TU Delft

Master of Science Thesis

Swaption Pricing Under Affine Interest Rate Models

By Chao Zheng

February, 2012

For the degree of Master of Science in Applied Mathematics

DIAM, Faculty of Electrical Engineering, Applied Mathematics and Computer Science, Delft University of Technology, the Netherlands.

Master of Science Thesis

Swaption Pricing Under Affine Interest Rate Models

By Chao Zheng

Daily supervisor:

Dr. Hans Van Der Weide

Committee members:

Dr. Hans van der Weide

Dr. Ludolf Meester

Prof. Kees Oosterlee

Prof. Arnold Heemink

Preface

This thesis has been submitted in partial fulfillment of the requirements of the Master's degree in Applied Mathematics department at Delft University of Technology, the Netherlands. The research is done at TU Delft, taking about eight months to complete this final version, almost according to my own research proposal.

I like challenges in academics, especially interested in discovering new things and developing models. I try to find the theoretical gap, propose my own ideas, search for relevant literatures, and verify them by experiments. Luckily, at last, I successfully work something out. Although I experienced some failures in the meanwhile, but for me, they are part of research. Anyway, I am glad my master thesis is done.

I would like to express my sincerely gratitude to several people who helped me intensively. Thanks to my supervisor Dr. Hans van der Weide, he offered me this nice thesis project, and provided me literatures that guide me through the swaption world. Besides, he is always trying to encourage me whenever I meet difficulties, point out mistakes in my thesis carefully, and discuss with me patiently. Without his help, my thesis cannot be pleasant for reading. I learnt a lot from previous course by Prof. Arnold Heemink, who is very pleasant to offer me a very nice Phd recommendation letter. My gratitude also goes to Dr Dorota Kurowicka for the same reason. Besides, I would like to thank Prof. FM Dekking, who is quite pleased to share his nice idea with me.

In addition, thanks to my master thesis defense committee members listed as follows for their present on my thesis defense.

Dr. Hans van der Weide
Dr. Ludolf Meester
Prof. Kees Oosterlee
Prof. Arnold Heemink

February 2012

Contents

1 Introduction

Swaption pricing is a hot topic in Financial Mathematics. In this thesis, we will analyze swaptions whose short term interest rates are assumed to follow some affine models with dimension of factors more than two, so called multiple-factor interest rate models. Considering there is no analytical swaption price, we attempt to approximate it, and our discussion will focus on one of these approximation method proposed by Collin-Dufresne and Goldstein. I implement and develop this method by providing an accurate measure of approximation errors. Besides, there are several of my innovation and research recommendations on my last chapter with respect to other methods to price swaptions.

The outline of this thesis is as follows.

In chapter 2, we briefly review some important concepts in Finance, such as LIBOR, bonds, swaptions. As a swaption can be regarded as an option on a coupon bond, our general formula for swaption pricing can be derived. Then, we introduce a density approximation method by Collin-Dufresne and Goldstein, in short CDG approximation, and propose some comments.

In chapter 3, we analyze two types of errors of CDG approximation and the intrinsic estimate of errors of densities. My innovation is to develop a measure of errors of densities based on this estimate. In experiments with respect to Gamma and Lognormal densities, we find that CDG approximation tends to have good approximation performance when underlying distributions are nearly Gaussian. Moreover, experiments show that this measure of errors is accurate.

In chapter 4, we introduce two affine interest rate models, Three-factor Gaussian model and CIR2++ model, and discuss how to price swaptions under these two models respectively using CDG approximation. Experiments will be carried out to see their approximation performance. Besides, considering pricing swaption requires computation of several probabilities, I develop a measure of errors of swaption price based on our previous measure of errors of densities. Experiments display that this measure is also accurate on swaption price, in terms of both absolute errors and relative errors.

In this chapter 5, other swaption pricing methods in literatures that provide different insights into swaption other than CDG approximation are discussed. My innovation is to derive these analytical solutions in a different way from existing solutions under one-factor interest rate models, but equivalent. However, in multi-factor models, there is no analytical solution, and thus swaption price has to be approximated. In this section, we will discuss the method by Singleton and Umantsev, and my research recommendations on two potential accurate measures of errors proposed by myself. Besides, a trick to use this method to price swaptions under multiple-factor Gaussian models is discussed, also my innovation.

2 Swaption and CDG Approximation

In this chapter, we will introduce the general swaption pricing formula and one of its approximation methods proposed by Pierre Collin-Dufresne and Robert Goldstein. For convenience, we call this approximation method CDG approximation in short.

2.1 Introduction on Swaptions

Before we explain the approximation method for the price of a swaption described in the paper Collin-Dufresne and Goldstein (2002), I will first discuss several concepts in Finance.

Discount factor

The discount factor, denoted by $D(t)$, is the amount at time 0 that is "equivalent" to one unit of currency payable at time t. It is defined by

$$D(t) = e^{-\int_0^t r(s)ds}$$

where $r(s)$ is the process of short term interest rate.

In fact, the present currency per unit is always more valuable than that in the future, and a simple example is that if you store money in the bank at present, you will get more in the future because this bank will not only pay you principal but also its interest. That is the reason why the discount factor is necessary.

Risk-neutral measure

The risk-neutral measure is a probability measure that is equivalent to real probability in market and makes the discounted stock price martingales.

Let $S_1(t)$, ..., $S_n(t)$ be process of n asset price, and p be a probability measure in real market, a probability measure $\widetilde{p}$ is said to be risk neutral if

- $\tilde{p}$ and p are equivalent.
- under $\tilde{p}$, the discounted asset prices $D(t)S_i(t)$ are martingales for every $i = 1, ..., n$.

Let $V(t)$ be the price process of any derivative with maturity T and payoff $V(T)$. According to Fundamental Theorem of Asset Pricing, under risk-neutral measure, $V(t)$ can be completely hedged by assets $S_1(t)$, ..., $S_n(t)$ if and only if there exists a unique risk-neutral measure. In this situation, an investment in a portfolio can become riskless if we choose the right hedge, and this is the reason why risk-neutral measure is essential in financial world. In this thesis, without specific statement, any measure we take is the risk-neutral measure.

Fundamental Theorem of Asset Pricing also implies that the discounted price process $D(t)V(t)$ is a martingale under the risk neutral measure, so

$$D(t)V(t) = \widetilde{E_t}(D(T)V(T)) \tag{1}$$

where $\widetilde{E_t}(D(T)V(T))$ is the conditional expectation of $D(T)V(T)$ given the information up to and including time t under the risk-neutral measure.

Formula (1) is also called the risk-neutral pricing formula, and can be rewritten as

$$V(t) = \widetilde{E}_t(\frac{D(T)}{D(t)}V(T)) \tag{2}$$

Bond

In finance, a bond is a debt security, in which the issuer owes the holder a debt and, depending on the terms of the bond, is obliged to repay the principal to the holder at maturity and/or pay interest periodically.

Thus, a bond is a formal contract to repay borrowed money, and can be divided into two categories, coupon bonds and zero coupon bonds according to whether they will repay interest before maturity. Coupon bonds pay interest before maturity, while zero coupon bonds do not. Most bonds are coupon bonds.

A bond has several terms. The nominal, or the principal, of a bond is amount repaid at the end of the term excluding the interest. The maturity is the time when the issuer of bonds has to pay nominal. The coupon is the interest rate that the issuer pays to the bond holders. For instance, a two-year coupon bond with nominal 100 euro, with interest paid annually with coupon at annual rate 6%. The maturity is this example is two years. At the end of the first year, a bond holder will receive 6 euro as the first year interest, and 106 euro, nominal plus the second year interest, at the end of the year. After the maturity, the issuer has no more obligation to the bond holders.

In academic research, we denote the price at time t of a zero coupon bond with maturity T and nominal one by $B(t,T)$ where $0 \leq t \leq T$. The convention of payment at maturity equals one is made for computational convenience. Clearly, $B(0,T)$, the current price of a bond, is visible in the market, and $B(T,T) = 1$. But, for $0 < t < T$, $B(t,T)$ with T fixed is a stochastic process and of course not observable at time zero. Since a bond is an asset, we have $D(t)B(t,T)$ is a martingale, so

$$B(t,T) = \widetilde{E}_t(\frac{D(T)}{D(t)}) = \widetilde{E}_t(e^{-\int_t^T r(s)ds}) \tag{3}$$

For all $0 \leq t \leq T$.

On the other hand, in literatures, a coupon bond can be interpreted as the sum of zero-coupon bonds with maturities when coupons are paid. Let $CB(T_0)$ be the T_0 price of a coupon bond, we have

$$CB(T_0) = \sum_{i=1}^{n} C_i B(T_0, T_i) \tag{4}$$

where $T_1, T_2 ... T_n$ are coupon-paying time.

The coupon of this underlying coupon bond can be either a fixed rate or a floating rate. If the coupon k is fixed , then $C_i = k\delta_i$, $i = 1, 2, 3, ...n-1$, $C_n = (k\delta_n + 1)$, where $\delta_i = T_i - T_{i-1}$, $\delta_n = T_n - T_{n-1}$. If the coupon is floating, then C_i are random variables.

LIBOR

LIBOR, London Interbank Offered Rate, is the interest rate at which the bank is prepared to make a large wholesale deposit with other banks. Alternatively, it is an interest rate at which banks borrow money from other banks.

However, in academic research, the meaning of LIBOR is more general. Denote LIBOR at time t for period $[T, T+\delta]$ by $L(t,T)$, where δ is called the tenor of the LIBOR, and it is usually either 0.25 or 0.5 years. $L(t,T)$ means the arbitrary free interest rate that is determined at time t for the time period $[T, T+\delta]$.

Since bonds indicate interest rate, we can formulate LIBOR as follows by bonds under some model.

$$L(t,T) = \frac{B(t,T) - B(t,T+\delta)}{\delta B(t,T+\delta)} \tag{5}$$

For all $0 \leq t \leq T$.When $t = T$, we call it spot LIBOR, and otherwise, forward LIBOR.

Remark: To explain formula (5): If you take a short position of, i.e. sell, size one in a T-maturity zero-coupon bond and a long position of, i.e. buy, size $\frac{B(t,T)}{B(t,T+\delta)}$ in $(T+\delta)$-maturity zero-coupon bonds at time t. The initial cost is $\frac{B(t,T)}{B(t,T+\delta)}B(t,T+\delta) - B(t,T) = 0$. At time T, you can invest one to cover the short position, and at time $T+\delta$, repay $\frac{B(t,T)}{B(t,T+\delta)}$ due to long position. The interest rate $L(t,T)$ is determined by the equation: investment$\times$(1+duration of investment $\times$ interest rate)= repayment, or in symbols:

$$1 + \delta L(t,T) = \frac{B(t,T)}{B(t,T+\delta)}$$

Then, we have LIBOR formula (5).

Interest Rate Swap

An interest rate swap is a derivative in which one party exchanges a stream of interest payments for another party's stream of cash flows.

They can be used by hedgers to manage their fixed or floating assets and liabilities, and thus are very popular and highly liquid instruments. There are many types of interest rate swaps, and the most common one is the Fixed-for-Floating rate swap. In this swap, a company agrees to pay cash flows equal to interest at a predetermined fixed rate over a nominal principle for a number of years. In return, it receives a floating rate on the same principle for the same period of time.

The floating interest rate in most interest swap agreements is LIBOR. Thus, in this thesis, we only discuss interest rate swap of fixed interest rate for LIBOR under the same currency.

In terms of the direction of payments, there are two types of swaps: receiver swaps and payer swaps. The holder of a receiver swap receives a fixed rate and pays a floating rate. For the holder of a payer swap, the payments go in the other direction.

We will introduce a type of swaps called $T_0 * (T_n - T_0)$ swaps, and the payments in such a receiver swap with nominal K and fixed interest rate k are as follows:

- Payments will be made and received at $T_i = \delta i$, $i = 1, ..., n$.
- For every elementary period $[T_i, T_{i+1}]$, $i = 0, ..., n-1$, the LIOBR rate is set at time T_i and the floating leg $K\delta L(T_i, T_i)$, spot LIBOR in duration $[T_i, T_i + \delta]$, is paid at T_{i+1}.
- For the same period the fixed leg $K\delta k$ is received at T_{i+1}.

where δ is constant tensor of the LIBOR.

Swaption

A swaption is an option granting its owner the right but not the obligation to enter into an underlying swap of interest rate. In short, it is an option on a swap. Corresponding to payer swaps and receiver swaps, there are payer swaptions and receiver swaptions.

A $T_0 * (T_n - T_0)$ receiver swaption is a contract which at the exercise date T_0 gives the holder the right to enter into a $T_0 * (T_n - T_0)$ swap. Since swaption is a type of derivatives of interest rate, by formula (2), we have the price of a $T_0 * (T_n - T_0)$ swaption at time t , $0 \leq t < T_0$

$$\widetilde{E_t}(\frac{D(T_0)}{D(t)}(FS(T_0))^+) \tag{6}$$

where T_0 is the maturity date of the swaption, and $FS(T_0)$ is the value of the swap on the swaption at time T_0. For a receiver swaption, we have

$$FS(T_0) = \widetilde{E}_{T_0}(\sum_{i=0}^{n-1} \frac{D(T_{i+1})}{D(T_0)}(L(T_i, T_i) - k)K\delta) \tag{7}$$

2.2 Swaption Interpreted as Option on Coupon Bond

Consider a $T_0 * (T_n - T_0)$ swaption, this kind of swaption can be interpreted as an option on a coupon bond, where the strike is equal to the nominal of the contract, and the coupon rate is equal to the swap rate strike of the swaption. This means problems on swaptions can be converted into those on coupon bonds.

We will provide its proof in the following part, but before that, I will first introduce the forward measure, which is quite a useful tool for this problem.

Forward measure

Let T be a fixed maturity date. We define the T-forward measure $\widetilde{p}^T$ by

$$\widetilde{p}^T(A) = \frac{1}{B(0,T)} \int_A D(T) d\widetilde{p}$$

for all $A \in \mathcal{F}$, where A is any subset of Ω in probability space $(\Omega, \mathcal{F}, \widetilde{p})$, and $\mathcal{F}$ is its σ-algebra, $\widetilde{p}$ is risk-neutral probability.

