and the buyer would not be obligated to make future premium payments. This reimbursement would be made at the time of default, and the buyer would need to make a final "accrual" payment at this time to account for the fact that she was "protected" between the most recent payment date and the time of default. If GM defaulted two months after the third payment date, the buyer would owe an additional (in addition to the three previous premium payments) amount of $10,000 at the time of default (one-sixth of a year at 6% simple interest).

Valuation of single-name CDS contracts is quite similar to valuation of futures contracts. The present value of the contract from the buyer's perspective is the difference between the present value of the default leg and the present value of the protection leg. The present value from the seller's perspective is simply the negative of the present value from the buyer's perspective. The contract is "fair" if the expected present values (under the risk-neutral measure) to both parties are equal, which occurs if and only if they are each equal to zero. The "fair spread" is that which equates the expected present values of each leg. Hull and White [73] show that this spread approximately eliminates arbitrage opportunities, and provide a "non-parametric" valuation method using prices of the reference credit and Treasury bonds. O'Kane and Turnbull [93] provide a good discussion of how CDS valuation is typically carried out in practice.

Ignoring the final accrual payment and assuming a constant risk-free interest rate $r$, we can express the present value of the protection leg in a single-name CDS on $1 notional as

$$\sum_{i=1}^{N} s\Delta e^{-rt_i} I\left(\tau \geq t_i\right)$$

where $s$ is the spread, $\Delta$ is the length of time between two payment dates, $t_i = i\Delta$

is the $i^{th}$ payment date and $\tau$ is the default time. Letting

$$\tau^* = \Delta \left\lfloor \frac{\tau}{\Delta} \right\rfloor$$

denote the payment date immediately preceding default (here $\lfloor x \rfloor$ is the integer part of $x$), we can write the present value of the final accrual payment as

$$s(\tau - \tau^*)e^{-r\tau} I\left(\tau \leq T\right)$$

The expected present value of the "protection leg" is therefore

$$sE\left[(\tau - \tau^*)e^{-r\tau} I\left(\tau \leq T\right)\right] + s\Delta \sum_{i=1}^{N} e^{-rt_i} P\left(\tau \geq t_i\right)$$

The present value of the "default leg" in the contract is simply

$$(1-R)e^{-r\tau} I\left(\tau \leq T\right) = (1-R)\sum_{i=1}^{N} e^{-r\tau} I\left(t_{i-1} < \tau \leq t_i\right)$$

If we assume that the recovery rate is constant the expected value of the default leg is simply

$$(1-R)E\left[e^{-r\tau} I\left(\tau \leq T\right)\right] = (1-R)\sum_{i=1}^{N} E\left[e^{-r\tau} I\left(t_{i-1} < \tau \leq t_i\right)\right]$$

Therefore the fair spread on the contract (which is independent of the notional) is given by

$$s = \frac{(1-R)E\left[e^{-r\tau} I\left(\tau \leq T\right)\right]}{E\left[(\tau - \tau^*)e^{-r\tau} I\left(\tau \leq T\right)\right] + \Delta \sum_{i=1}^{N} e^{-rt_i} P\left(\tau \geq t_i\right)} \tag{A.1}$$

If $\tau$ is a discrete random variable taking on values in $\{\Delta, 2\Delta, \ldots, \}$ with $p_i = P\left(\tau = i\Delta\right)$, then (A.1) reduces considerably to

$$s = \frac{(1-R)\sum_{i=1}^{N} p_i e^{-rt_i}}{\Delta \sum_{i=1}^{N} q_i e^{-rt_i}} \tag{A.2}$$

where $q_i = P\left(\tau \geq i\Delta\right)$. In the more general case where $\tau$ could be continuous (or discrete with a different support), O'Kane and Turnbull [93] show that equation (A.2) with $p_i = P\left(t_{i-1} < \tau \leq t_i\right)$ often provides a fast and accurate approximation to (A.1).

There are many variations of the single-name CDS. These include multi-name versions, often referred to as basket default swaps, an example of which would be a $k^{th}$-to-default swap. In such a contract reimbursement from the seller is only

made after the $k^{th}$ default in a portfolio of $n$ reference credits take place. No reimbursement is made for the first $(k-1)$ defaults. A general expression for the fair spread in this situation, assuming constant recovery rates, would be

$$s = \frac{E\left[(1-R_{(k)})D_{(k)}e^{-r\tau_{(k)}}I\left(\tau_{(k)} \leq T\right)\right]}{E\left[(\tau_{(k)} - \tau^*_{(k)})e^{-r\tau_{(k)}}I\left(\tau_{(k)} \leq T\right)\right] + D\Delta\sum_{i=1}^{N} e^{-rt_i}P\left(\tau_{(k)} \geq t_i\right)}$$

where $D_i$ and $R_i$ are the notional principal and recovery rate on the $i^{th}$ credit, $D = \sum_{i=1}^{N} D_i$ is the total notional. Here $\tau_{(k)}$ denotes the $k^{th}$ order statistic of the default times, and $D_{(k)}$ and $R_{(k)}$ denote the notional principal and recovery rate of the $k^{th}$ firm to default (*not* the $k^{th}$ largest values of the $D_i$ and $R_i$). This simplifies significantly if principals and recovery rates are constant across the reference credits, although one would still require the marginal distribution of the order statistic. The significant difference between basket default swaps and single-name swaps is that, in general, the distribution of the $k^{th}$ order statistic will depend on the entire joint distribution of the default times. We note in passing that Brasch [22] shows how a $k^{th}$-to-default basket swap can be exactly replicated with first-to-default swaps on various sub-baskets. This means that pricing a $k^{th}$-to-default swap is possible if one only knows the distribution of the first order statistic for each sub-portfolio. Unfortunately the number of first-to-default swaps needed can grow quite rapidly with $k$ and $n$, although in the homogeneous case only $k-1$ "sub-valuations" are required.

At the present time the market for single-name default swaps is quite liquid, indeed D'Amato and Gyntelberg [36] comment that these contracts are more liquid than most corporate bonds. As these markets have developed several indices have been introduced to track spreads on large portfolios of investment grade corporations. In particular Dow Jones has introduced the CDX and iTraxx series, which report average spreads on portfolios of 125 investment grade US and European corporations, respectively. Indices are available for various maturities, as well as various sub-portfolios such as financials or industrials. The composition of the underlying portfolios is updated semi-annually and is determined by dealers.

There is also a liquid market for standardized basket default swaps on these indices, referred to as CDS index contracts. With these products the seller of protection reimburses the buyer for each default in the index, and each time a default occurs the buyer's premium payments are reduced (the buyer pays the same spread, but on a reduced principal). Each firm is equally weighted, in the sense that the notional principals are equal. See D'Amato and Gyntelberg [36] for a good description of the indices and index contracts.

We conclude this section with a brief discussion of collateralized debt obligations (CDO) and CDS index tranches. The reader is referred to Banks et al. [10] for a comprehensive treatment. In a CDO an underlying portfolio of assets is defined, along with several "tranches." Each tranche is defined by two percentages: an attachment point ($K_A$) and a detachment point ($K_D$), for example 3% and 7%. The buyer of protection in this tranche is not reimbursed for any defaults until cumulative losses on the portfolio have reached 3% of the original notional principal. Once cumulative losses have hit this attachment point, the buyer is reimbursed for all defaults until cumulative losses have reached 7%. Once cumulative losses exceed the detachment point the buyer on this tranche is not reimbursed for any future losses. In return for this protection the buyer makes regular payments to the seller. As with default swaps, these payments are typically referred to as the "tranche spread," and quoted as a simple annual rate. In our example the buyer is initially covered for an amount equal 4% of the total notional, and until the attachment point is hit the spread is assessed on this amount. As defaults occur which are absorbed by this tranche, the amount to which the spread is applied decreases by amounts equal to each loss.