By this definition, $\widetilde{p}^T$ is a probability measure equivalent to $\widetilde{p}$ for any T since $\widetilde{E}(\frac{D(T)}{B(0,T)}) = \frac{1}{B(0,T)}\widetilde{E}(\frac{D(T)}{D(0)}) = 1$ and $\frac{D(T)}{B(0,T)} > 0$, refer to Shreve(2004). Forward measures are quite popular in Finance, and there are two properties related.

Let $\widetilde{w}(t)$ be a Brownian motion on the probability space $(\Omega, \mathcal{F}, \widetilde{p})$. Since $B(t,T)$ is the price process of an asset, it follows that $D(t)B(t,T)$ is a martingale under $\widetilde{p}$. Then, according to the Martingale Representation Theorem, there is a volatility process $\sigma^*(t,T)$ for the bond , a process in t, T is fixed, such that

$$d(D(t)B(t,T)) = -\sigma^*(t,T)D(t)B(t,T)d\widetilde{w}(t) \tag{8}$$

Property 1.1. Define the process $\widetilde{w}^T(t)$ to be

$$\widetilde{w}^T(t) = \int_0^t \sigma^*(u,T)du + \widetilde{w}(t) \tag{9}$$

Then, $\widetilde{w}^T(t)$ is a Brownian motion under T-forward measure $\widetilde{p}^T$.

Definition (Forward Price). T-forward price at time t is the strike price of the forward contract that causes this contract at time t to have value zero. Suppose $V(t)$ is the price of an asset, then $\frac{V(t)}{B(t,T)}$ is the T-forward price of $V(t)$ at time t.

Property 1.2. T-forward price is a martingale under T-forward measure $\widetilde{p}$.

It implies that

$$\widetilde{E}_t^T(\frac{V(T_0)}{B(T_0,T)}) = \frac{V(t)}{B(t,T)}$$

where $t \leq T_0 \leq T$.

This gives us a simple formula

$$V(t) = B(t,T)\tilde{E}_t^T(\frac{V(T_0)}{B(T_0,T)}) \tag{10}$$

Combining (2) and (10), we get

$$\widetilde{E_t}(\frac{D(T_0)}{D(t)}V(T_0)) = B(t,T)\tilde{E}_t^T(\frac{V(T_0)}{B(T_0,T)}) \tag{11}$$

(11) states the relationship between expectation under risk-neutral measure and that under forward measure. For the proof of both properties, see Shreve(2004).

Theorem 1.1. LIBOR $L(t,T_i)$, $t \leq T_i$, is a martingale under $T_{i+1}-$forward measure

Proof. Let $0 \leq t < T_0 \leq T_i$, by equation (5), definition of $L(T_0, T_i)$, we have

$$\begin{aligned} \tilde{E}_t^{T_{i+1}}(L(T_0,T_i)) &= \tilde{E}_t^{T_{i+1}}(\frac{B(T_0,T_i)-B(T_0,T_{i+1})}{\delta B(T_0,T_{i+1})}) \\ &= \frac{1}{\delta}\tilde{E}_t^{T_{i+1}}(\frac{B(T_0,T_i)}{B(T_0,T_{i+1})}) - \frac{1}{\delta} \end{aligned} \tag{12}$$

By (11), where $V(T_0) = B(T_0, T_i)$, we obtain

$$\begin{aligned} \tilde{E}_t^{T_{i+1}}(\frac{B(T_0,T_i)}{B(T_0,T_{i+1})}) &= \frac{1}{B(t,T_{i+1})}\widetilde{E_t}(\frac{D(T_0)}{D(t)}B(T_0,T_i)) \\ &= \frac{1}{B(t,T_{i+1})}\widetilde{E_t}(\frac{D(T_0)}{D(t)}\tilde{E_{T_0}}(\frac{D(T_i)}{D(T_0)})) \end{aligned} \tag{13}$$

As $T_0 > t$, information on T_0 is more specific than that on t, by Tower rule, we have

$$\begin{aligned} \widetilde{E_t}(\frac{D(T_0)}{D(t)}\tilde{E_{T_0}}(\frac{D(T_i)}{D(T_0)})) &= \widetilde{E_t}(\tilde{E_{T_0}}(\frac{D(T_i)}{D(t)})) \\ &= \widetilde{E_t}(\frac{D(T_i)}{D(t)}) \\ &= B(t,T_i) \end{aligned} \tag{14}$$

Combining equation (13) and (14), equation (12) yields

$$\begin{aligned} \tilde{E}_t^{T_{i+1}}(L(T_0,T_i)) &= \frac{1}{\delta}\tilde{E}_t^{T_{i+1}}(\frac{B(T_0,T_i)}{B(T_0,T_{i+1})}) - \frac{1}{\delta} \\ &= \frac{1}{\delta}\frac{B(t,T_i)}{B(t,T_{i+1})} - \frac{1}{\delta} \\ &= \frac{B(t,T_i)-B(t,T_{i+1})}{\delta B(t,T_{i+1})} \\ &= L(t,T_i) \end{aligned}$$

In addition, we assume $L(t, T_i)$ is bounded and thus $\tilde{E}^{T_{i+1}}(L(t,T_i)) < \infty$. This assumption is reasonable, because in real market LIBOR is always a bounded value. Therefore, LIBOR $L(t, T_i)$ is martingale under T_{i+1}–forward measure.

■

Theorem 1.2. A swaption is equivalent to an option on coupon bond, where the strike is equal to the nominal of the contract, and the coupon rate is equal to the swap rate strike of the swaption.

Proof: Consider a $T_0 * (T_n - T_0)$ payer swap exchanging fixed rate k for floating rate LIBOR and its nominal is K, from (7), we have

$$\begin{aligned} FS(T_0) &= \tilde{E}_{T_0}(\sum_{i=0}^{n-1} \frac{D(T_{i+1})}{D(T_0)}(L(T_i,T_i)-k)K\delta) \\ &= \sum_{i=0}^{n-1} \tilde{E}_{T_0}(\frac{D(T_{i+1})}{D(T_0)}(L(T_i,T_i)-k)K\delta) \\ &= \sum_{i=0}^{n-1} B(T_0,T_{i+1})\tilde{E}_{T_0}^{T_{i+1}}((L(T_i,T_i)-k)K\delta) \end{aligned}$$

The last step above uses Property 1.1. By Theorem 1.1, we have $\tilde{E}_{T_0}^{T_{i+1}}(L(T_i,T_i)) = L(T_0,T_i)$ .Thus

$$\begin{aligned} FS(T_0) &= \sum_{i=0}^{n-1} B(T_0,T_{i+1})((L(T_0,T_i)-k)K\delta) \\ &= \sum_{i=0}^{n-1} B(T_0,T_{i+1})((\frac{B(T_0,T_i)-B(T_0,T_{i+1})}{\delta B(T_0,T_{i+1})}-k)K\delta) \\ &= K\sum_{i=0}^{n-1}(B(T_0,T_i)-(k\delta+1)B(T_0,T_{i+1})) \end{aligned}$$

The last equation can be rewritten as

$$\begin{aligned} FS(T_0) &= KB(T_0,T_0) - \sum_{i=1}^{n} C_i B(T_0,T_i) \\ &= K - \sum_{i=1}^{n} C_i B(T_0,T_i) \end{aligned}$$

where $C_i = k\delta K$,$i = 1,2,3,...n-1$; $C_n = (k\delta+1)K$;

Recall the formula for the coupon bond $CB(T_0)$ is

$$CB(T_0) = \sum_{i=1}^{n} C_i B(T_0,T_i) \tag{15}$$

Therefore, by formula (6), the price of payer swaption at time t with maturity T_0 is

$$\widetilde{E_t}(\frac{D(T_0)}{D(t)}(K - CB(T_0))^+)$$

Thus, this payer swaption can be regarded as a put option on coupon bond $CB(T_0)$ with strike K and maturity T_0.

Similarly, consider a $T_0 * (T_n - T_0)$ receiver swap exchanging fixed rate k for floating rate LIBOR and its nominal is K, we have

$$\begin{aligned} FS(T_0) &= \tilde{E}_{T_0}(\sum_{i=0}^{n-1} \frac{D(T_{i+1})}{D(T_0)}(k - L(T_i, T_i))K\delta) \\ &= K\sum_{i=0}^{n-1}(-B(T_0, T_i) + (k\delta + 1)B(T_0, T_{i+1})) \\ &= \sum_{i=1}^{n} C_i B(T_0, T_i) - K \end{aligned}$$

where $C_i = k\delta K$,$i = 2, 3, ...n - 1$; $C_n = (k\delta + 1)K$;

For a receiver swaption, its price at time t with maturity T_0 is

$$\widetilde{E}_t(\frac{D(T_0)}{D(t)}(CB(T_0) - K)^+) \tag{16}$$

Thus, this receiver swaption can be regarded as a call option on coupon bond $CB(T_0)$ with strike price K and maturity T_0.

■

Therefore, the problems on a swaption can always be converted into those on an option of a coupon bond, as long as this swaption is on exchange of fixed interest rate for LIBOR under the same currency. For other cases, such as ones on swaption based on interest rate under different currencies, this theorem may not apply because it contains foreign exchange rate which turns out to be more complicated.

2.3 Swaption Pricing Formula and CDG Approximation

In this part, we will formulate a formula for swaption pricing, and then, introduce Collin-Dufresne and Goldstein approximation method.

Consider a receiver swaption discussed above whose price at time t is denoted as $Swn(t)$, from (16), we have

$$
\begin{aligned}
Swn(t) &= \widetilde{E_t}(\frac{D(T_0)}{D(t)}(CB(T_0)-K)^+) \\
&= \widetilde{E_t}(\frac{D(T_0)}{D(t)}(CB(T_0)1_{(CB(T_0)>K)} - K1_{(CB(T_0)>K)})) \\
&= \sum_{i=1}^{n} C_i\widetilde{E_t}(\frac{D(T_0)}{D(t)}1_{(CB(T_0)>K)}B(T_0,T_i) - K1_{(CB(T_0)>K)})) \\
&= \sum_{i=1}^{n} C_i\widetilde{E_t}(\frac{D(T_0)}{D(t)}1_{(CB(T_0)>K)}\tilde{E_{T_0}}(\frac{D(T_i)}{D(T_0)}) - K1_{(CB(T_0)>K)})) \\
&= \sum_{i=1}^{n} C_i\widetilde{E_t}(\frac{D(T_i)}{D(t)}1_{(CB(T_0)>K)}) - K\widetilde{E_t}(\frac{D(T_0)}{D(t)}1_{(CB(T_0)>K)}) \qquad (17)
\end{aligned}
$$

where $\widetilde{P}_t^{T_i}$ on the last line of the equation is the probability conditioned on information at time t under T_i forward measure. Use Property 1.2 , where $V(T_0) = B(T_0,T_i)1_{(CB(T_0)>K)}$, we can rewrite (17) as

$$
\begin{aligned}
Swn(t) &= \sum_{i=1}^{n} C_iB(t,T_i)\widetilde{E_t}^{T_i}(1_{(CB(T_0)>K)}) - KB(t,T_0)\widetilde{E_t}^{T_0}(1_{(CB(T_0)>K)}) \\
&= \sum_{i=1}^{n} C_iB(t,T_i)\widetilde{P}_t^{T_i}(CB(T_0)>K) - KB(t,T_0)\widetilde{P}_t^{T_0}(CB(T_0)>K) \qquad (18)
\end{aligned}
$$

Equation (18) is our general receiver swaption pricing formula. It can be interpreted as stating that the price of a swaption is related to a series of probabilities $\widetilde{P}_t^{T_i}(CB(T_0)>K)$ of the underlying coupon bond $CB(T_0)$, and thus, we have to approximate these probabilities under each forward measures $T_0, T_1, ...T_n$ if there is no closed form solutions. In the year of 2002, Collin-Dufresne and Goldstein worked out an approximation method to solve this problem, we call it CDG approximation in short. The basic idea of their method is to use cumulants of the underlying coupon bond $CB(T_0)$ to approximate these probabilities, where the cumulants are obtained by their corresponding moments. In the following part, we will be present CDG approximation specifically.

CDG approximation is based on an assumption that the price of the bond can be written in the following exponential,

$$B(t,T_i) = e^{-B_0(T_i-t)-\sum_{j=1}^{J} B_j(T_i-t)X_j(t)}$$

where $X_1(t)$, ...,$X_J(t)$ are factors of the interest rate, which are all stochastic process satisfying certain stochastic differential equations. $B_0(\tau)$, $B_1(\tau)$, ..., $B_J(\tau)$ are deterministic functions with variable τ.

Under this assumption, for any positive integer m, we can conclude from (15), the formula of $CB(T_0)$, that$(CB(T_0))^m$ can be written as sum of exponentials

$$(CB(T_0))^m = \sum_{i_1,i_2,...,i_m=1}^{n} (C_{i_1}...C_{i_m}) e^{-F_0 - \sum_{j=1}^{J} X_j(T_0) F_j} \tag{19}$$

where any coefficient $F_j = \sum_{k=1}^{m} B_j(T_{i_k} - T_0)$, $j = 0, 1, ...J$. Note that F_j relies on the choice of i_1, i_2, ...,i_m, where $i_k \in \{1, 2, ...n\}$.

According to Duffie, Pan, and Singleton (2000), the expectation of $(CB(T_0))^m$ conditioned on information up to time t also has an exponentially affine solution

$$\tilde{E}_t^{T_i}(CB(T_0))^m = \sum_{i_1,i_2,...,i_m=1}^{n} (C_{i_1}...C_{i_m}) e^{-H_0(T_0-t) - \sum_{j=1}^{J} X_j(T_0) H_j(T_0-t)} \tag{20}$$

where the deterministic functions $H_j(\tau)$, $j = 0, 1, 2...J$, satisfy a set of Riccati equations. Hence, the equation above demonstrates that all moments of coupon bond prices have analytic solutions within an affine framework.

Having known the moments of $CB(T_0)|\mathcal{F}(t)$ by formula (20), it is easy to calculate the corresponding cumulants by moments, since there is an one-to-one correspondence between moments and cumulants. More specifically, cumulants can be uniquely determined by moments with formulas showed in Appendix A, and vice versa. The following is the introduction of Riccati equations as well as Fourier Inversion Theorem.

Riccati Equations

A Riccati equation is any ordinary differential equation of the form: $y'(x) = q_0(x) + q_1(x)y(x) + q_2(x)y^2(x)$. It can always be reduced to a second order linear ODE, and thus the problems of solving Riccati equations can be reduced to solving a second order linear ODEs. Besides, if a particular solution is known, there is a method for its other solutions. However, theoretically, a Riccati equation might have no solution or more than one solutions depending on parameters $q_0(x)$, $q_1(x)$, $q_2(x)$. It is well known that even for a linear ODE, if the parameters are not all constants, solving this equation is in general rather difficult. In CDG approximation, it is required to obtain a unique solution for system of Riccati equations, and if its solution is not analytical, we have to seek numerical approach, which might be time consuming.

Fourier Inversion Theorem: Let φ be the characteristic function of the distribution F and suppose that $\varphi \in L^1$ [1] Then, F has bounded continuous density f given by

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-i\zeta x} \varphi(\zeta) d\zeta$$

where the characteristic function of a random variable Y is defined as $\varphi(\zeta) = E(e^{i\zeta y})$.

[1] $\varphi \in L^1$ if and only if φ is integrable over $(-\infty, +\infty)$.

After a short review of Riccati equations and Fourier Inversion Theorem, we will go back to CDG approximation. Let $G(k)$ to be the characteristic function of the random variable $CB(T_0)|\mathcal{F}(t)$ with t and T_0 fixed. Using Taylor expansion at the original, or Maclaurin expansion, we can get

$$log(G(k)) = \sum_{j=1}^{\infty} c_j \frac{(ik)^j}{j!}$$

where c_j is defined as the j-th order cumulant of $CB(T_0)|\mathcal{F}(t)$.

It follows from Fourier Inversion Theorem that the density function of $CB(T_0)|\mathcal{F}(t)$ is

$$\begin{aligned} f_t(y) &= \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-iky}G(k)dk \\ &= \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-iky}e^{\sum_{j=1}^{\infty} c_j \frac{(ik)^j}{j!}}dk \\ &= \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-iky}e^{ikc_1-\frac{k^2}{2}c_2}e^{\Lambda}dk \end{aligned} \tag{21}$$

where $\Lambda = \sum_{j=3}^{\infty} c_j \frac{(ik)^j}{j!}$.

If the characteristic function $G(k)$ of $CB(T_0)|\mathcal{F}(t)$ is known, some Fast Fourier Transform techniques can be applied for densities, such as Carr and Madan (1999). However, in many cases of interest, $G(k)$ is unknown, and thus we have to turn to (21) for a general discussion.

Collin-Dufresne and Goldstein approximates

$$\Lambda \approx \sum_{j=3}^{m} c_j \frac{(ik)^j}{j!}$$

where m is an integer and $m > 3$, they also approximates

$$e^{\Lambda} \approx 1 + \Lambda + \frac{\Lambda^2}{2}$$

In addition, $1 + \Lambda + \frac{\Lambda^2}{2}$ is further approximated by ignoring terms with order higher than m. [2]

Denoted $g_t(y)$ as the approximation of density function $f_t(y)$. After integration, $g_t(y)$ can be written as

$$g_t(y) = \frac{1}{\sqrt{2\pi c_2}} e^{-\frac{(y-c_1)^2}{2c_2}} (\sum_{j=0}^{m} \gamma_j (y-c_1)^j) \tag{22}$$

[2] The order of terms is defined based on the order of cumulants in this term. If a term is c_i, we define the order to be i. For a term is $c_i c_j$, the order is defined as $i + j$.

In addition

$$\int_K^{\infty} g_t(y)dy = \sum_{j=0}^{m} \gamma_j \lambda_j \tag{23}$$

where coefficient γ_j is some function of cumulants and λ_j is the combination of some normal distribution functions of variable K, nominal of the swaption contract.

Formulas (22) and (23) are the final result of CDG approximation that allow us to access the approximated swaption price. We omits the detailed calculation, because CDG approximation is a special case of Edgeworth expansion, which will be discussed in section 3.1 specifically.

Collin-Dufresne and Goldstein indicate their approximation offers an excellent balance between speed and accuracy if $m = 7$, alternatively, approximation by the first seven cumulants. So, we will only deal with CDG approximation with $m = 7$ in the rest of my thesis. Obviously, with different m, even the same coefficient with the same index would have the different term. For coefficients γ_j and λ_j when $m = 7$, refer to Appendix B.

In addition, we can see in CDG approximation, one has to approximate $\widetilde{P}_t^{T_i}(CB(T_0) > K)$ under each T_i forward measures $i = 0, 1, ..., n$. For every such a forward probability, there is a correspondent expectation $\tilde{E}_t^{T_i}(CB(T_0))^m$, which contains $\binom{n+m-1}{m}$ different terms $(C_{i_1}...C_{i_m})e^{-H_0(T_0-t)-\sum_{j=1}^{J} X_j(T_0)H_j(T_0-t)}$, and every such term may have to be calculated separately. For instance, when $n = 10$ and $m = 7$, $\binom{16}{7} = 11440$. Figure 1.1 illustrates how large number of terms will become as n increases, where we fix $m = 7$. We can see that the computation complexity is quite sensible to n, the number of time of swap. Alternatively, with n goes up, the computation cost will increase sharply. Therefore, a potential disadvantage of CDG approximation is that the computation may be expensive in cases when n is quite large, and functions $H_j(\tau)$, $j = 0, 1...m$ satisfying certain Raccati equations, have no analytical solutions, thus have to be calculated numerically by Runge Kutta method.

However, if $H_j(\tau)$ have analytically solutions, the computation is very fast, for instance, under Three-factor Gaussian interest rate model. Collin-Dufresne and Goldstein claim it takes less than 0.05 seconds to calculate the swaption price under Three-factor Gaussian model when $n = 20$, but in all of their experiments, they set cumulants $c_6 = c_7 = 0$ to save computational cost. The paper by Collin-Dufresne and Goldstein is published in the year 2002, and we assume that the speed of computes doubles every two years, so in 2012, the speed is 32 times faster. Considering on that, we are more interested in the accuracy of CDG approximation, and thus we will never set cumulants $c_6 = c_7 = 0$ instead to avoid losing accuracy while applying CDG approximation in the rest of my thesis.

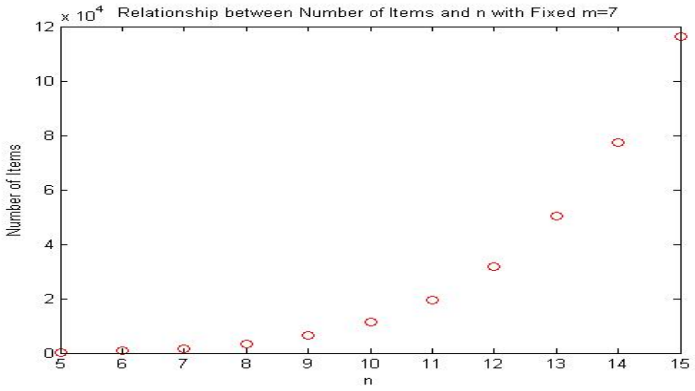


Figure 1.1. Relationship between Number of Items and n with Fixed m=7

3 Implement to CDG Approximation

3.1 Edgeworth Expansion

CDG approximation is somewhat similar to Edgeworth expansion, but there is a large difference between their ideas. We will first introduce Edgeworth expansion, and then make a comparison with CDG approximation.

Edgeworth expansion is a density approximation method for the distribution of the sum of large number of standardized[3] independent and identical random variables in statistics. It can be regarded as an improvement to the Central Limit Theorem. The idea was proposed by Edgeworth at the early of 20th century. However, it was not until the year 1978, that the fundamental work was done to make the classical formalism rigorous, see Bhayyacharya and Ghosh (1978).

To further understand this, let X_j $j = 1, 2, ..., n$ be iid random variables with mean μ and variance σ^2. Let $Y_j = \frac{X_j - \mu}{\sigma}$ for each j, and $S_n = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} Y_j$. We have by Central Limit Theorem

$$S_n = \frac{\sum_{j=1}^{n} X_j - n\mu}{\sqrt{n}\sigma} \to N(0, 1)$$

as $n \to \infty$. Where $N(0, 1)$ is the standard Normal distribution function and $S_n \to N(0, 1)$ means convergence in distribution.

Actually, in statistics, the distribution of S_n is frequently approximated by the standard Normal distribution for confidence intervals of some parameter estimator of X, if n is sufficiently large. However, Edgeworth expansion provides a more accurate approximation in general for the distribution of S_n using cumulants of Y.

We will discuss Edgeworth Expansion as follows, see Hall (1992).

Let $\chi(t)$ to be the characteristic function of Y and $\chi_n(t)$ to be that of S_n. It follows that

$$\begin{aligned} \chi_n(t) &= E(e^{it\frac{1}{\sqrt{n}}\sum_{j=1}^{n} Y_j}) \\ &= [E(e^{it\frac{1}{\sqrt{n}}Y_j})]^n \\ &= [\chi(\frac{t}{\sqrt{n}})]^n \end{aligned}$$

On the other hand, since Y is standardized, we have $c_1 = E(Y) = 0$, $c_2 = Var(Y) = 1$, then

$$\chi(\frac{t}{\sqrt{n}}) = exp(-\frac{1}{2n}t^2 + \sum_{j=3}^{\infty} c_j \frac{(it)^j}{n^{j/2} j!})$$

Thus, by Taylor expansion of exponential function $e^X = 1 + X + \frac{1}{2}X^2 + ... + \frac{1}{n!}X^n + ...$, we have

[3] A standardized random variable is the one with mean zero and variance one.

$$\begin{aligned}
\chi_n(t) &= [\chi(\frac{t}{\sqrt{n}})]^n \\
&= exp(-\frac{1}{2}t^2 + \sum_{j=3}^{\infty} c_j n \frac{(it)^j}{n^{j/2} j!}) \\
&= e^{-t^2/2} exp(\sum_{j=1}^{\infty} c_{j+2} \frac{(it)^{j+2}}{n^{j/2}(j+2)!}) \\
&= e^{-t^2/2}(1 + \sum_{j=1}^{\infty} \frac{r_j(it)}{n^{j/2}})
\end{aligned}$$

where r_j is a polynomial with real coefficients, of degree $3j$, depending on $c_3, c_4, ..., c_{j+2}$, but not on n.

As

$$\begin{aligned}
\frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-ity} e^{-t^2/2} (it)^k dt &= \frac{(-1)^k}{2\pi}\frac{d^k}{dy^k}\int_{-\infty}^{+\infty} e^{-ity} e^{-t^2/2} dt \\
&= (-1)^k He_k(y)\phi(y) \qquad (24)
\end{aligned}$$

where $\phi(y)$ is the density function of standard Normal distribution, and $He_k(y)$ is the $k-th$ probabilists' Hermite polynomial. [4]

Thus, by inverse Fourier transform, the density function of S_n is

$$\begin{aligned}
f_n(y) &= \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-ity}\chi_n(t)dt \\
&= \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-ity} e^{-t^2/2}(1 + \sum_{j=1}^{\infty}\frac{r_j(it)}{n^{j/2}})dt \\
&= \phi(y) + \sum_{j=1}^{\infty}\frac{p_j(y)}{n^{j/2}}\phi(y) \\
&= \phi(y) + \sum_{j=1}^{m}\frac{p_j(y)}{n^{j/2}}\phi(y) + o(n^{-m/2}) \qquad (25)
\end{aligned}$$

where p_j is some combination of Hermit polynomial. Besides, it is also a polynomial of degree $3j$, depending on $c_3, c_4, ..., c_{j+2}$ similar to r_j.

[4] Probabilists' Hermite polynomials are a classical orthogonal polynomial sequence that arises in probability, named after Charles Hermite. $He_n(y) = (-1)^n e^{y^2/2}\frac{d^n}{dx^n}e^{-y^2/2}$ by definition. There are two valuable properties related. One is orthogonal property: $\int_{-\infty}^{+\infty} He_m(y)\ He_n(y)e^{-y^2/2}dy = \sqrt{2\pi}n!\delta_{nm}$, where $\delta_{nm} = 0$ if $m \neq n$ and $\delta_{nm} = 1$ if $m = n$. The other is recursion: $He_{n+1}(y) = yHe_n(y) - He_n^{'}(y)$. We list some of probabilists' Hermite polynomials: $He_0(y) = 1$, $He_1(y) = y$, $He_2(y) = y^2 - 1$, $He_3(y) = y^3 - 3y$.

Due to the following property of Hermite polynomial

$$\frac{d}{dy}(He_k(y)\Phi(y)) = -He_{k+1}(y)\phi(y)$$

where $\Phi(y)$ is the standard Normal distribution function.

The distribution function of S_n is

$$\begin{aligned} F_n(y) &= \Phi(y) + \sum_{j=1}^{\infty} \frac{P_j(y)}{n^{j/2}}\Phi(y) \\ &= \Phi(y) + \sum_{j=1}^{m} \frac{P_j(y)}{n^{j/2}}\Phi(y) + o(n^{-m/2}) \end{aligned} \tag{26}$$

where P_j is a polynomial of degree $3j-1$. Besides, Blinnikov and Moessner (1998) have given a simple algorithm to calculate P_j in higher-order.

We illustrate Edgeworth expansion only for standardized random variables Y, not only because it is an easier approach, but also because it is more practical in statistics. But, for random variables without standardizing, Edgeworth expansion still works for density and distribution function approximation with results similar to (25) and (26).