The key concepts in CDO valuation are "tranche losses" and "tranche principal." With $NP$ denoting the original notional principal on the portfolio, consider a tranche with attachment point $K_A$ and detachment point $K_D$. As the buyer of protection on this tranche is insured up to a maximum of $NP\,(K_D - K_A)$, we call this the "initial tranche" principal. Now with $L(t)$ denoting the total cumulative losses on the portfolio up to time $t$, the "cumulative tranche loss" is defined as

$$\begin{array}{rll} 0 & \text{if} & L(t) < NP \cdot K_A \\ L(t) - NP \cdot K_A & \text{if} & NP \cdot K_A < L(t) < NP \cdot K_D \\ NP\,(K_D - K_A) & \text{if} & L(t) > NP \cdot K_D \end{array}$$

The original amount to which the tranche spread is applied is $NP(K_D - K_A)$(this the amount for which the buyer of protection is insured). Each time the tranche absorbs a loss, this amount is reduced by the amount of that loss. Therefore at any time $t$, the amount for which the buyer is insured is given by

$$\begin{array}{rll} NP\,(K_D - K_A) & \text{if} & L(t) < NP \cdot K_A \\ K_D - L(t) & \text{if} & NP \cdot K_A < L(t) < NP \cdot K_D \\ 0 & \text{if} & L(t) > NP \cdot K_D \end{array}$$

We call this the "tranche notional" and note the outstanding tranche principal at any time is simply the initial amount for which the buyer is covered, less the value of any reimbursements paid out.

The fair spread for a CDO tranche is determined in the same way as a default swap, namely it is the spread which equates the present values of the default and protection legs. As with default swaps, accrual payments are due at each default time in addition to the regular quarterly payments. For example, suppose that as of the third payment date the tranche principal is $4 million, and that the tranche absorbs a $500,000 loss one month after this date. The buyer was insured for this $500,000 for a one-month period and if the tranche spread is 12% a payment of $5,000 would be due at the time of the loss. Should no more losses be absorbed by the tranche by the next payment date, an amount of $105,000 would be due on the next payment date. This is because the buyer was insured for the $3.5 million of "undefaulted tranche principal" for the full three-month period. Including accrual payments in an expression for the present value of protection payments can entail significant complications as one must not only keep track of the tranche principal at payment dates, but default times as well.

To simplify evaluation of the fair spread it is quite common to ignore accrual payments and assume the following simplified payment structure (see [27]). At each payment date $t_i$, the spread is assessed on the simple average of the outstanding tranche principal over $[t_{i-1}, t_i]$. Thus we would simply taken the beginning and end-of-period tranche principals and apply the spread to their average value. In addition it is also common to assume default reimbursements are made only on the payment dates, with no accounting for the timing of defaults within an interval.

While the specific features of CDOs can vary widely from deal to deal, there is a very liquid market for "CDS index tranches." These standardized contracts are essentially CDOs on an equally-weighted portfolio of default swaps. The names in the portfolio are the companies in the CDX and iTraxx indices. For the CDX investment grade index the available tranches are 0-3%, 3-7%, 7-10%, 10-15% and 15-30%. For the iTraxx investment grade index the available tranches are 0-3%, 3-6%, 6-9%, 9-12% and 12-22%.

## A.1 Valuation of CDO tranches

Consider a portfolio of $N$ names, with nominal exposures $E_i$ and recovery rates $R_i$. The cumulative percentage loss on the portfolio up to time $t$ is given by

$$L_N(t) = \sum_{i=1}^{N} w_i(1 - R_i) I\left(\tau_i \leq t\right)$$

where $w_i = E_i / \sum_{i=1}^{N} E_i$. When valuing tranches of CDOs written on this portfolio, we may assume without loss of generality that the total notional is one dollar, that is $\sum_{i=1}^{N} E_i = 1$. For a tranche with attachment and detachment points $K_A$ and $K_D$, respectively, the cumulative loss experienced by the tranche up to time $t$ is given by

$$f\left(L_N(t), K_A, K_D\right) = \max\left[0, \min\left(L_N(t), K_D\right) - K_A\right] \tag{A.3}$$

Similarly, the outstanding tranche principal as of time $t$ is given by

$$g\left(L_N(t), K_A, K_D\right) = \max\left[0, K_D - \max\left(L_N(t), K_A\right)\right] \tag{A.4}$$

and we note that

$$g\left(L_N(t), K_A, K_D\right) = \left(K_D - K_A\right) - f\left(L_N(t), K_A, K_D\right)$$

Suppose that our tranche is initiated today and has a maturity of $T$ years with $N$ equally-spaced payment dates $t_i = i\Delta$. Here $\Delta = T/N$ is the length of time between payment dates. As mentioned in the previous section, a common assumption when valuing tranches is to assume that all losses are reimbursed at the payment dates. With this assumption the present value of the default leg (i.e. the present value of cash flows received by the buyer of protection on this tranche) is given by

$$PV\left(\text{Default}\right) = \sum_{i=1}^{N} e^{-\int_0^{t_i} r(s)ds} \left[f\left(L_N(t_i)\right) - f\left(L_N(t_{i-1})\right)\right] \tag{A.5}$$

where we have suppressed the dependence of $f$ on $K_A$ and $K_D$, and where $r(t)$ is the risk-free interest rate. A further simplifying assumption commonly made when valuing tranches is to ignore accruals in protection payments. Assuming the spread on the tranche is $S$ and the protection payment due at $t_i$ is based on the (simple) average outstanding principal over $[t_{i-1}, t_i]$ we obtain that the present value of the protection leg (i.e. the present value of cash flows received by the seller of protection on this tranche) is given by

$$PV\left(\text{Prot}\right) = \sum_{i=1}^{N} e^{-\int_0^{t_i} r(s)ds} \left(S\Delta\right) \frac{g\left(L_N(t_i)\right) + g\left(L_N(t_{i-1})\right)}{2} \tag{A.6}$$

The fair spread is then taken to be that value of $S$ which equates the expected present values (under the risk-neutral pricing measure) of the default and protection legs. The formula for the fair spread, say $\hat{S}$, is then given by

$$\hat{S} = \frac{\sum_{i=1}^{N} E\left[e^{-\int_0^{t_i} r(s)ds} \left[f\left(L_N(t_i)\right) - f\left(L_N(t_{i-1})\right)\right]\right]}{\Delta \sum_{i=1}^{N} E\left[e^{-\int_0^{t_i} r(s)ds} \left[g\left(L_N(t_i)\right) + g\left(L_N(t_{i-1})\right)\right]\right] / 2}$$

As a final note, we point out that for the investment grade CDX and iTraxx tranches, there is a slightly different payment structure for the equity tranches (i.e. tranches with an attachment point of 0%). For these tranches the running spread is typically fixed at 500 basis points, and parties to the contract negotiate an upfront fee which is paid at inception. This fee is expressed as a percentage (quoted either in basis or percentage points) of the initial tranche notional, in which case the present value of the protection leg becomes

$$PV\left(\text{Prot}\right) = SK_D + \sum_{i=1}^{N} e^{-\int_0^{t_i} r(s)ds} \left(.05\Delta\right) \frac{g\left(L_N(t_i)\right) + g\left(L_N(t_{i-1})\right)}{2} \tag{A.7}$$

and the fair spread is given by

$$\hat{S} = \frac{E\left[PVD\right] - .05E\left[PVP\right]}{K_D}$$

where

$$\begin{aligned} PVD &= \sum_{i=1}^{N} e^{-\int_0^{t_i} r(s)ds} \left[f\left(L_N(t_i)\right) - f\left(L_N(t_{i-1})\right)\right] \\ PVP &= \Delta \sum_{i=1}^{N} E\left[e^{-\int_0^{t_i} r(s)ds} \left[g\left(L_N(t_i)\right) + g\left(L_N(t_{i-1})\right)\right]\right] /2 \end{aligned}$$

In order to evaluate tranche spreads, one need only compute the expectations $E\left[f\left(L_N(t_i)\right)\right]$ for each payment date. When using Monte Carlo one may simulate trajectories for portfolio losses, compute the present value of each leg for each trajectory, and approximate the fair spread by the ratio of their averages.