Although suffered from some theoretical problems in convergence, Edgeworth expansion is still a powerful tool in statistics for constructing confidence bounds in parameters estimation, for instance, the expectation, of some distributions. Edgeworth expansion is a classical but still hot topic now and there are many research related, such as, Bickel, Götze and Zwet (1986), Bloznelis and Götze (2000) due to the fact that it is an asymptotic expansion in many cases of interest. In addition, from (25) and (26), we know its error is $o(n^{-m/2})$. It implies n should be sufficiently large in order to keep this expansion accurate.

However, although CDG approximation also deals with cumulants for distribution approximation in a similar way, the largest difference in between is that it approximates the density or distribution function of a single random variable, instead of sum of iid random variables. That's to say, CDG approximation is the Edgeworth expansion with $n = 1$ using only the first seven cumulants and second order Taylor expansion. We know that the power of Edgeworth expansion depends on the large number of samples n due to its error $o(n^{-m/2})$, so it seems that CDG approximation is not asymptotic in general.

Actually, there is another expansion similar to CDG approximation called Gram-Charlier expansion, which also approximates density functions by cumulants, but they differ in arrangement of terms and thus the accuracy of truncation. Tanaka (2005) provides a method to price swaption using Gram-Charlier expansion, in which the third order cumulant approximation is the best for computational accuracy and efficiency instead.

Except swaptions, similar approximations based on cumulants are used to price options, see Jarrow and Rudd (1982), Asian options, see Turnbull and Wakeman (1991), in which the approximation price is a close-form solution. However, the accuracy of approximation is rather limited for these cases, see Ju (2001).

3.2 Assumption Analysis

As a special case of Edgeworth expansion, CDG approximation also suffers from some convergent problems. In literatures, relevant researches mainly focus on studying when Edgeworth series of a certain random variable will converge to its distribution and its rate of convergence, see Hall (1992). However, there is something in common among these problems, if we view them from a different perspective.

In CDG approximation, we calculate the density function using Fourier Inversion Theorem

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-itx}\varphi(t)dt$$

but replace the characteristic function $\varphi(t)$ through (27)

$$log(\varphi(t)) = \sum_{j=1}^{\infty} c_j \frac{(it)^j}{j!} \tag{27}$$

if there is a Taylor expansion of $log(\varphi(t))$ on the origin.

Theoretically, CDG approximation requires that the right side of (27) always converges to $log(\varphi(t))$ for any $t \in (-\infty, +\infty)$, because the calculation of its correspondent density function $f(x)$ requires an integration of $\varphi(t)$ over $t \in (-\infty, +\infty)$.

We take it as an assumption of CDG approximation. However, Taylor expansion is a local expansion at a certain point, and thus condition (27) can only be satisfied in some domain $|t| < r$, where $0 \leq r \leq \infty$. Note that r is not larger than the radius of convergence of Taylor series, denoted as R,[5]right side of (27).

In the following part, we will seek a way to determine its radius r in a more general Taylor expansion.

Definition (Holomorphic function). A holomorphic function is a complex function that is differentiable in a neighborhood of every point in its domain.

Taylor Theorem. Suppose a complex function $f(x)$ is holomorphic in the domain D. For any point $x \in B(x_0, R') \subseteq D$, we have $f(x) = \sum_{j=1}^{\infty} \frac{f^{(j)}(x_0)}{j!}(x - x_0)^j$.
where $B(x_0, R') = \{x \,|\, |x - x_0| < R'\}$ and R' is an arbitrary real number.

Followed by Taylor Theorem and a property that $\sum_{j=1}^{\infty} \frac{f^{(j)}(x_0)}{j!}(x - x_0)^j$ is holomorphic if convergent, we have Lemma 3.1 below.

[5]**Radius of Convergence of Taylor Series**

Let $\sum_{j=1}^{\infty} a_j x^j$ be a Taylor expansion of $f(x)$, real or complex, where $a_j = \frac{f^{(j)}(x_0)}{j!}$. It is true that $\sum_{j=1}^{\infty} a_j x^j$ converges within $|x| < R$, $0 \leq R \leq \infty$., where R is called the radius of convergence of Taylor series. By Cauchy-Hadamard formula, $R^{-1} = \overline{\lim_{j\to\infty}} \sqrt[j]{|a_j|}$. In addition, $R^{-1} = \lim_{j\to\infty} |\frac{a_{j+1}}{a_j}|$ if this limit exists, and thus, in (27), $R = \lim_{j\to\infty} |\frac{(j+1)c_j}{c_{j+1}}|$.

Lemma 3.1 Suppose $f(x)$ is holomorphic at $x = x_0$, and x_1 is the nearest singular from x_0, then the radius r is $|x_0 - x_1|$.

This lemma provides a simple approach to obtain r in (27). Firstly extend the domain of the random variable t of $log(\varphi(t))$ into a complex field, then find out the nearest singular from origin, and the radius r is the distance in between.

Table 3.1 below displays characteristic functions of different distributions, and their corresponding radius r

	Characteristic Function $\varphi(t)$	r
Gamma $\Gamma(\alpha, \beta)$	$(1 - \frac{it}{\beta})^{-\alpha}$	β
Normal $N(\mu, \sigma^2)$	$e^{it\mu - \frac{1}{2}\sigma^2 t^2}$	∞

Table 3.1. Characteristic Functions and Radius of Gamma and Normal Distributions

Clearly there is only one singular $t = -\beta i$ in $log(\varphi(t))$ of Gamma distribution. Thus its $r = \beta$. But, for Normal distribution, there is no singular, and hence $r = \infty$. Although in Gamma distribution, $r = \beta$, other than ∞, not satisfied with our assumption, later experiments in this chapter show it is still possible to get an accurate approximation for its density function by CDG approximation.

When Cumulants Determine Unique Distribution

Readers may be interested in such a question: Is it possible that two different distributions will have the same cumulants, and thus the same CDG approximation? The answer is yes. In this case, cumulants cannot determine a unique distribution.

The problem when cumulants determine unique distribution is equivalent to the problem when moments do, since there is also a one-to-one correspondence between cumulants and moments. Unfortunately, in most cases, moments cannot determine the distribution uniquely, and there is a classic example.

Suppose X is Lognormal distributed and its probability density f is

$$f(x) = \frac{1}{\sqrt{2\pi}} x^{-1} e^{-0.5(logx)^2}, x > 0$$

We can construct another probability density f_a based on f

$$f_a(y) = f(y)(1 + a * sin(2\pi logy))$$

where $-1 \leq a \leq 1$, and $y > 0$. Then, f and f_a have the same moments, and thus have the same cumulants. See Feller (1971).

We provide two theorems about when moments can determine unique distributions. All of them are sufficient but not necessary, so even moments satisfy neither of theorems above, it is still possible that its corresponding distribution is unique.

Theorem 3.1 let F be a distribution function on $(-\infty, +\infty)$, F is uniquely determined by its moments if

$$\sum_n \mu_{2n}^{-1/2n} = \infty$$

where μ_i is the i-th moment of F.

Theorem 3.2: F is uniquely determined by its moments whenever

$$\sum_n \mu_{2n} t^n/(2n)!$$

converges in some interval: $|t| < t_0$ where $t_0 > 0$.

Remark: The general idea of proof of both theorems is to see when moments can determine a unique characteristic function, since there is a one-to-one correspondence in distribution and characteristic function. See Feller (1971).

Theorem 3.1 and Theorem 3.2 coincide in the way that both theorems state that its $2n$-th moments cannot increase too fast as n increases if the distribution is uniquely determined. But it seems that Theorem 3.1 is too strong thus we prefer Theorem 3.2 in most situations.

Actually, there is no close relationship between this moment property and our assumption (27). We will show in next part that cumulants, or moments, from Gamma distribution can determine a unique distribution, while those from Lognormal distribution cannot, which satisfies neither conditions in Theorem 3.1 nor Theorem 3.2. Unfortunately, most distributions cannot be determined uniquely by cumulants, and it is true that two different distributions with the same cumulants would lead to the same distribution by CDG approximation if their cumulants exist and are finite. This is a flaw of CDG approximation, and obviously, of Edgeworth expansion as well.

3.3 Error Analysis

In this part, we will discuss sources of errors of CDG approximation and an indication of their magnitude.

There are two types of errors originated from two sources. The first type is from the assumption of $log(G(k)) = \sum_{j=1}^{\infty} c_j \frac{(ik)^j}{j!}$, for any $t \in (-\infty, +\infty)$ as we talked previously. The second type of errors is truncation errors.

To understand the origin of the second type of errors, in (21)

$$f_t(y) = (\frac{1}{2\pi}) \int_{-\infty}^{+\infty} e^{-iky} e^{ikc_1 - \frac{k^2}{2}c_2} e^{\Lambda} dk$$

where $\Lambda = \sum_{j=3}^{\infty} c_j \frac{(ik)^j}{j!}$, we approximate

$$\Lambda \approx \sum_{j=3}^{7} c_j \frac{(ik)^j}{j!}$$

$$e^{\Lambda} = \sum_{n=0}^{\infty} \frac{\Lambda^n}{n!} \approx 1 + \Lambda + \frac{\Lambda^2}{2}$$

and cut off terms with order higher than seven.

Collin-Dufresne and Goldstein propose an indication of magnitude for this truncation errors. Let $g_t(y)$ to be the approximation of the density function $f_t(y)$, we have

$$\begin{aligned} g_t(y) &= (\frac{1}{2\pi}) \int_{-\infty}^{+\infty} e^{-\frac{k^2}{2}c_2 - ik(y-c_1)} \times \\ &\quad [1 + (-\frac{ic_3}{3!}k^3 + \frac{c_4}{4!}k^4 + \frac{ic_5}{5!}k^5 - \frac{c_6}{6!}k^6 - \frac{ic_7}{7!}k^7) + \frac{1}{2}(-\frac{c_3^2}{(3!)^2}k^6 - \frac{2ic_3c_4}{3!4!}k^7)]dk \end{aligned}$$

Define $Z = \frac{y-c_1}{\sqrt{c_2}}$, then $g_t(y)$ can be written as:

$$\begin{aligned} g_t(Z) &= (\frac{1}{2\pi}) \int_{-\infty}^{+\infty} e^{-\frac{k^2}{2}c_2 - ikZ\sqrt{c_2}} \times [1 + (-\frac{ic_3}{3!}k^3 + \frac{c_4}{4!}k^4 + \frac{ic_5}{5!}k^5 - \frac{c_6}{6!}k^6 - \frac{ic_7}{7!}k^7) \\ &\quad + \frac{1}{2}(-\frac{c_3^2}{(3!)^2}k^6 - \frac{2ic_3c_4}{3!4!}k^7)]dk \\ &= (\frac{1}{2\pi}) \int_{-\infty}^{+\infty} e^{-l^2 - ilZ} \times [1 + (-\frac{ic_3}{3!c_2^{3/2}}l^3 + \frac{c_4}{4!c_2^{4/2}}l^4 + \frac{ic_5}{5!c_2^{5/2}}l^5 - \frac{c_6}{6!c_2^{6/2}}l^6 - \frac{ic_7}{7!c_2^{7/2}}l^7) \\ &\quad + \frac{1}{2}(-\frac{c_3^2}{(3!)^2c_2^{6/2}}l^6 - \frac{2ic_3c_4}{3!4!c_2^{7/2}}l^7)]dl \end{aligned} \tag{28}$$

where $l = \sqrt{c_2}k$. Compute the integration in (28) by formula from (24), we can get

$$g_t(Z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{Z^2}{2}} (1 + \sum_{i=3}^{7} \pi(i))$$

where

$$\pi(3) = (\frac{c_3}{3!c_2^{3/2}}) He_3(Z)$$

$$\pi(4) = (\frac{c_4}{4!c_2^{4/2}}) He_4(Z)$$

$$\pi(5) = (\frac{c_5}{5!c_2^{5/2}}) He_5(Z)$$

$$\pi(6) = (\frac{c_6}{6!c_2^{6/2}} + \frac{1}{2}(\frac{c_3}{3!c_2^{3/2}})^2) He_6(Z)$$

$$\pi(7) = (\frac{c_7}{7!c_2^{7/2}} + (\frac{c_3}{3!c_2^{3/2}})(\frac{c_4}{4!c_2^{4/2}}))He_7(Z)$$

in which $He_i(Z)$ is the i-th probabilists' Hermite polynomials.

Definition (Scaled Cumulant). Define $\frac{c_i}{i!c_2^{i/2}}$ as the i-th scaled cumulant, where i is a positive integer and $i > 2$.

Dufresne and Goldstein imply that from $\pi(3)$ to $\pi(7)$, it can be concluded that scaled cumulants can be taken as an indication of accuracy of CDG approximation, since they indicate truncation errors: The faster they decay, the better the approximation is. However, they do not provide more explanation. We will discuss it more specifically as follows.

Consider the Edgeworth expansion for $f_t(Z)$, which is similar to (28), we can find that $f_t(Z)$ and $g_t(Z)$ share the same term $\pi(i)$, $i \leq 7$. So, $f_t(Z)$ can be written as

$$f_t(Z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{Z^2}{2}} (1 + \sum_{i=3}^{\infty} \pi(i))$$

where the order of each term in each $\pi(i)$ is i.[6]

Define $er_t(Z) = f_t(Z) - g_t(Z)$ to be the truncation error of CDG approximation, we have

$$er_t(Z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{Z^2}{2}} (\sum_{i=7}^{\infty} \pi(i)) \tag{29}$$

Generally, it seems that if scaled cumulants $\frac{c_i}{i!c_2^{i/2}}$ decay fast with i increases, where decay means convergence to zero, then $\pi(i)$ will also decrease fast in absolute value. In this situation, the truncation error $er_t(Z)$ will be small compared with the true density $f_t(Z)$. This implies scaled cumulants can be an indication of accuracy of the truncation error. Therefore, Collin-Dufresne and Goldstein propose to take scaled cumulants as an indication of the error of CDG approximation.

Compared with Jarrow and Rudd (1982), Turnbull and Wakeman (1991) and Tanaka (2005), all of which seem not provide any error estimation, CDG approximation is superior in this way. However, Dufresne and Goldstein's error estimation is neither precise nor mathematically solid.