# Appendix B

# Appendix to Chapter 2

## B.1 Transformation Details

In this section we carry out the details of the transformation in Section 2.1. As mentioned in that section, the overall transformation is best understood as the composition of the following three individual transformations

1. The linear transformation

$$U(t) = \begin{bmatrix} \frac{1}{\sqrt{1-\rho^2}} & -\frac{\rho}{\sqrt{1-\rho^2}} \\ 0 & 1 \end{bmatrix} W(t)$$

   It is easy to verify that $U(t)$ is a standard planar Brownian motion started at the origin. In addition, the vertical line at $w_1 = a_1$ becomes the line

$$u_2 = \frac{a_1}{\rho} - \frac{\sqrt{1-\rho^2}}{\rho} u_1$$

   and the horizontal line at $w_2 = a_2$ becomes the horizontal line at $u_2 = a_2$. Note that these lines intersect at the point

$$(u_1, u_2) = \left( \frac{a_1 - \rho a_2}{\sqrt{1-\rho^2}}, a_2 \right)$$

   Note also that if $\rho = 0$ this step is not carried out.

2. The translation

$$V(t) = U(t) - \begin{bmatrix} \frac{a_1 - \rho a_2}{\sqrt{1-\rho^2}} \\ a_2 \end{bmatrix}$$

$V(t)$ is a standard planar Brownian motion started at

$$\left(\frac{\rho a_2 - a_1}{\sqrt{1-\rho^2}}, -a_2\right)$$

The line

$$u_2 = \frac{a_1}{\rho} - \frac{\sqrt{1-\rho^2}}{\rho} u_1$$

becomes the line

$$v_2 = -\frac{\sqrt{1-\rho^2}}{\rho} v_1$$

while the horizontal line $u_2 = a_2$ becomes the horizontal axis $v_2 = 0$. It is worthwhile to note that the polar co-ordinates of the starting point are given by

$$\begin{aligned} r_0^{'} &= \sqrt{\tfrac{a_1^2+a_2^2-2\rho a_1 a_2}{1-\rho^2}} \\ \theta_0^{'} &= \begin{cases} \tan^{-1}\left(\tfrac{a_2\sqrt{1-\rho^2}}{a_1-\rho a_2}\right) & a_1 < \rho a_2 \\ \pi + \tan^{-1}\left(\tfrac{a_2\sqrt{1-\rho^2}}{a_1-\rho a_2}\right) & a_1 \geq \rho a_2 \end{cases} \end{aligned}$$

3. The rotation by the angle $\pi$

$$Z(t) = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} V(t)$$

Due to the rotational invariance of Brownian motion, $Z(t)$ is a standard planar Brownian motion. The line

$$v_2 = -\frac{\sqrt{1-\rho^2}}{\rho} v_1$$

becomes the (same) line

$$z_2 = -\frac{\sqrt{1-\rho^2}}{\rho} z_1$$

## B.2 Correcting to Iyengar's Formula for $P(\tau > t)$.

In this section we will need the following results concerning Bessel functions, which are available in Abramowitz and Stegun [2]

$$2I_\nu'(z) = I_{\nu-1}(z) + I_{\nu+1}(z) \tag{B.1}$$

$$\int_0^\infty e^{-\beta t^2} I_\nu(\alpha t) dt = \frac{1}{2}\sqrt{\frac{\pi}{\beta}} \exp\left(\frac{\alpha^2}{8\beta}\right) I_{\nu/2}\left(\frac{\alpha^2}{8\beta}\right) \tag{B.2}$$

We now derive Equation (2.3) from Equation (2.2). Equation (2.2) can be written

$$P^{z_0}\left(\tau > t,\ Z(t) \in dz\right) = \frac{2}{t\alpha} e^{-r_0^2/2t} \sum_{n=0}^{\infty} \sin\frac{n\pi\theta_0}{\alpha} g_n(r,\theta) dr d\theta$$

where, for given values of $t, \alpha, r_0$ we define

$$g_n(r,\theta) = r e^{-r^2/2t} \sin\frac{n\pi\theta}{\alpha} I_{n\pi/\alpha}\left(\frac{rr_0}{t}\right)$$

The distribution of $\tau$ is obtained by integrating (2.2) over $r$ and $\theta$ (recall that $z = (r\cos\theta, r\sin\theta)$)

$$\begin{aligned} P^{z_0}\left(\tau > t\right) &= \int_0^{\infty}\int_0^{\alpha} P^{z_0}\left(\tau > t,\ Z(t) \in dz\right) d\theta dr \\ &= \frac{2}{t\alpha} e^{-r_0^2/2t} \sum_{n=0}^{\infty} \sin\frac{n\pi\theta_0}{\alpha} \int_0^{\infty}\int_0^{\alpha} g_n(r,\theta) d\theta dr \end{aligned}$$

Since

$$\int_0^{\alpha} \sin\frac{n\pi\theta}{\alpha} d\theta = \begin{cases} 0 & n \text{ even} \\ \frac{2\alpha}{n\pi} & n \text{ odd} \end{cases}$$

we obtain

$$P^{z_0}\left(\tau > t\right) = \frac{4}{\pi t} e^{-r_0^2/2t} \sum_{n \text{ odd}}^{\infty} \frac{1}{n} \sin\frac{n\pi\theta_0}{\alpha} \int_0^{\infty} r e^{-r^2/2t} I_{n\pi/\alpha}\left(\frac{rr_0}{t}\right) dr$$

Using integration by parts and identities (B.1) and (B.2) we find that for $\nu \neq 0$

$$\int_0^{\infty} r e^{-r^2/2t} I_{\nu}\left(\frac{rr_0}{t}\right) dr = \frac{r_0\sqrt{2\pi t}}{4} \exp\left(\frac{r_0^2}{4t}\right)\left[I_{(\nu+1)/2}\left(\frac{r_0^2}{4t}\right) + I_{(\nu-1)/2}\left(\frac{r_0^2}{4t}\right)\right]$$

Therefore

$$P^{z_0}\left(\tau > t\right) = \frac{2r_0}{\sqrt{2\pi t}} e^{-r_0^2/4t} \sum_{n \text{ odd}}^{\infty} \frac{1}{n} \sin\left(\frac{n\pi\theta_0}{\alpha}\right)\left[I_{(\nu_n-1)/2}(r_0^2/4t) + I_{(\nu_n+1)/2}(r_0^2/4t)\right]$$

where $\nu_n = n\pi/\alpha$.

## B.3 Inverse Transform for Exit Location

In this section we derive inverses for the functions

$$\begin{aligned} F(r) &= 1 + \frac{1}{\pi - 2\theta_0}\left[\tan^{-1}\left(\frac{r - x_0}{y_0}\right) - \tan^{-1}\left(\frac{r + x_0}{y_0}\right)\right] \\ G(r) &= 1 + \frac{1}{2\theta_0}\left[\tan^{-1}\left(\frac{r - y_0}{x_0}\right) - \tan^{-1}\left(\frac{r + y_0}{x_0}\right)\right] \end{aligned}$$

where $0 < \theta_0 < \pi/2$ and $r_0 > 0$ are the polar co-ordinates of the point $(x_0, y_0)$. Note that this implies $x_0, y_0 > 0$.