To be more specific, firstly, "scaled cumulants decay fast" is not mathematically precise. Second, there is neither theoretically rigorous proof nor convincing experiments to show that the faster scaled cumulants decay, the more accurate the CDG approximation is. In fact, it seems not possible to get a theoretical bound of error because CDG approximation is generally not asymptotic. Even it is asymptotic on some special cases, we only have seven cumulants and thus still need experiments to see how large the truncation error is.

Therefore, we have to analyze errors by experiments and my work is to try to develop an accurate measure of errors of swaption price based on Collin-Dufresne and Goldstein's

[6] If a term is $\frac{c_i}{i!c_2^{i/2}}$, we define the order to be i. For a term $\frac{c_i}{i!c_2^{i/2}} \frac{c_j}{j!c_2^{j/2}}$, the order is defined as $i + j$.

indication of errors, and see whether it is accurate in experiments. In addition, we will see the performance of CDG approximation on several probability distributions.

Measure of Error

Dufresne and Goldstein do not provide a measure to measure how fast the scaled cumulants decay. Instead, I define a measure l as follows

$$l = log|\frac{sc_7}{sc_3}| \tag{30}$$

where sc_i is the i-th scaled cumulant, and $|\frac{sc_7}{sc_3}|$ is the absolute value of $\frac{sc_7}{sc_3}$.

We choose $|\frac{sc_7}{sc_3}|$ because $|\frac{sc_7}{sc_3}|$ is a combination of all five scaled cumulants.

$$|\frac{sc_7}{sc_3}| = |\frac{sc_4}{sc_3}| * |\frac{sc_5}{sc_4}| * |\frac{sc_6}{sc_5}| * |\frac{sc_7}{sc_6}|$$

It is clear $|\frac{sc_7}{sc_3}|$ can indicate the speed that scaled cumulants decay: the faster they decay, the smaller $|\frac{sc_7}{sc_3}|$ is. So, it corresponds with Dufresne and Goldstein's error estimation, but more mathematically precise. In addition, we prefer $log|\frac{sc_7}{sc_3}|$ over $|\frac{sc_7}{sc_3}|$ for visual convenience, since $|\frac{sc_7}{sc_3}|$ might be very too small, for instance 0.0001. Obviously, smaller $log|\frac{sc_7}{sc_3}|$ also means faster decay of scaled cumulants.

Of course, there are many other measures that can also describe the speed that scaled cumulants decay, such as $|\frac{sc_4}{sc_3}|$. Later experiments show that l is an accurate measure of errors of densities and distribution functions, and thus on the rest of my thesis, we will use this measure l to measure the speed of scaled cumulants decay.

3.4 Approximation Performance

In this part, we will carry out experiments to see the overall performance of CDG approximation on different probability densities, such as Normal, Gamma and Log-normal densities. Besides, our measure l will be evaluated to see whether it is accurate in estimating errors.

Normal Distribution

Let a random variable X to be normally distributed, that is, $X \sim N(\mu, \sigma)$, we can write $X = U + \sigma\mu$, where U has a standard Normal distribution.

It is easy to get its cumulants directly from its characteristic function instead of calculating them by moments. Since its characteristic function is

$$E(e^{itx}) = E(e^{it(u+\mu)}) = e^{it\mu - \sigma^2 t^2/2}$$

then, we have

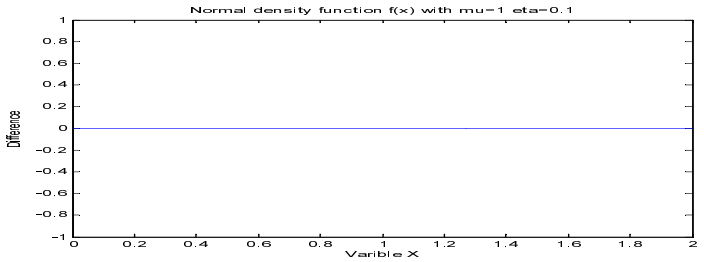
$$log(E(e^{itx})) = it\mu - \sigma^2 t^2/2 \tag{31}$$

It shows that the first cumulant of a Normal distribution is μ, and the second is σ^2, while other cumulants of order higher than two are all zeros. Hence, there is no truncation error in CDG approximation. Besides, we already know the radius $r = \infty$ in Normal distributions,

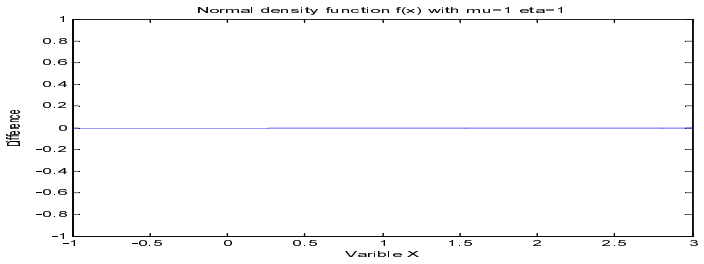
in other words, (31) holds for any $t \in (-\infty, +\infty)$. Therefore, there should be absolutely no errors between the CDG approximated density function of a Normal distribution and the true density function.

In our next few experiments, we will choose several Normal density functions with different parameters, and compare the CDG approximated densities with the true ones. Experiment results are displayed as follows.

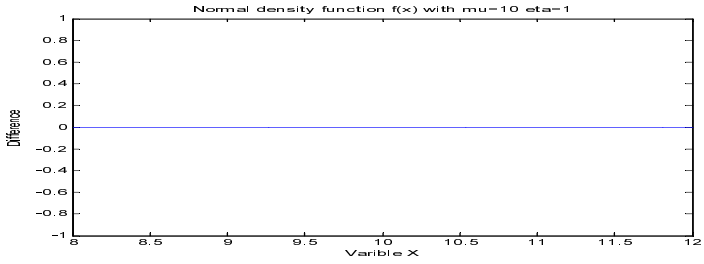
Example 1: $\mu = 1$, $\sigma = 0.1$



Example 2: $\mu = 1$, $\sigma = 1$



Example 3: $\mu = 10$, $\sigma = 1$



Conclusion:

As what we expected, in Normal distributions, the approximation errors of density functions are always zero no matter what values of parameters are. In addition, obviously, if we compare CDG approximated Normal distribution functions with the true ones, there is also no error.

By intuition, it seems that a random variable with distribution close to Normal may tend to have good CDG approximation performance. In fact, this is one of the nice property of Edgeworth expansion. In the following part, we will test this rule of CDG approximation on Gamma and Lognormal distributions.

Gamma Distribution

Suppose X is Gamma distributed, or, $X \sim \Gamma(\alpha, \beta)$ where $\alpha > 0$, $\beta > 0$, then its density function is

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$$

where $x \geqslant 0$, and $\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy$

It is not difficult to compute the n-th moment of X

$$\begin{aligned} E(x^n) &= \int_{-\infty}^{\infty} x^n f(x) dx \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_{-\infty}^{\infty} x^{n+\alpha-1} e^{-\beta x} dx \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_{-\infty}^{\infty} \frac{1}{\beta^{n+\alpha}} y^{n+\alpha-1} e^{-y} dy \\ &= \frac{1}{\beta^n \Gamma(\alpha)} \int_{-\infty}^{\infty} y^{n+\alpha-1} e^{-y} dy \\ &= \frac{\Gamma(n+\alpha)}{\beta^n \Gamma(\alpha)} \end{aligned} \tag{32}$$

As $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$ and $\Gamma(1) = 1$, so if α is a positive integer, then we have $\Gamma(\alpha) = (\alpha-1)!$ It follows that

$$E(x^n) = \frac{1}{\beta^n}(n+\alpha-1)(n+\alpha-2)...\alpha$$

The cumulants are calculated by moments, with formulas of their relationship showed in Appendix A. Besides, by Theorem 3.2, $\sum \mu_{2n} t^n/(2n)!$ converges in some interval $(-t_0, t_0)$, where

$$t_0 = \lim_{n\to\infty} \frac{(2n+2)!\mu_{2n}}{(2n)!\mu_{2n+2}} = \beta^2 > 0$$

Therefore, moments of a Gamma distribution uniquely determine its distribution.

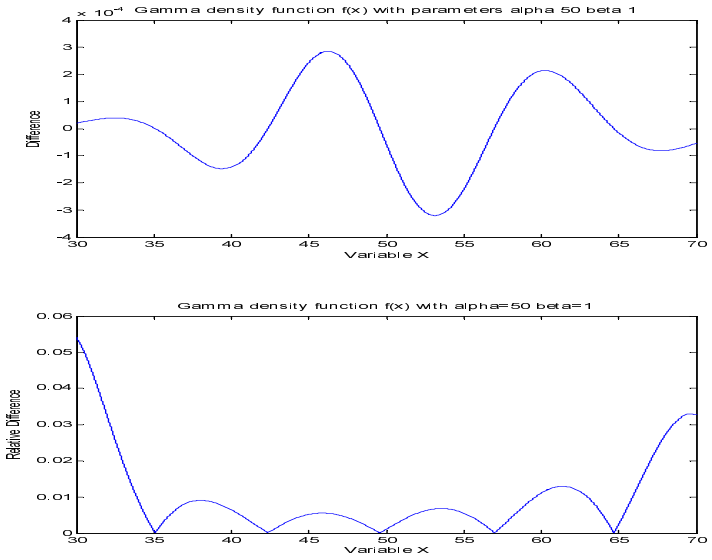
Next, we will see CDG approximation performance on Gamma distributions by Matlab experiments. In those experiments, in order to make a comparison between true densities of Gamma distributions and their approximated density functions, we calculate their errors in two ways, difference and relative difference.[7] Note that the performance of CDG approximation is mainly showed by relative differences other than differences, and the reason behind is that the swaption price for approximation is a combination of several probabilities,

[7] Let $f(x)$ be the true density function, and $g(x)$ is its approximation by CDG approximation. The difference is defined as $f(x) - g(x)$, and relative difference is $|f(x) - g(x)|/f(x)$.

with approximation accuracy determined more by relative differences of densities. Usually, in Statistics, small relative errors of densities would result in small errors of probabilities, both absolute errors and relative errors. Therefore, we will focus more on their relative differences.

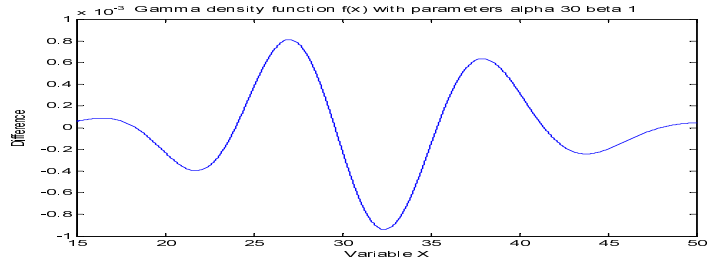
In addition, we will only plot in the domain of variable X where the true density is not almost zero. In fact, we can find that in this almost-zero-density domain, the absolute error is also very close to zero, although there might be large relative errors due to too small true density of X. We make this setting is to avoid large relative errors on the tail. The results are illustrated as follows.

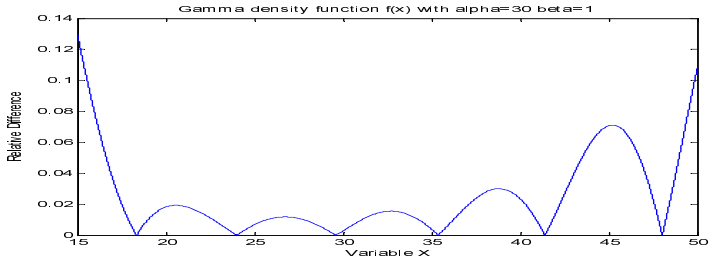
Example 1: $\alpha = 50$, $\beta = 1$



sc = [0.0471 0.0050 0.0006 0.0001 0.0000], $l = -8.6713$
where sc_i is the i-th scaled cumulant.