For a given $b > 0$ it is straightforward to verify that

$$\tan^{-1}(a) - \tan^{-1}(b) = \begin{cases} \tan^{-1}\left(\frac{a-b}{1+ab}\right) - \pi & a < -\frac{1}{b} \\ -\frac{\pi}{2} & a = -\frac{1}{b} \\ \tan^{-1}\left(\frac{a-b}{1+ab}\right) & a > -\frac{1}{b} \end{cases}$$

Therefore the equation $u = F(r)$ (which, for a given value of $u$ has a unique solution in $r$) can be re-arranged as follows

$$(u-1)(\pi - 2\theta_0) = \begin{cases} \tan^{-1}\left(\frac{-\sin(2\theta_0)}{(r/r_0)^2 - \cos(2\theta_0)}\right) - \pi & r^2 < r_0^2 \cos 2\theta_0 \\ \tan^{-1}\left(\frac{-\sin(2\theta_0)}{(r/r_0)^2 - \cos(2\theta_0)}\right) & r^2 < r_0^2 \cos 2\theta_0 \end{cases}$$

so that

$$\tan\left((u-1)(\pi - 2\theta_0)\right) = \frac{-\sin 2\theta_0}{(r/r_0)^2 - \cos 2\theta_0}$$

Re-arranging we obtain

$$F^{-1}(u) = r_0 \sqrt{\cos(2\theta_0) - \frac{\sin(2\theta_0)}{\tan\left((\pi - 2\theta_0)(u-1)\right)}}$$

An analogous argument can be used to show that

$$G^{-1}(u) = r_0 \sqrt{-\cos(2\theta_0) - \frac{\sin(2\theta_0)}{\tan\left((2\theta_0)(u-1)\right)}}$$

## B.4 Derivation of the Joint Density of $(\tau_1, \tau_2)$

To begin note that for $s < t$ we have

$$\begin{aligned} P^{z_0}\left(\tau_1 \in ds,\ \tau_2 \in dt\right) &= P^{z_0}\left(\tau_1 \in ds,\ \tau_2 \in dt\right) \\ &= P^{z_0}\left(\tau \in ds,\ \tau' \in dt,\ \Theta(\tau) = \alpha\right) \\ &= P^{z_0}\left(\tau \in ds,\ (\tau' - \tau) \in d(t-s), \Theta(\tau) = \alpha\right) \end{aligned}$$

And so the probability in question can be obtained by integrating (with respect to $r$)

$$P^{z_0}\left(\tau \in ds,\ (\tau' - \tau) \in d(t-s), \Theta(\tau) = \alpha,\ R(\tau) \in dr\right)$$

This integrand can be re-written as the product

$$P^{z_0}\left((\tau' - \tau) \in d(t-s) | R(\tau) \in dr\right) P^{z_0}\left(\tau \in ds,\ \Theta(\tau) = \alpha,\ R(\tau) \in dr\right)$$

In Section 2.1.3 it was shown that the first term here is the inverse Gaussian density

$$P^{z_0}\left(\tau' - \tau \in d(t-s) | R(\tau) \in dr\right) = \frac{r \sin\alpha}{\sqrt{2\pi}(t-s)^{3/2}} \exp\left(-\frac{r^2 \sin^2\alpha}{2(t-s)}\right)$$

while in Section 2.1.2 we saw that the second term is given by

$$P^{z_0}\left(\tau \in ds,\ R(\tau) \in dr,\ \Theta(\tau) = \alpha\right) = \frac{\pi e^{-(r^2+r_0^2)/2s}}{\alpha^2 sr} \sum_{n=1}^{\infty} g_n\left(r_0, \alpha - \theta_0, \alpha, r, s\right)\ dr\ ds$$

where $g_n$ is given by (2.10). Multiplying these two terms and re-arranging we obtain

$$\frac{\pi \sin\alpha}{\sqrt{2\pi}\alpha^2 s(t-s)^{3/2}} e^{-r_0^2/2s} \sum_{n=1}^{\infty} n \sin(n\pi(\alpha - \theta_0)/\alpha) e^{-\beta r^2} I_{n\pi/\alpha}(rr_0/s) \tag{B.3}$$

where

$$\beta = \frac{t - s\cos^2\alpha}{2s(t-s)}$$

Using identity (B.2) we find (after a significant amount of simplification)

$$\begin{aligned}\int_0^{\infty} e^{-\beta r^2} I_{n\pi/\alpha}(rr_0/s) dr &= \sqrt{\frac{\pi}{2}} \sqrt{\frac{s(t-s)}{t - s\cos^2\alpha}} \\ &\quad \times \exp\left(\frac{r_0^2}{2s} \frac{(t-s)}{(t-s) + (t - s\cos 2\alpha)}\right) \\ &\quad \times I_{n\pi/2\alpha}\left(\frac{r_0^2}{2s} \frac{(t-s)}{(t-s) + (t - s\cos 2\alpha)}\right)\end{aligned}$$

Integrating (B.3) and inserting this expression we get, after a tremendous amount of simplification, that for $s < t$ the probability $P^{z_0}\left(\tau_1 \in ds,\ \tau_2 \in dt\right)$ is given by

$$\begin{aligned}&\frac{\pi \sin\alpha}{2\alpha^2 \sqrt{s}(t-s)\sqrt{t - s\cos^2\alpha}} \exp\left(-\frac{r_0^2}{2s} \frac{t - s\cos 2\alpha}{(t-s) + (t - s\cos 2\alpha)}\right) \\ &\times \sum_{n=1}^{\infty} n \sin\left(\frac{n\pi(\alpha - \theta_0)}{\alpha}\right) I_{n\pi/2\alpha}\left(\frac{r_0^2}{2s} \frac{t-s}{(t-s) + (t - s\cos 2\alpha)}\right)\end{aligned}$$

Now, to see that this explodes as $s \nearrow t$ we first note that

$$\frac{1}{t-s} I_{n\pi/2\alpha}\left(\frac{r_0^2}{2s} \frac{t-s}{(t-s) + (t - s\cos 2\alpha)}\right) \sim k_n \left[(t-s) + (t - s\cos 2\alpha)\right]^{-n\pi/2\alpha} (t-s)^{\frac{n\pi}{2\alpha} - 1}$$

where $k_n$ is a constant depending on $n$. Thus bringing the term $1/(t-s)$ inside the series we see that for $\frac{\pi}{2} < \alpha < \pi$, the leading term explodes as $s \nearrow t$ while the rest of the terms tend to zero.

To derive the joint density for $s > t$ we must integrate the product of the terms

$$\begin{aligned} P^{z_0}\left(\tau \in dt,\ R(\tau) \in dr,\ \Theta(\tau) = 0\right) &= \frac{\pi e^{-(r^2+r_0^2)/2t}}{\alpha^2 t r} \sum_{n=1}^{\infty} g\left(r_0, \theta_0, \alpha, r, s\right) \\ P^{z_0}\left(\tau' - \tau \in d(s-t) | R(\tau) \in dr\right) &= \frac{r \sin\alpha}{\sqrt{2\pi}(s-t)^{3/2}} \exp\left(-\frac{r^2 \sin^2\alpha}{2(s-t)}\right) \end{aligned}$$

Following the same steps we obtain that for $s > t$ the probability

$$P^{z_0}\left(\tau_1 \in ds,\ \tau_2 \in dt\right)$$

is given by

$$\begin{aligned} &\frac{\pi \sin\alpha}{2\alpha^2 \sqrt{t}(s-t)\sqrt{s - t\cos^2\alpha}} \exp\left(-\frac{r_0^2}{2t} \frac{s - t\cos 2\alpha}{(s-t) + (s - t\cos 2\alpha)}\right) \\ &\times \sum_{n=1}^{\infty} n \sin\left(\frac{n\pi\theta_0}{\alpha}\right) I_{n\pi/2\alpha}\left(\frac{r_0^2}{2t} \frac{s-t}{(s-t) + (s - t\cos 2\alpha)}\right) \end{aligned}$$

These are the equations presented in the main body of the thesis.

# Appendix C

# The Laplace and log-Laplace Distributions

In this appendix we discuss the Laplace distribution, which is a three-parameter family with several pleasing features, including heavy tails, infinite divisibility and asymmetry. We also discuss the related log-Laplace, which also has pleasing features such as power tails at both zero and infinity. In addition we discuss the construction of ergodic diffusion processes possessing the Laplace and log-Laplace as their invariant densities.

### The Laplace Distribution

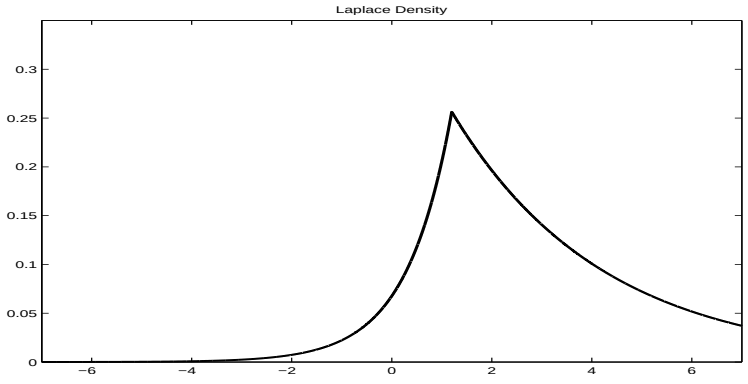
A random variable $X$ is said to have a *Laplace distribution* if it has a probability density function of the form

$$f(x) = \begin{cases} [\beta_1 + \beta_2]^{-1} \, e^{(x-\alpha)/\beta_2} & -\infty < x \leq \alpha \\ \\ [\beta_1 + \beta_2]^{-1} \, e^{(\alpha-x)/\beta_1} & \alpha \leq x < \infty \end{cases} \tag{C.1}$$

This is a three-parameter family with a location parameter $\alpha \in \mathbb{R}$ and scale-type parameters $\beta_1, \beta_2 > 0$. Figure C.1 plots a typical member of this family.

The Laplace density obtains as the density of the random variable $X = \alpha + \beta_1 Z_1 - \beta_2 Z_2$, where $Z_1$ and $Z_2$ are independent unit-mean exponential variates. Infinite divisibility is immediate from this representation, as is the fact that the mean of (C.1) is $\alpha + \beta_1 - \beta_2$. Several studies have found the Laplace to be quite capable of describing financial quantities such as log-changes in interest rates (see

Figure C.1: A Laplace Density




[82]). A more compact parametrization of (C.1) is given by

$$f(x) = \frac{\sigma^2 - \kappa^2}{2\sigma} \exp\left(\kappa(x-\alpha) - \sigma\left|x-\alpha\right|\right)$$

where

$$\sigma = \frac{1}{2}\frac{\beta_1 + \beta_2}{\beta_1\beta_2} \qquad \kappa = \frac{1}{2}\frac{\beta_1 - \beta_2}{\beta_1\beta_2}$$

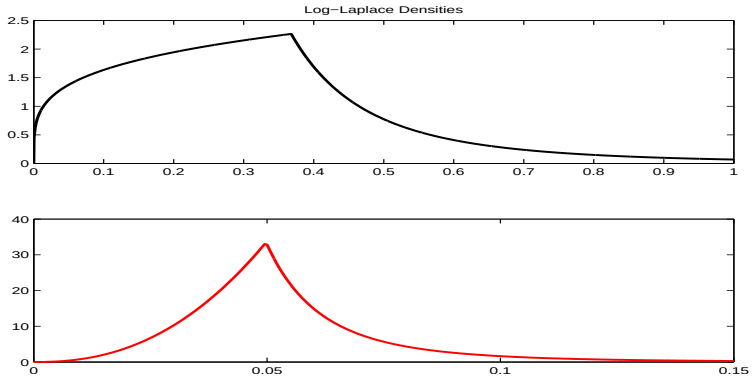
A comprehensive treatment of this and related distributions is available in Kotz et al. [82].

The moment generating function for (C.1) is given by

$$m(u) = E\left[\exp(uX)\right] = \frac{e^{u\alpha}}{(1-u\beta_1)(1+u\beta_2)} \qquad -\frac{1}{\beta_2} < u < \frac{1}{\beta_1} \tag{C.2}$$

which will be important when discussing moments of the log-Laplace distribution.

Figure C.2: Log-Laplace Densities




### The log-Laplace Distribution

If $X$ has Laplace density (C.1), the distribution of $Y = e^X$ is called the log-Laplace and is easily seen to have density

$$g(y) = \begin{cases} [\beta_1 + \beta_2]^{-1} \, e^{-\alpha/\beta_2} y^{(1-\beta_2)/\beta_2} & 0 < y \leq e^{\alpha} \\ \\ [\beta_1 + \beta_2]^{-1} \, e^{\alpha/\beta_1} y^{-(1+\beta_1)/\beta_1} & y \geq e^{\alpha} \end{cases} \tag{C.3}$$

Figure C.2 illustrates the wide variety of shapes that are possible within this family.

Interesting features of the log-Laplace distribution are that it has power tails at both zero and infinity, in particular it has heavier right-hand tails than the gamma distribution. In addition this family is clearly invariant to scaling and exponentiation.

Existence of the moments of (C.3) is governed solely by the parameter $\beta_1$. Recalling (C.2) we see that for integer $k$, $E\left[Y^k\right] < \infty$ if and only if $k < \beta_1^{-1}$. In particular (C.3) has finite mean provided $\beta_1 < 1$ and finite variance when $\beta_1 < 1/2$.

When it exists the mean of (C.3) is given by

$$\frac{e^{\alpha}}{(1-\beta_1)(1+\beta_2)}$$

which obtains easily using (C.2). As a final note we point out that (C.3) is bounded provided $\beta_2 \leq 1$.

### Constructing Ergodic Laplace Diffusion Processes

Our goal in this section is to construct an ergodic diffusion process $X_t$ possessing (C.1) as its invariant density. Our main tool will be the results of the Bibby et al. [17], which we outline here.

Suppose that $f$ is a probability density which is continuous, bounded, strictly positive on the interval $(l, u)$ and zero outside $(l, u)$. Here $-\infty \leq l < u \leq \infty$ and we require $f$ to have a finite mean $\mu$. For a given $\theta > 0$ define the function

$$\sigma^2(x) := \frac{2\theta \int_l^x (\mu - z)\, f(z) dz}{f(x)} \tag{C.4}$$

The domain of $\sigma$ is taken as $(l, u)$. Bibby et al. [17] consider the stochastic differential equation

$$dX_t = \theta\,(\mu - X_t)\, dt + \sigma\,(X_t)\, dW_t \tag{C.5}$$

and show that (C.5) has a unique weak solution with invariant density $f$. Clearly, if $X_0 \sim f$ then $X_t$ is stationary as well.

The Laplace density (C.1) satisfies the conditions required for this result, and it is straightforward to verify that in this case $\mu = \alpha + \beta_1 - \beta_2$ and

$$\sigma^2(x) = \begin{cases} 2\theta\beta_2\,[\beta_1 + \alpha - x] & x \leq \alpha \\ 2\theta\beta_1\,[\beta_2 + x - \alpha] & x > \alpha \end{cases} \tag{C.6}$$

When simulating large numbers of trajectories for such a process using an Euler-type scheme one would need, at each time step, to identify which values of $X_{t_i}$ are above and below $\alpha$. The time required for such a computation, though quite simple, can significantly slow down the simulation process. In order to eliminate the need to identify those sample paths for which $X_{t_i} > \alpha$, say, it is useful to re-express (C.6) as follows

$$\sigma^2(x) = \frac{2\theta}{\sigma^2 - \kappa^2}\left[1 + \sigma\,|x - \alpha| + \kappa\,(x - \alpha)\right]$$

where

$$\sigma = \frac{1}{2}\frac{\beta_1 + \beta_2}{\beta_1\beta_2} \qquad \kappa = \frac{1}{2}\frac{\beta_1 - \beta_2}{\beta_1\beta_2}$$

**Constructing Ergodic log-Laplace Diffusion Processes**

For the Laplace distribution one can construct an ergodic diffusion with (C.1) as its invariant density for all values of the parameters. The same cannot be said of the log-Laplace density (C.3). In order that a log-Laplace density have finite mean we require $\beta_1 < 1$, and in order that such a density be bounded we require $\beta_2 < 1$.