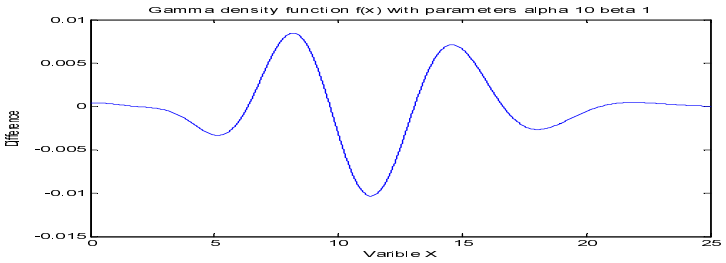
Example 2: $\alpha = 30$, $\beta = 1$

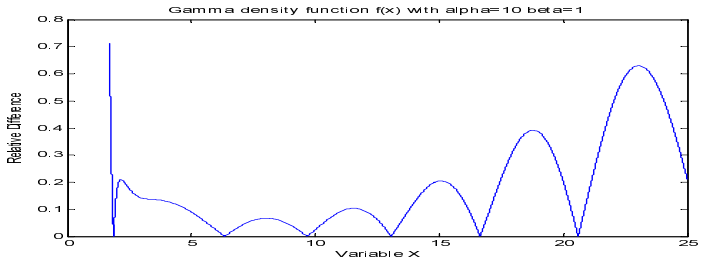




sc = [0.0609 0.0083 0.0012 0.0002 0.0000], $l = -7.6497$

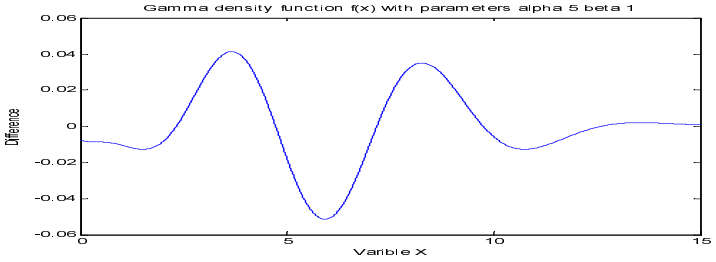
Example 3: $\alpha = 10$, $\beta = 1$

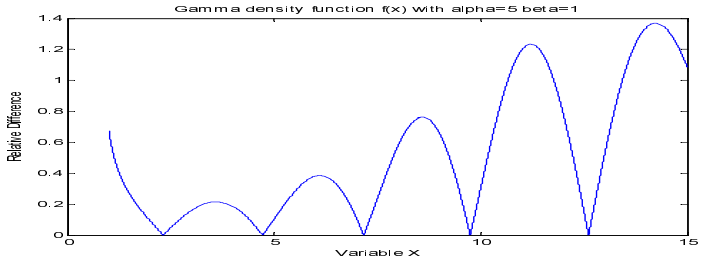




sc = [0.1054 0.0250 0.0063 0.0017 0.0005] , $l = -5.4525$

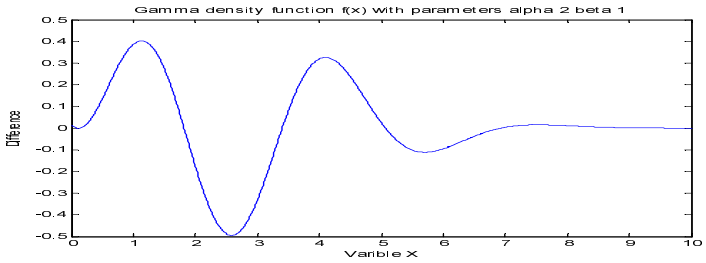
Example 4: $\alpha = 5$, $\beta = 1$

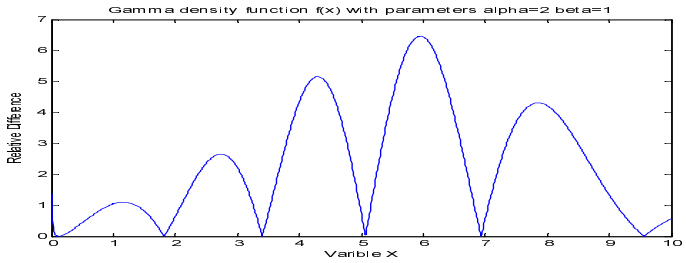




sc = [0.1491 0.0500 0.0179 0.0067 0.0026], $l = -4.0662$

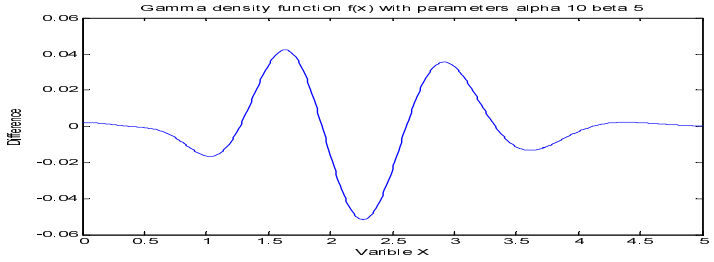
Example 5: $\alpha = 2$, $\beta = 1$

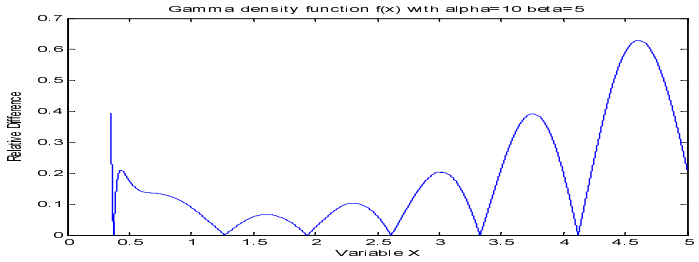




sc = [0.2357 0.1250 0.0707 0.0417 0.0253], $l = -2.2336$

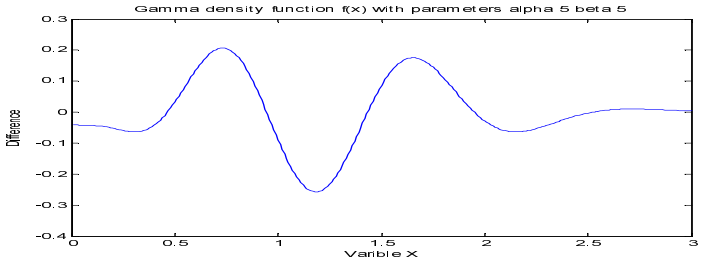
Example 6: $\alpha = 10$, $\beta = 5$

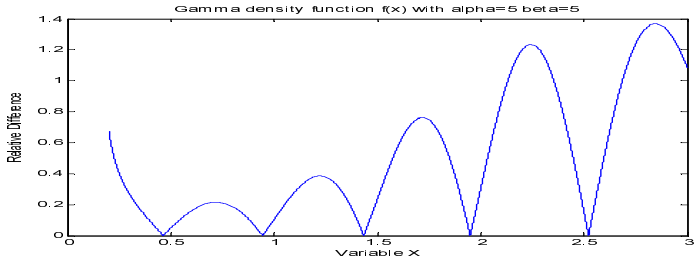




sc = [0.1054 0.0250 0.0063 0.0017 0.0005], $l = -5.4525$

Example 7: $\alpha = 5$, $\beta = 5$





sc = [0.1491 0.0500 0.0179 0.0067 0.0026], $l = -4.0662$

Conclusion:

1. The faster the scaled cumulants decay, and the more accurate the approximation is, see from the first five examples. Example 1 shows the most accurate result with relative difference about 0.01 and the fastest decay of scaled cumulants with $l = -8.6713$, followed by Example 2 with relative difference around 0.02 and $l = -7.6497$, then Example 3, with relative difference around 0.2 and $l = -5.4525$. The approximation performance in Example 5 is worst with relative difference around 1, due to rather slow decay of scaled cumulants with $l = -2.2336$. In other words, the smaller l is, the more accurate CDG approximation is. It shows that our measure l is accurate in cases of Gamma distributions. Besides, the accuracy becomes less sensitive to l as it decreases. For instance, while l decreases from -5.4525 to -7.6497, relative error

decreases rapidly from 0.2 to 0.02. However, when l decreases from -7.6497 to -8.6713, relative error only decreases from 0.02 to 0.01.

2. With respect to parameter influence on density approximation performance, the larger the α is, the more accurate the approximation is, while β has little affection on it. By comparing Example 3 and Example 6, or Example 4 and Example 7, in any pair of which, α keeps the same, only β differs, we can find that the scaled cumulants are exactly the same in any pair, and the relative errors are also at the same level. The property that β has no affection on scaled cumulants can be proved theoretically, see Appendix D.

Previously, we indicate that a random variable with distribution close to Normal tends to have good approximation performance. Then, we will study when a Gamma distribution looks like a Normal distribution.

We know that while $\alpha = 1$, a Gamma distribution is an Exponential distribution, that is, $\Gamma(1, \beta) = Exp(\beta)$. Besides, a random variable with Gamma distribution $\Gamma(\alpha, \beta)$ is the sum of α iid random variables distributed with $Exp(\beta)$ if α is an integer. According to Central Limit Theorem, if α is sufficiently large, $\Gamma(\alpha, \beta)$ can be approximated by a Normal distribution. This also implies that larger α is, the closer a Gamma distribution is to a Normal distribution.

In addition, experiments show that the larger the α is, the more accurate the approximation is, thus it is true that a Gamma distribution close to Normal will have good CDG approximation result.

Lognormal Distribution

Suppose random variable X is lognormal distributed with parameters μ and σ, then its density function is:

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(Inx-\mu)^2}{2\sigma^2}}, x > 0$$

Commonly, it is more convenient to write $X = e^{\mu+\sigma U}$, where random variable U is standard Normal distributed.

It is well known that the expectation of X is

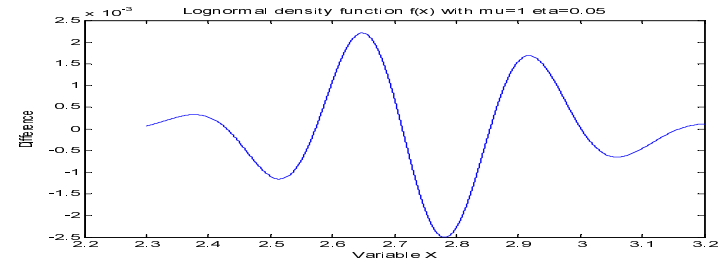
$$E(x) = E(e^{\mu+\sigma u}) = e^{\mu+\sigma^2/2}$$

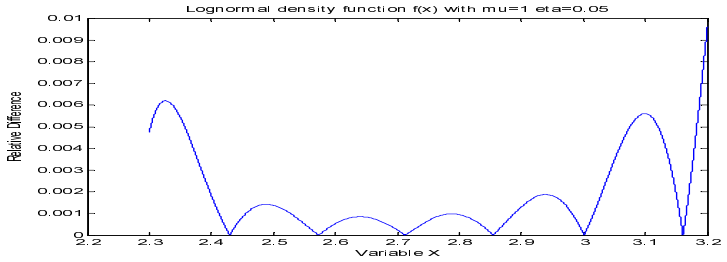
Thus, the n-th moment of a lognormal distribution is:

$$E(x^n) = E(e^{n\mu+n\sigma u}) = e^{n\mu+\sigma^2 n^2/2}$$

In the following part, Matlab experiments will be carried out to see performance of CDG approximation on density functions of Lognormal distributions. The experimental results are displayed below.

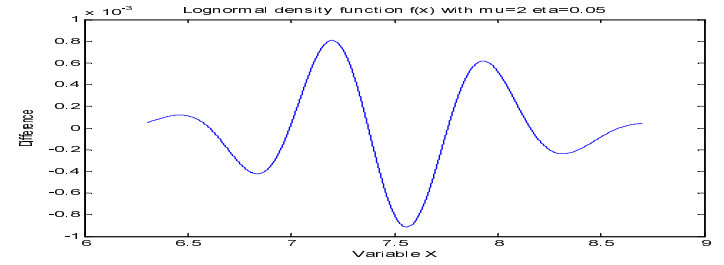
Example 1: $\mu = 1$, $\sigma = 0.05$

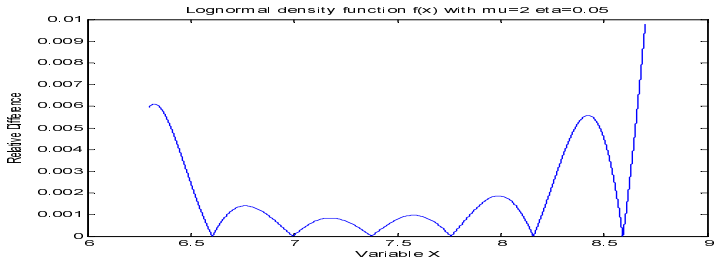




sc $= [0.0250\ 0.0017\ 0.0001\ 0.0000\ 0.0000]$, $l = -9.9934$

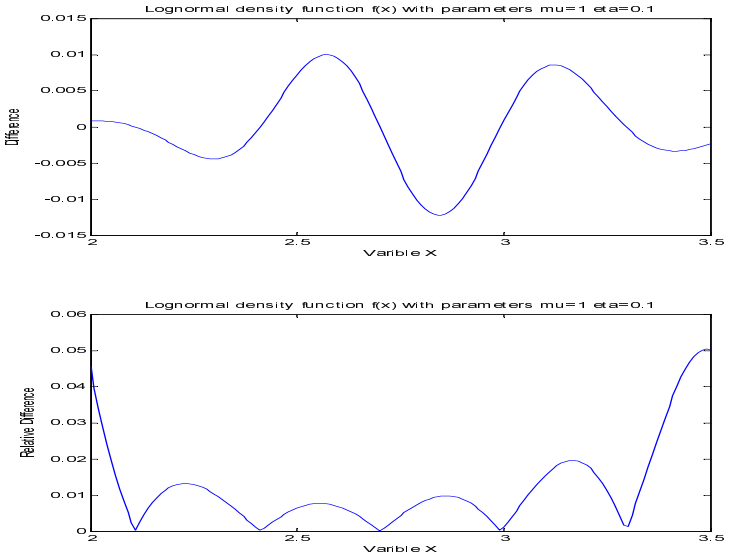
Example 2: $\mu = 2$, $\sigma = 0.05$





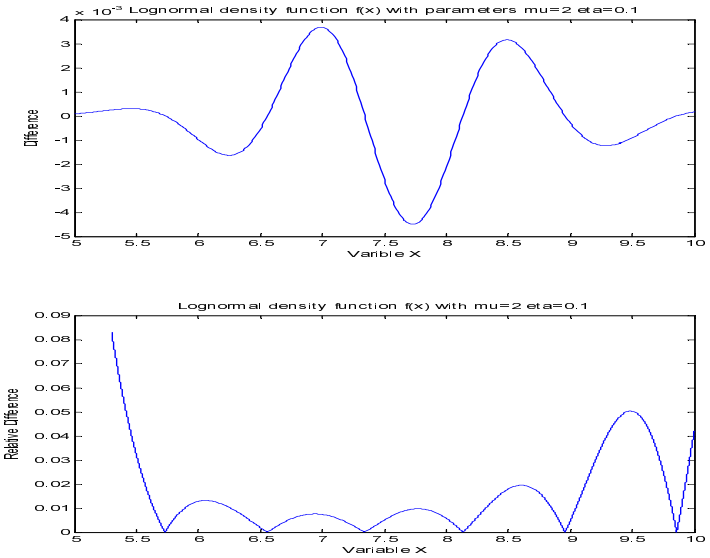
sc $= [0.0250\ 0.0017\ 0.0001\ 0.0000\ 0.0000]$, $l = -10.1596$

Example 3: $\mu = 1$, $\sigma = 0.1$



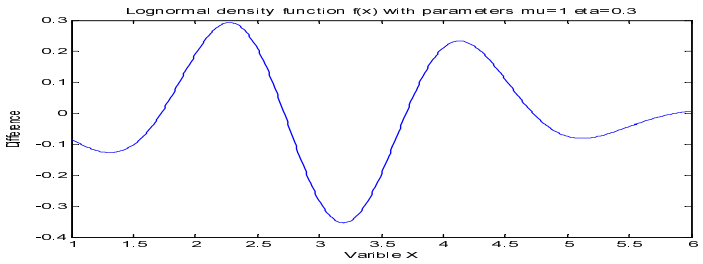
sc $= [0.0503\ 0.0068\ 0.0011\ 0.0002\ 0.0000]$, $l = -7.2652$