Using the results from Bibby et al. [17] discussed in the previous section, and with $f$ given by (C.3) with $\max(\beta_1, \beta_2) < 1$ we have that for $\theta > 0$ the process

$$dX_t = \theta\left(\mu - X_t\right) dt + \sigma\left(X_t\right) dW_t$$

has $f$ as its invariant density, where $\mu = e^{\alpha}/\left(1-\beta_1\right)\left(1+\beta_2\right)$ and

$$\sigma^2(x) = \begin{cases} 2\theta\beta_2 x\left[\mu - \frac{x}{1+\beta_2}\right] & 0 \leq x \leq e^{\alpha} \\ 2\theta\beta_1 x\left[\frac{x}{1-\beta_1} - \mu\right] & x \geq e^{\alpha} \end{cases} \tag{C.7}$$

This can be expressed in the more compact form

$$\sigma^2(x) = 2\theta x\left[\sigma\left(x - e^{\alpha}\right) + \kappa\left|x - e^{\alpha}\right| + \mu\beta_1\beta_2\right]$$

where we have re-defined $\mu$ and $\sigma$ via

$$\begin{aligned} \sigma &= \frac{\beta_1 + 2\beta_1\beta_2 - \beta_2}{2\left(1-\beta_1\right)\left(1+\beta_2\right)} \\ \kappa &= \frac{\beta_1 + \beta_2}{2\left(1-\beta_1\right)\left(1+\beta_2\right)} \end{aligned}$$

# Appendix D

# Optimal Importance Sampling for Ratios

In this appendix we discuss optimal importance sampling for ratio estimators. More specifically, suppose that $\mathbf{X}$ is a $d$-dimensional random vector with density function $f$ and consider the problem of estimating, via simulation, the ratio

$$\frac{E\left[h_1\left(\mathbf{X}\right)\right]}{E\left[h_2\left(\mathbf{X}\right)\right]} \tag{D.1}$$

where $h_1, h_2$ are real-valued functions defined on $\mathbb{R}^d$ such that $h_i\left(\mathbf{X}\right)$ has non-zero first moment and finite second moment. A crude estimation procedure would simulate independent copies of $\mathbf{X}$, say $\mathbf{X}_i$, from the density $f$ and approximate (D.1) via

$$\frac{\frac{1}{n}\sum_{i=1}^{n} h_1\left(\mathbf{X}_i\right)}{\frac{1}{n}\sum_{i=1}^{n} h_2\left(\mathbf{X}_i\right)} = \frac{\sum_{i=1}^{n} h_1\left(\mathbf{X}_i\right)}{\sum_{i=1}^{n} h_2\left(\mathbf{X}_i\right)}$$

In order to enhance the efficiency of such a procedure one might consider an importance sampling scheme which simulates $\mathbf{X}_i$ from an alternative density $g$ and estimate (D.1) via

$$\frac{\sum_{i=1}^{n} h_1\left(\mathbf{X}_i\right)\frac{f(\mathbf{X}_i)}{g(\mathbf{X}_i)}}{\sum_{i=1}^{n} h_2\left(\mathbf{X}_i\right)\frac{f(\mathbf{X}_i)}{g(\mathbf{X}_i)}} \tag{D.2}$$

An optimal importance density $\hat{g}$ will minimize the variance of (D.2), that is

$$\hat{g} = \arg\min \mathrm{Var}_g \left( \frac{\sum_{i=1}^{n} h_1\left(\mathbf{X}_i\right)\frac{f(\mathbf{X}_i)}{g(\mathbf{X}_i)}}{\sum_{i=1}^{n} h_2\left(\mathbf{X}_i\right)\frac{f(\mathbf{X}_i)}{g(\mathbf{X}_i)}} \right)$$

where $\mathrm{Var}_g$ denotes variance when $\mathbf{X}_i$ has density $g$ and the minimum is taken over all densities $g$ which have the same support as $f$.

An approximation to $\hat{g}$, say $\tilde{g}$, may be obtained as follows. To begin, it can be shown (indeed we will demonstrate this shortly) that for sufficiently large $n$ we have

$$\mathrm{Var}_g\left(\frac{\sum_{i=1}^n h_1(\mathbf{X}_i)\frac{f(\mathbf{X}_i)}{g(\mathbf{X}_i)}}{\sum_{i=1}^n h_2(\mathbf{X}_i)\frac{f(\mathbf{X}_i)}{g(\mathbf{X}_i)}}\right) \approx \frac{1}{n}\left(\frac{\mu_1}{\mu_2}\right)^2 \mathrm{Var}_g\left(\frac{f(\mathbf{X})}{g(\mathbf{X})}h(\mathbf{X})\right) \tag{D.3}$$

where

$$h(\mathbf{x}) = \frac{h_1(\mathbf{x})}{\mu_1} - \frac{h_2(\mathbf{x})}{\mu_2} \tag{D.4}$$

and $\mu_i = E_f[h_i(\mathbf{X})]$. Thus a reasonable approximation to $\hat{g}$ is given by

$$\tilde{g} = \arg\min \mathrm{Var}_g\left(\frac{f(\mathbf{X})}{g(\mathbf{X})}h(\mathbf{X})\right)$$

Now, since

$$E_g\left[\frac{f(\mathbf{X})}{g(\mathbf{X})}h(\mathbf{X})\right] = E_f[h(\mathbf{X})] = 0$$

we may use Jensen's inequality to obtain

$$\begin{aligned}
\mathrm{Var}_g\left(\frac{f(\mathbf{X})}{g(\mathbf{X})}h(\mathbf{X})\right) &= E_g\left[\left(\frac{f(\mathbf{X})}{g(\mathbf{X})}h(\mathbf{X})\right)^2\right] \\
&= E_g\left[\left(\frac{f(\mathbf{X})}{g(\mathbf{X})}|h(\mathbf{X})|\right)^2\right] \\
&\geq \left(E_g\left[\frac{f(\mathbf{X})}{g(\mathbf{X})}|h(\mathbf{X})|\right]\right)^2 \\
&= \left(E_f[|h(\mathbf{X})|]\right)^2
\end{aligned}$$

and since this lower bound is obtained by the following density, we see that $\tilde{g}$ is given by

$$\tilde{g}(\mathbf{x}) = \frac{f(\mathbf{x})|h(\mathbf{x})|}{E_f[|h(\mathbf{X})|]} \tag{D.5}$$

The crucial step in this development was the approximation (D.3). To see the validity of this approximation we note that if $X$ and $Y$ are random variables with finite second moments and $E[Y] \neq 0$, then (see Crámer [35]) the limiting distribution, as $n \to \infty$, of

$$\sqrt{n}\left(\frac{\sum_{i=1}^n X_i}{\sum_{i=1}^n Y_i} - \frac{\mu_X}{\mu_Y}\right)$$

is normal with mean zero and variance

$$\left(\frac{1}{\mu_Y}\right)^2\left[\sigma_X^2 - 2\frac{\mu_X}{\mu_Y}\sigma_X\sigma_Y\rho + \left(\frac{\mu_X}{\mu_Y}\sigma_Y\right)^2\right] \tag{D.6}$$

Here $(X_i, Y_i)$ are independent copies of $(X, Y)$, $\mu_X$ and $\sigma_X$ denote the mean and variance of $X$, $\mu_Y$ and $\sigma_Y$ denote the mean and variance of $Y$, and $\rho$ is the correlation between $X$ and $Y$. Noting that (D.6) is equal to

$$\left(\frac{\mu_X}{\mu_Y}\right)^2 \text{Var}\left(\frac{X}{\mu_X} - \frac{Y}{\mu_Y}\right)$$

it follows that for sufficiently large $n$ we have

$$\text{Var}\left(\frac{\sum_{i=1}^n X_i}{\sum_{i=1}^n Y_i}\right) \approx \frac{1}{n}\left(\frac{\mu_X}{\mu_Y}\right)^2 \text{Var}\left(\frac{X}{\mu_X} - \frac{Y}{\mu_Y}\right)$$

# References


[1] Aarts, E.H.L. and P.J.M. van Laarhoven. *Simulated Annealing: Theory and Applications*. D. Reidel, 1987.

[2] Abramowitz, M. and I.A. Stegun, editors. *Handbook of Mathematical Functions*. U.S. Department of Commerce, 1967.

[3] Abundo, M. Limit at zero of the first-passage time density and the inverse problem for one-dimensional diffusions. *Stochastic Analysis and Applications*, 24:1119–1145, 2006.

[4] Andersen, J., S. Basu, and J. Sidenius. All your hedges in one basket. *Risk*, November 2003.

[5] Andersen, L. Portfolio losses in factor models: Term structure and intertemporal loss dependence. Bank of America, 2006.