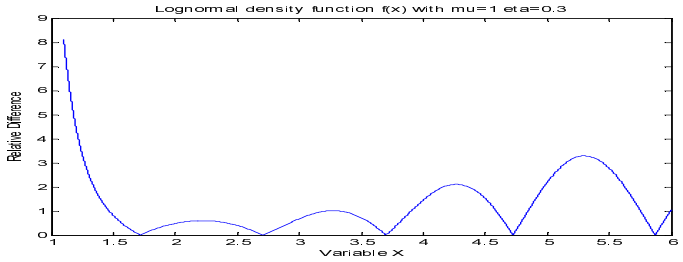
Example 4: $\mu = 2$, $\sigma = 0.1$



sc $= [0.0503\ 0.0068\ 0.0011\ 0.0002\ 0.0000]$, $l = -7.2652$

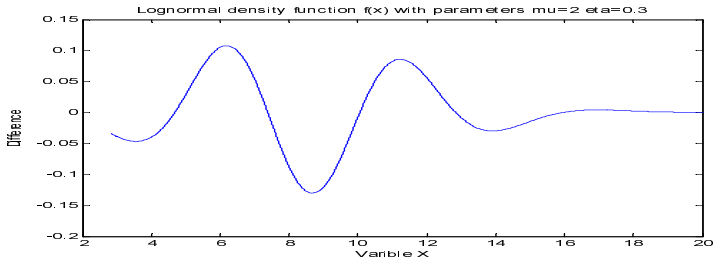
Example 5: $\mu = 1$, $\sigma = 0.3$

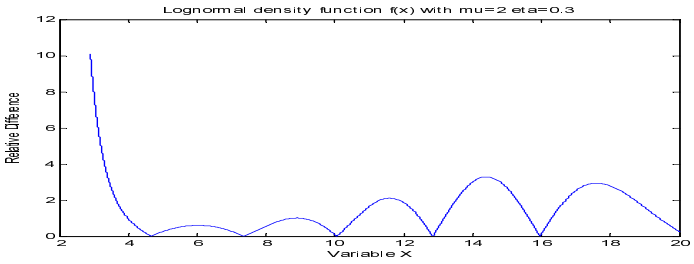




sc $= [0.1583\ 0.0685\ 0.0356\ 0.0209\ 0.0134]$, $l = -2.4667$

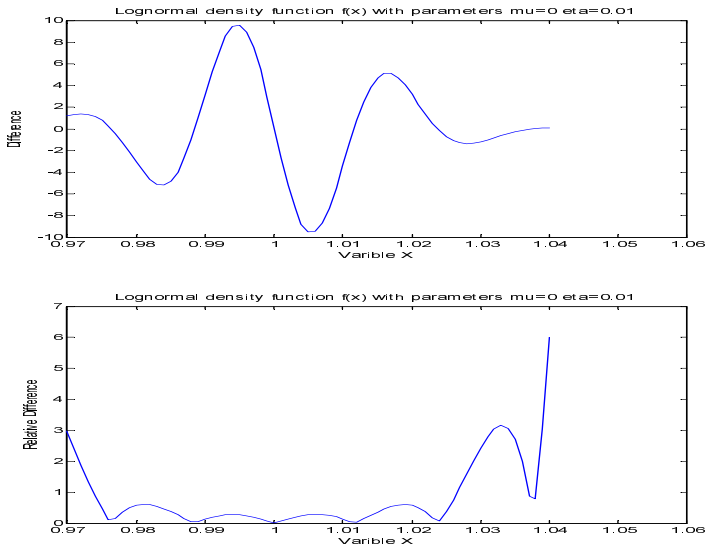
Example 6: $\mu = 2$, $\sigma = 0.3$





sc $= [0.1583\ 0.0685\ 0.0356\ 0.0209\ 0.0134]$, $l = -2.4667$

Example 7: $\mu = 0$, $\sigma = 0.01$



sc = [0.0050 0.0001 0.0000 -0.0000 0.0068], $l = 0.3020$

Conclusion:

1. Generally, the faster the scaled cumulants decay, or the smaller the measure l is, the more accurate the approximation is. We can see from Example 1 to Example 7, the speed of decay of scaled cumulants becomes slower with measure l increases, and the corresponding density approximation becomes less accurate. In Example 1, the relative difference is about 0.001 with $l = -9.9934$, and in Example 3, the relative error is around 0.01 with $l = -7.2652$. While in Example 5, the relative difference turns into around 1 with $l = -2.4667$. It implies that our measure l is also accurate in cases of Lognormal distributions, as in Gamma distributions.

2. The accuracy of CDG approximation is determined majorly by parameter σ. In general, the smaller σ is, the more accurate the approximation is. On the other hand, μ affects little on scaled cumulants, and thus influences little on CDG approximation. For instance, by comparing Example 1 and Example 2, in which $\sigma = 0.05$, only μ differs, we can only observe a slight difference in measure l, while their relative errors are almost the same. The reason behind is that it is σ instead of μ that makes a Lognormal distribution look like Normal. However, we can learn from Example 7 that too small σ might lead to bad approximation.

Remark: Too small σ tends to make the second cumulant $c_2 = E(X^2) - E(X)^2$ quite small, where $E(x^n) = E(e^{n\mu+n\sigma u}) = e^{n\mu+\sigma^2 n^2/2}$. It is clear that if c_2 is tiny, it is quite possible that the scaled cumulants of higher order will be large, since the n-th scaled cumulant is $\frac{c_n}{n! c_2^{n/2}}$ by definition. For instance, in Example 7, $sc_7 = 0.0068$, which is even larger than $sc_3 = 0.0050$. Thus, σ cannot be too small if we expect an accurate approximation. In figure 3.1, we plot the curve that describes the relationship between σ, standard derivation of $log(X)$ and measure l, when $\mu = 0$.

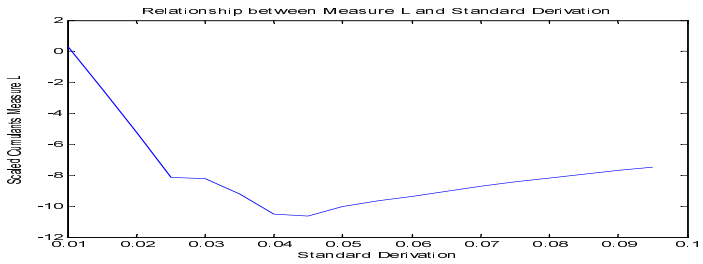


Figure 3.1. Relationship Between Measure l and σ

We will explore when a Lognormal distribution looks like Normal. Let X be a random variable of Lognormal distribution, we can write $X = e^{\mu+\sigma U}$, where U is standard Normal distributed. By Taylor expansion, we have $X = e^{\mu}(1+\sigma U+\frac{\sigma^2}{2}U^2+...)$. If σ is small, then $X \approx e^{\mu}(1+\sigma U)$, which implies X looks like a Normal distribution. Generally, the smaller σ is, the better the CDG approximation is, and the closer X is to a Normal distribution. Thus, a lognormal distribution tends to have good approximation performance when it resembles a Normal distribution.

However, it does not necessary mean $\sigma = 0.01$ is closer to Normal than $\sigma = 0.05$ since the above explanation is based on the general situation. Anyway, we are more interested in researching when a Lognormal distribution has good CDG approximation than when it looks like Normal.

Conclusion

In the experiments of Normal, Gamma and Lognormal distributions, there is one thing in common: Our measure l is accurate in estimating the accuracy of CDG approximation within the same type of distributions. However, when the types of distributions are different, it is not always the case, although in general it is. For instance, in the second example of Gamma distributions, $l = -7.6497$, with relative difference around 0.02, while in the third example of Lognormal distributions, $l = -7.2652$, but with relative difference about 0.01, a little more accurate than the former one. This is reasonable, because measure l is a general accurate measure.

Another interesting finding is that if a distribution is close to Normal, it tends to have good approximation performance by CDG approximation.

In next chapter, we will use CDG approximation to price a swaption, which is a combination of several probabilities with more complicated distributions.

Remark: Although we discussed CDG approximation on densities instead of on distribution functions, however, in Statistics, a good approximation on densities with small relative errors would usually lead to good approximation on their corresponding distribution functions also small in relative errors. Besides, since a distribution function is an increasing function from 0 to 1, it is natural that small relative errors usually bring about small absolute errors of distribution functions. This is true to CDG approximation, based on our experiment results on Gamma and Lognormal distribution functions (partly reported on Appendix C). On Appendix C, we plot errors of CDG approximation on several Lognormal distribution

functions. In addition, we illustrate in Appendix E that CDG approximation dominates over Normal approximation of density functions of random variables.

4 Swaption Pricing by CDG Approximation

In this chapter, we will use Collin-Dufrersne and Goldstein approximation to price swaptions, with the pricing formula (17), to see its approximation performance. Collin-Dufresne and Goldstein (2002) only show two accurate cases, one under Three-factor Gaussian model, and the other under CIR2++ model, but not analyze them in detail. Instead, we will provide more specific analysis and intensive insight. We assume the interest rate in formula (17) follows some affine model with number of factors more than one.

4.1 Affine Term Structure Interest Rate Models

An affine interest rate model is the model where its short-term interest rate is a stochastic process and an affine function of factors. Term structure means the parameters of this interest rate model are calibrated from the yield curve available from the real market, where the yield curve is the relation between long term interest rate and the time to maturity.

Let $x_1(t)$, $x_2(t)$,...$x_n(t)$ be factors of an affine model, satisfying certain stochastic differential equations and boundary conditions.

$$dx_i(t) = \beta_i(t, x_1, x_2, ...x_n)dt + \gamma_i(t, x_1, x_2, ...x_n)d\widetilde{w}(t)$$

for any $1 \leq i \leq n$. Let $r(t)$ be the interest rate at time t, an affine interest rate model assumes

$$r(t) = \delta_0(t) + \sum_{i=1}^{n} \delta_i(t)x_i(t)$$

where $\delta_i(t)$ is a deterministic function for any i.

The above is the general n-factor model. The simplest type is a one-factor model, which only contains one factor, and it is more convenient to write it in this form

$$dr(t) = \beta(t, r(t))dt + \gamma(t, r(t))d\widetilde{w}(t)$$

There are many kinds of affine interest rate models, and the most common ones are Hull-White model and CIR model.

Hull-White model

In the Hull-White model, described by John Hull and Allen White in 1990, the evolution of the interest rate is given by

$$dr(u) = (a(u) - b(u)r(u))du + \sigma(u)d\widetilde{w}(u) \tag{33}$$

where $a(u)$, $b(u)$, and $\sigma(u)$ are nonrandom positive functions of time. $\sigma(u)$ is called the volatility.

Equation (33) with the initial condition $r(t) = r_0$ has an analytical solution.

$$r(T) = e^{-\int_t^T b(v)dv}r_0 + \int_t^T e^{-\int_u^T b(v)dv}a(u)du + \int_t^T e^{-\int_u^T b(v)dv}\sigma(u)d\widetilde{w}(u) \tag{34}$$

It is obvious that $r(T)$ is Normal distributed, and its variance can be calculated by Ito isometry

$$\widetilde{E}(\int_t^T \Delta(u)d\widetilde{W}(u))^2 = \widetilde{E}(\int_t^T \Delta^2(u)du)$$

where $\Delta(u) = e^{-\int_u^T b(v)dv}\sigma(u)$ in our calculation. Then, we have

$$r(T) \sim N(e^{-\int_t^T b(v)dv}r + \int_t^T e^{-\int_u^T b(v)dv}a(u)du, \int_t^T e^{-2\int_u^T b(v)dv}\sigma^2(u)du)$$

When parameters a, b and σ are all constants, we can get

$$r(T) \sim N(e^{-b(T-t)}r_0 + \frac{a}{b}(1-e^{-b(T-t)}), \frac{\sigma^2}{2b}(1-e^{-2b(T-t)})) \tag{35}$$

In this situation, Hull-White model is called Vasicek model. In long term, when T goes to infinity, we have $\lim_{T\to\infty} r(T) \sim N(\frac{a}{b}, \frac{\sigma^2}{2b})$, which does not rely on the initial interest rate r_0. In literatures, $\frac{a}{b}$ is called the long term mean level, and $\frac{\sigma^2}{2b}$ called long term variance.

Hull-White model is frequently used in financial world, due to its advantage that the interest rate has Normal distribution with expectation and variance known. However, there is a positive probability that the interest rate is negative in the model, and this is one of the principal objections to the Hull-White model. Anyway, if we model a net interest rate, Hull-White model may still apply, because a net interest rate allows negative value, which is defined to the interest rate after eliminating the effect of inflation.

CIR model

CIR model, short for Cox-Ingersoll-Ross interest rate model, was introduced in 1985 by John Cox, Jonathan Ingersoll and Stephen Ross as an extension of the Vasicek model. In this model, the interest rate is given by the stochastic differential equation

$$dr(u) = (a - br(u))du + \sigma\sqrt{r(u)}d\widetilde{w}(u) \tag{36}$$

where a, b, and σ are all positive constants.

Given an initial condition $r(t) = r_0$, the interest rate $r(T)$ follows a non-central chi-squared distribution with distribution function

$$F_t(x) = \chi^2(c_T x, v, \lambda_T)$$

where

$$c_T = \frac{4b}{\sigma^2(1-exp(-b(T-t)))}$$

$$v = 4a/\sigma^2$$

$$\lambda_T = c_T r_0 exp(-b(T-t)))$$

and $\chi^2(x, v, \lambda)$ is the non-central chi-squared distribution with degree of freedom v and non-centrality parameter λ.