[6] Andersen, L. and J. Sidenius. Extensions to the Gaussian copula: Random recovery and random factor loadings. *Journal of Credit Risk*, 1:29–70, 2004.

[7] Asmussen, S., D.P. Kroese, and R.Y. Rubinstein. Heavy tails, importance sampling and cross-entropy. *Stochastic Models*, 21:57–76, 2005.

[8] Avellaneda, M. and J. Zhu. Modeling the distance-to-default process of a firm. Working Paper, 2001.

[9] Bañuelos, R. and R. Smits. Brownian motion in cones. *Probability Theory and Related Fields*, 108:299–319, 1997.

[10] Banks, E., M. Glantz, and P. Siegel. *Credit Derivatives*. McGraw-Hill, 2007.

[11] Baxter, M. Lévy process dynamic modeling of single-name credits and CDO tranches. Nomura Fixed Income Quant Group, 2006.

[12] Baxter, M. Lévy simple structural models. Working Paper, Nomura Fixed Income Quant Group, 2006.

[13] Beekman, J.A. and C. Park. Stochastic barriers for the Wiener process. *Journal of Applied Probability*, 20:338–348, 1983.

[14] Bennani, N. The forward loss model: A dynamic term structure approach for the pricing of portfolio credit derivatives. Royal Bank of Scotland, 2005.

[15] Bennani, N. A note on Markov functional loss models. Royal Bank of Scotland, 2006.

[16] Bhansali, V. and M. Wise. Correlated random walks and the joint survival probability. Working Paper, 2004.

[17] Bibby, B.M., I.M. Skovgaard, and M. Sørensen. Diffusion-type models with given marginal distribution and autocorrelation function. *Bernoulli*, 11:191–220, 2005.

[18] Bielecki, T.R. and M. Rutkowski. *Credit Risk: Modeling, Valuation and Hedging*. Springer, 2002.

[19] Billingsley, P. *Convergence of Probability Measures*. Wiley, 2nd edition, 1999.

[20] Black, F. and J.C. Cox. Valuing corporate securities: Some effects of bond indenture provisions. *Journal of Finance*, 31:351–367, 1976.

[21] Borovkov, K. and A. Novikov. Explicit bounds for approximation rates of boundary crossing probabilities for the Wiener process. *Journal of Applied Probability*, 42:82–92, 2005.

[22] Brasch, H.-J. Exact replication of kth-to-default swaps with first-to-default swaps. Working Paper, 2006.

[23] Brigo, D. and M. Morini. CDS calibration with tractable structural models. Working Paper, 2005.

[24] Brigo, D. and M. Morini. CDS calibration with tractable structural models under uncertain credit quality. Working Paper, 2005.

[25] Brigo, D., A. Pallavicini, and R. Torresetti. Implied correlation in CDO tranches: A paradigm to be handled with care. Working Paper, 2006.

[26] Brigo, D., A. Pallavicini, and R. Torresetti. Calibration of CDO tranches with the dynamical generalized Poisson loss model. Working Paper, 2007.

[27] Brigo, D., A. Pallavicini, and R. Torresetti. Implied expected tranched loss surface from CDO data. Working Paper, 2007.

[28] Buckholtz, P. and M. Wasan. Characterization of certain first passage processes. *Sankhya*, 38:326–339, 1976.

[29] Buckholtz, P. and M. Wasan. First passage probabilities of a two dimensional Brownian motion in an anisotropic medium. *Sankhya*, 41:198–206, 1979.

[30] Buffet, E. Credit risk: The structural approach revisited. In *Proceedings of the Third Seminar on Stochastic Analysis, Random Fields and Applications, Ascona 1999*. Birkhäuser, 2000.

[31] Buonocore, A., A.G. Nobile, and L.M. Ricciardi. A new integral equation for the evaluation of first-passage-time probability densities. *Advances in Applied Probability*, 19:784–800, 1987.

[32] Chadam, J., X. Chen, L. Cheng, and D. Saunders. Analysis of an inverse first passage problem from risk management. *SIAM Journal on Mathematical Analysis*, 38:845–873, 2006.

[33] Collin-Dufresne, P. and R.S. Goldstein. Do credit spreads reflect stationary leverage ratios? *Journal of Finance*, 56:1929–1957, 2001.

[34] Corana, A., M. Marchesi, C. Martini, and S. Ridella. Minimizing multimodal functions of continuous variables with the "simulated annealing" algorithm. *ACM Transactions on Mathematical Software*, 13:262–280, 1987.

[35] Crámer, H. *Mathematical Methods of Statistics*. Princeton University Press, 1946.

[36] D'Amato, J. and J. Gyntelberg. CDS index tranches and the pricing of credit risk correlations. BIS Quarterly Review, March 2005.

[37] Daniels, H.E. Sequential tests constructed from images. *Annals of Statistics*, 10:394–400, 1982.

[38] Dembo, A. and O. Zeitouni. *Large Deviations Techniques and Applications*. Springer, 2nd edition, 1998.

[39] Di Nardo, E., A.G. Nobile, E. Pirozzi, and L.M. Ricciardi. A computational approach to first-passage-time problems for Gauss-Markov processes. *Advances in Applied Probability*, 33:453–482, 2001.

[40] DiCesare, G. and D.L. McLeish. Importance sampling and imputation for diffusion models. Working Paper, University of Waterloo, 2006.

[41] DiGraziano, G. and L.C.G. Rogers. A dynamic approach to the modeling of correlation credit derivatives using Markov chains. Working Paper, 2006.

[42] Ding, X., K. Giesecke, and P. Tomecek. Time-changed birth processes and multi-name credit. Working Paper, 2006.

[43] Duffie, D., A. Eckner, G. Horel, and L. Saita. Frailty correlated default. Working Paper, 2006.

[44] Duffie, D. and D. Lando. Term structures of credit spreads with incomplete accounting information. *Econometrica*, 69:633–664, 2001.

[45] Duffie, J.D. and K.J. Singleton. Modeling term structures of defaultable bonds. *Review of Financial Studies*, 12:687–720, 1999.

[46] Durbin, J. Boundary-crossing probabilities for the Brownian motion and Poisson processes and techniques for computing the power of the Kolmogorov-Smirnov test. *Journal of Applied Probability*, 8:431–453, 1971.

[47] Elliott, R., L. Chan, and T.K. Siu. Option pricing and Esscher transform under regime switching. *Annals of Finance*, 1:423–432, 2005.

[48] Embrechts, E., A. McNeil, and R. Frey. *Quantitative Risk Management*. Princeton University Press, 2005.

[49] Ericsson, J., J. Reneby, and H. Wang. Can structural models price default risk? Evidence from bond and credit derivative markets. Working Paper, 2006.

[50] Errais, E., K. Giesecke, and L. Goldberg. Pricing credit from the top down with affine point processes. Working Paper, 2007.

[51] Fabozzi, F., S. Rachev, and D. Wang. Pricing of credit default index swap tranches with one-factor heavy-tailed copula models. Working Paper, 2006.

[52] Fabozzi, F.J., S.T. Rachev, and D. Wang. Pricing tranches of a CDO and a CDS index: Recent advances and future research. Working Paper, 2006.

[53] Feller, W. *Introduction to Probability Theory and its Applications, Volume II.* John Wiley and Sons, 2nd edition, 1971.

[54] Ferrarese, C. A comparative analysis of correlation skew modeling techniques for CDO index tranches. Master's thesis, King's College London, 2006.

[55] Finger, C. Issues in the pricing of synthetic CDOs. RiskMetrics Group Working Paper 04-01, 2004.

[56] Fortet, R. Les fonctions aléatoires du type de Markoff associées à certaines équations linéaires aux derivées partielles du type parabolique. *Journal de Mathématiques Pures et Appliquées*, 22:177–243, 1943.

[57] Fouque, J.-P., R. Sircar, and K. Sølna. Stochastic volatility effects on defaultable bonds. *Applied Mathematical Finance*, 13:215–244, 2006.

[58] Fouque, J.-P., B.C. Wignall, and X. Zhou. Modeling correlated defaults: First passage models under stochastic volatility. *Journal of Computational Finance*, 11:43–78, 2007.

[59] Frey, R. and A. McNeil. Dependent defaults in models of portfolio credit risk. Working Paper, 2003.

[60] Frishling, V., N. Kordzakhia, and A. Novikov. Approximations of boundary crossing probabilities for a Brownian motion. *Journal of Applied Probability*, 36:1019–1030, 1999.

[61] Genest, C., B. Remillard, and J.-F. Quessy. Goodness-of-fit procedures for copula models based on the probability integral transformation. *Scandinavian Journal of Statistics*, 33:337–366, 2006.

[62] Giesecke, K. Default and information. *Journal of Economic Dynamics and Control*, 30:2281–2303, 2006.

[63] Giesecke, K. and L. Goldberg. Sequential defaults and incomplete information. *Journal of Risk*, 7, 2004.

[64] Glasserman, P., P. Heidelberger, and P. Shahabuddin. Asymptotically optimal importance sampling and stratification for pricing path-dependent options. *Mathematical Finance*, 9:117–152, 1999.

[65] Gregory, J. and J.-P. Laurent. Basket default swaps, CDOs and factor copulas. *Journal of Risk*, 7, 2005.

[66] Guasoni, P. and S. Robertson. Optimal importance sampling with explicit formulas in continuous time. *Finance and Stochastics*, 12:1–19, 2008.

[67] Guo, X., R. Jarrow, and Y. Zeng. Information reduction in credit risk models. Working Paper, 2005.

[68] Hager, S. and R. Schobel. A note on the correlation smile. Working Paper, 2006.

[69] Huh, J. and A. Kolkiewicz. Efficient computation of multivariate barrier crossing probability and its applications in credit risk models. Working Paper, 2006.

[70] Hui, C.-H., C.-F. Lo, and S.-W. Tsang. Pricing corporate bonds with dynamic default barriers. *Journal of Risk*, 5:17–37, 2003.

[71] Hui, C.-H., C.F. Lo, and M.X. Huang. Are corporates' target leverage ratios time-dependent? *International Review of Financial Analysis*, 15:220–236, 2006.

[72] Hull, J., M. Predescu, and A. White. The valuation of correlation-dependent credit derivatives using a structural model. Working Paper, 2005.

[73] Hull, J. and A. White. Valuing credit default swaps I: No counterparty default risk. *Journal of Derivatives*, Fall 2000.

[74] Hull, J. and A. White. Valuation of a CDO and an n-th to default CDS without Monte Carlo simulation. *Journal of Derivatives*, 12:8–23, 2004.

[75] Hurd, T.R. Credit risk modeling using time-changed Brownian motion. Working Paper, 2007.

[76] Iyengar, S. Hitting lines with two-dimensional Brownian motion. *SIAM Journal on Applied Mathematics*, 45:983–989, 1985.

[77] Jarrow, R., P. Protter, and A. Deniz Sezer. Information reduction via level crossings in a credit risk model. *Finance and Stochastics*, 11:195–212, 2007.

[78] Jobert, A. and L.C.G. Rogers. Option pricing with Markov-modulated dynamics. *SIAM Journal on Control and Optimization*, 44:2063–2078, 2006.

[79] Kalemanova, A., B. Schmid, and R. Werner. The normal inverse Gaussian distribution for synthetic CDO pricing. Working Paper.

[80] Karatzas, I. and S. Shreve. *Brownian Motion and Stochastic Calculus*. Springer, 2nd edition, 1991.

[81] Kendall, D. On finite and infinite sequences of exchangeable events. *Studia Scientiarum Mathematicarum Hungarica*, 2:319–327, 1967.

[82] Kotz, S., T.J. Kozubowski, and K. Podgórski. *The Laplace Distribution and Generalizations*. Birkhäuser, 2001.

[83] Krekel, M. Pricing distressed CDOs with base correlation and stochastic recovery. Working Paper, 2008.

[84] Kuen, T.K., C. Erlwein, and R. Mamon. The pricing of credit default swaps under a Markov-modulated Merton's structural model. *North American Actuarial Journal*, 12:19–46, 2008.

[85] Lando, D. On Cox processes and credit risky securities. *Review of Derivatives Research*, 2:99–120, 1998.

[86] Leland, H. and K. Toft. Optimal capital structure, endogenous bankruptcy, and the term structure of credit spreads. *Journal of Finance*, 51:987–1019, 1996.

[87] Li, D. On default correlation: A copula function approach. *Journal of Fixed Income*, March 2000.

[88] Longstaff, F. and E. Schwartz. A simple approach to valuing risky fixed and floating rate debt. *Journal of Finance*, 50:789–819, 1995.

[89] Luciano, E. and W. Schoutens. A multivariate jump-driven financial asset model. *Quantitative Finance*, 6:385–402, 2006.

[90] McLeish, D.L. Estimating the correlation of processes using extreme values. *Fields Institute Communications*, 44:447–467, 2004.

[91] Merton, R. On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance*, 29:449–470, 1974.

[92] O'Kane, D. and L. Schloegl. A note on the large homogeneous portfolio approximation with the Student-t copula. *Finance and Stochastics*, 9:577–584, 2005.

[93] O'Kane, D. and S. Turnbull. Valuation of credit default swaps. Lehman Brothers Fixed Income Quantitative Credit Research, 2003.

[94] Overbeck, L. and W. Schmidt. Modeling default dependence with threshold models. *Journal of Derivatives*, Summer:10–19, 2005.

[95] Park, C. and F.J. Schuurmann. Evaluations of barrier-crossing probabilities of Wiener paths. *Journal of Applied Probability*, 13:267–275, 1976.

[96] Peskir, G. and A. Shiryaev. *Optimal Stopping and Free-Boundary Problems*. Birkhäuser Verlag, 2006.

[97] Pötzelberger, K. and L. Wang. Boundary crossing probability for Brownian motion and general boundaries. *Journal of Applied Probability*, 34:54–65, 1997.

[98] Pötzelberger, K. and L. Wang. Boundary crossing probability for Brownian motion. *Journal of Applied Probability*, 38:152–164, 2001.

[99] Rebholz, J. *Planar Diffusions with Applications to Mathematical Finance*. PhD thesis, University of California, Berkely, 1994.

[100] Ricciardi, L.M., L. Sacerdote, and S. Sato. On an integral equation for first-passage-time probability densities. *Journal of Applied Probability*, 21:302–314, 1984.

[101] Rogers, L.C.G. and D.Williams. *Diffusions, Markov Processes and Martingales. Volume 2: Itô Calculus*. Cambridge University Press, 2nd edition, 2000.

[102] Rosen, D. and D. Saunders. Analytical methods for hedging systematic credit risk with linear factor portfolios. *Journal of Economic Dynamics and Control*, To Appear.

[103] Schmidt, T. A structural model with random default boundary. Working Paper, 2006.

[104] Schönbucher, P. Taken to the limit: Simple and not-so-simple loan loss distributions. Working Paper.

[105] Shepp, L. The joint density of the maximum and its location for a Wiener process with drift. *Journal of Applied Probability*, 16:423–427, 1979.

[106] Shevchenko, P.V. Addressing the bias in Monte Carlo pricing of multi-asset options with multiple barriers through discrete sampling. *Journal of Computational Finance*, 6:1–20, 2003.

[107] Spătaru, S. An absolutely continuous function whose inverse function is not absolutely continuous. *Note di Matematica*, 23:47–49, 2004.

[108] Stout, W. *Almost Sure Convergence*. Academic Press, 1974.

[109] Vasicek, O. Loan portfolio value. *Risk*, December 2002.

[110] Yildirim, Y. Modeling default risk: A new structural approach. *Finance Research Letters*, 3:165–172, 2006.

[111] Zhou, C. An analysis of default correlations and multiple defaults. *Review of Financial Studies*, 14:555–576, 2001.

[112] Zhou, C. The term structure of credit spreads with jump risk. *Journal of Banking and Finance*, 25:2015–2040, 2001.