Besides, many of its properties can be determined. For instance, the mean and variance of $r(T)$ conditioned on initial value $r(t) = r_0$.

$$\begin{aligned}
\widetilde{E}_t(r(T)) &= e^{-b(T-t)}r_0 + \frac{a}{b}(1 - e^{-b(T-t)}) \\
\widetilde{Var}_t(r(T)) &= \frac{\sigma^2}{b}r_0(e^{-b(T-t)} - e^{-2b(T-t)}) + \frac{a\sigma^2}{2b^2}(1 - 2e^{-b(T-t)} + e^{-2b(T-t)})
\end{aligned}$$

It follows the long term expectation and variance

$$\begin{aligned}
\lim_{T\to\infty} \widetilde{E}_t(r(T)) &= \frac{a}{b} \\
\lim_{T\to\infty} \widetilde{Var}_t(r(T)) &= \frac{a\sigma^2}{2b^2}
\end{aligned}$$

Unlike the interest rate in Hull-White model, the interest rate in the CIR model can never be negative. This is because when the interest rate approaches zero, the term $\sigma\sqrt{r(u)}d\widetilde{w}(u)$ also approaches zero. With the volatility disappearing, the behavior of the interest rate near zero depends on the drift term $a - br(u)$, and this is $a > 0$ when $r(u) = 0$. The positive drift prevents the interest rate from crossing zero into negative territory. Besides, we have Feller Condition that tells whether $r(u)$ will hits zero or not.

Feller Condition

Feller condition says CIR process $r(u)$ will never hits zero, if and only if $a \geq \frac{1}{2}\sigma^2$. On the other hand, if $0 < a < \frac{1}{2}\sigma^2$, $r(u)$ hits zero repeatedly but after each hit it becomes positive again. Under some circumstances, it is convenient to approximation $r(u)$ as Normal if it never hits zero. However, if it hits zero, in general, it can not be regarded as Normal.

Opposite to Hull-White model, the advantage of CIR model is that the interest rate is always non-negative, but there is no explicit formula for its solution [8], and thus there might be some inconvenience in computation.

4.2 CDG Approximation under Three-factor Gaussian Model

In this part, we will make a specific description of CDG approximation under three-factor Gaussian model with factor dynamics as follows

$$dx_i(u) = -\alpha_i x_i(u)du + \sigma_i d\widetilde{w}_i(u) \tag{37}$$

and

[8] Strickly speaking, non-central chi-squared distribution function is not in closed form by defination because it is sum of infinite number of functions. Thus, Shreve (2004) says 'there is no explicit formula'. However, in some literatures, for instance, Brigo and Mercurio (2006), the solution of CIR is considered to be analytical.

$$d\widetilde{w}_i(u)d\widetilde{w}_j(u) = \rho_{ij}du$$

$$r(u) = \delta_0 + \sum_{i=1}^{3} x_i(u) \tag{38}$$

where α_i, σ_i for any $i = 1, 2, 3$ and δ_0 are all positive constants.

Three-factor Gaussian model can actually be proved to be equivalent to three-factor Hull-White model with parameters all constants. However, the formulation with Gaussian model leads to less complicated formulas and it is easier to implement in practice although we may lose some insight and intuition on the nature and the interpretation of the three factors. The outline of this section is as follows: first formulate the bond price under three-factor Gaussian model, then calculate the moments of coupon bond under forward measure.

Recall the zero coupon bond price $CB(T_0) = \sum_{i=1}^{n} C_i B(T_0, T_i)$ in chapter two, we have to generate the formula for each bond price $B(T_0, T_i)$ under three-factor Gaussian model and the following lemmas and theorems are necessary.

Lemma 4.2.1. If X is a lognormal distributed variable, then its expectation is

$$E(X) = e^{\mu + \frac{1}{2}\sigma^2}$$

where μ and σ are the mean and standard variance of variable $logX$ respectively.

Then, we have the following lemma and theorems, see Brigo and Mercurio (2006).

Lemma 4.2.2. For each t, T, the random variable defined as

$$I(t,T) = \int_t^T [x_1(u) + x_2(u) + x_3(u)]du$$

conditioned on information at t is normally distributed with mean $M(t,T)$ and $V(t,T)$ given by

$$M(t,T) = B_{\alpha_1}(T-t)x_1(t) + B_{\alpha_2}(T-t)x_2(t) + B_{\alpha_3}(T-t)x_3(t)$$

and

$$V(t,T) = \sum_{i,j} \frac{\sigma_i \sigma_j \rho_{ij}}{\alpha_i \alpha_j}[(T-t) - B_{\alpha_i}(T-t) - B_{\alpha_j}(T-t) + B_{\alpha_{i+}\alpha_j}(T-t)]$$

where

$$B_{\alpha_i}(\tau) = \frac{1 - e^{-\alpha_i \tau}}{\alpha_i}$$

Proof. Similar to Hull-White model, in there-factor Gaussian model, processes $x_1(u)$, $x_2(u)$, $x_3(u)$ are all normally distributed. Thus, their sum $x_1(u) + x_2(u) + x_3(u)$ is also Normal. The proof of this lemma is straightforward.

■

Note that $V(t,T)$ and be written as $V(T-t)$ since it can be regarded as function of $T-t$.

Theorem 4.4.1. The price of at time t of a zero-coupon bond maturing at time T is

$$B(t,T) = e^{-A(T-t) - \sum_{i=1}^{3} C_i(T-t)x_i(t)}$$

where

$$C_i(\tau) = B_{\alpha_i}(\tau) \tag{39}$$

$$A(\tau) = \delta_0 \tau - \frac{1}{2} V(\tau) \tag{40}$$

Proof. Combine Lemma 4.2.1 and Lemma 4.2.2, and bond formula $B(t,T) = \widetilde{E_t}(e^{\int_t^T r(u)du})$, the proof is almost immediate.

■

We have discussed the T forward measure in the section 2.2, under which, dynamics of factors changes. We will state the following theorem without proof under three-factor Gaussian model .

Theorem 4.4.2. The processes $x_1(u)$, $x_2(u)$, $x_3(u)$ under the T forward measure evolve according to

$$dx_i(u) = (-\alpha_i x_i(u) - \sum_{j=1}^{3} \sigma_i \sigma_j \rho_{ij} \frac{1 - e^{-\alpha_j(T-u)}}{\alpha_j})du + \sigma_i d\tilde{w_i}^T(u)$$

for any $i = 1, 2, 3$. where the correlation of Brownian motions will not change. Equivalently, $d\tilde{w_i}^T(u) d\tilde{w_j}^T(u) = d\widetilde{w}_i(u) d\widetilde{w}_j(u) = \rho_{ij}$.

■

If the Brownian motion terms are mutually independent, the above equations have less complicated forms.

$$dx_i(u) = (-\alpha_i x_i(u) - \frac{1-e^{-\alpha_i(T-u)}}{\alpha_i}\sigma_i^2)du + \sigma_i d\tilde{w_i}^T(u)$$

Obviously, interest rate $x_i(u)$ still follows the general Hull-White model under T-forward measure, and thus it is also normally distributed at time $u = T_0$ given initial value $x_i(t) = x_0$

$$x_i(T_0) \sim N(e^{-\alpha_i(T_0-t)}x_0 - (1-e^{-\alpha_i(T_0-t)})\frac{\sigma_i^2}{\alpha_i^2} + \frac{\sigma_i^2}{2\alpha_i^2}e^{-\alpha_i(T-T_0)} - \frac{\sigma_i^2}{2\alpha_i^2}e^{-\alpha_i(T+T_0-2t)}, \frac{\sigma_i^2}{2\alpha_i}(1-e^{-2\alpha_i(T_0-t)})) \tag{41}$$

This can be used for Monte Carlo simulation in our next experiments. Besides, with this assumption of independence of Brownian motions, $V(\tau)$ can be simplified as

$$V(\tau) = \sum_{i=1}^{3} \frac{\sigma_i^2}{\alpha_i^2}[\tau - 2B_{\alpha_i}(\tau) + B_{2\alpha_i}(\tau)]$$

Then, we will turn to the formula for the moments. suppose F_i, $i = 1, 2, 3$ all to be constants, the expectation of products of zero-coupon bond prices at time t can be computed using Lemma 4.1.1, Theorem 4.1.1 and Theorem 4.1.1, under the T forward measure.

$$\tilde{E_t^T}(e^{-\sum_{i=1}^3 F_i x_i(T_0)}) = e^{M(T_0-t) - \sum_{i=1}^3 N_i(T_0-t)x_i(t)} \tag{42}$$

where

$$N_i(\tau) = F_i e^{-\alpha_i \tau}$$

$$\begin{aligned} M(\tau) &= \sum_{i,j} \frac{\sigma_i\sigma_j\rho_{ij}}{\alpha_j} F_i[B_{\alpha_i}(\tau) - e^{-\alpha_j(T-T_0)}B_{\alpha_i+\alpha_j}(\tau)] + \\ &\quad \sum_{i \geq j} \sigma_i\sigma_j\rho_{ij}F_iF_jB_{\alpha_i+\alpha_j}(\tau) \end{aligned}$$

If the Brownian motions are mutually independent, we have

$$\begin{aligned} M(\tau) &= \sum_{i=1}^{3} \frac{\sigma_i^2}{\alpha_i} F_i[B_{\alpha_i}(\tau) - e^{-\alpha_i(T-T_0)}B_{2\alpha_i}(\tau)] + \\ &\quad \frac{1}{2}\sum_{i=1}^{3} \sigma_i^2 F_i^2 B_{2\alpha_i}(\tau) \end{aligned}$$

This formula (42) above allows us to compute all the moments of a coupon bond under three-factor model. Together with the general swaption pricing formula (17), we are capable of obtaining the swaption price by CDG approximation.

Besides, in order to evaluate the accuracy of CDG approximation by experiments, we need the Monte Carlo swaption price, which is usually taken as the true value of swaption. In the following part, we will discuss variance reduction techniques for Monte Carlo simulation.

Standard Variance Reduction Techniques

From the swaption formula (18) in chapter two

$$Swn(t) = \sum_{i=1}^{n} C_i B(t,T_i)\widetilde{P}_t^{T_i}(CB(T_0) > K) - KB(t,T_0)\widetilde{P}_t^{T_0}(CB(T_0) > K)$$

we can conclude that the price of swaption is quite sensitive to the probability $\widetilde{P}_t^{T_n}(CB(T_0) > K)$ and $\widetilde{P}_t^{T_0}(CB(T_0) > K)$. Thus, it might be necessary to increase the efficiency of Monte Carlo simulation by reducing the variance of simulation estimator $CB(T_0)$. These methods are called Variance Reduction Techniques, and we will discuss two kinds of them: Control Variates and Antithetic Variates. See Glasserman (2003).

Control Variates

Suppose random variable Y is the simulation estimator, we take a variable X such that its expectation $E(X)$ is known. Define a variable

$$Z = Y - b(X - E(X))$$

where b is

$$b = \frac{Cov(X,Y)}{Var(X)}$$

Then, Z is unbiased control estimator of Y, or $E(Y) = E(Z)$, but $\frac{Var(Z)}{Var(Y)} = 1 - \rho_{XY}^2$, where ρ_{XY} is the correlation between X and Y. The idea of this variance reduction technique is to replace Y by Z for simulation. It is obvious the higher the correlation is in absolute value, the better the Variance Reduction Technique works.

In our case of swaption simulation, we have to calculate $\widetilde{P}_t^T(CB(T_0) > K) = \widetilde{E}_t^T(1_{CB(T_0)>K})$. Let $Y = 1_{(CB(T_0)>K)}$ and $X = \frac{CB(T_0)}{B(T_0,T)}$, so from Property 1.2 in chapter two, we have $E_t^T(X) = \frac{CB(t)}{B(t,T)}$. In addition, the parameter b above is estimated by

$$\hat{b} = \frac{\sum_{i=1}^{n}(X_i - \hat{X})(Y_i - \hat{Y})}{\sum_{i=1}^{n}(X_i - \hat{X})^2}$$

where $\hat{X}$ is the expectation of simulations of X, and $\hat{Y}$ is that of Y.

In practice, we find that this technique is not quite useful for our swaption price. This is because by the nature of our target variable $Y = 1_{(CB(T_0)>K)}$, it is difficult to find a variable X, which has known expectation and high correlation with Y.

Antithetic Variates

Suppose we have to calculate the expectation $E(Y)$ of the random variable Y by Monte Carlo simulation, we can create another variable $\widetilde{Y}$ such that Y and $\widetilde{Y}$ have the same distribution. Define

$$Z = \frac{Y + \widetilde{Y}}{2}$$

Then, Z is the unbiased estimator of Y, because

$$E(Z) = \frac{1}{2}(E(Y) + E(\widetilde{Y})) = E(Y)$$

Besides,

$$\begin{aligned} Var(Z) &= \frac{1}{2}Var(Y) + \frac{1}{2}Var(Y)Cor(Y, \widetilde{Y}) \\ &\leq Var(Y) \end{aligned}$$

where $Cor(Y, \widetilde{Y})$ is the correlation between Y and $\widetilde{Y}$.

Obviously, the key idea of this method is to find a variable $\widetilde{Y}$ with the same distribution as Y, but their correlation should be as negatively small as possible. In our swaption pricing, we sample x_i from state variables $X_i(T_0)$, $i = 1, 2, 3$, whose distribution is Normal. Then, use x_1,x_2, x_3 to calculate Monte Carlo price of the underlying coupon bond $CB(T_0)$, set $Y = 1_{(CB(T_0)>K)}$. Let $\tilde{x}_i$ to be symmetric to x_i, or

$$\tilde{x}_i = 2\mu_{x_i} - x_i$$

where μ_{x_i} is the expectation of $X_i(T_0)$. By $\tilde{x}_1$,$\tilde{x}_2$, $\tilde{x}_3$, we have another price of coupon bond, $\tilde{CB}(T_0)$, and set $\widetilde{Y} = 1_{(\tilde{CB}(T_0)>K)}$. Clearly, Y and $\widetilde{Y}$ have the same distribution, because the densities of the state variable $X_i(T_0)$ at x_i and at $\tilde{x}_i$ are exactly the same. In addition, we can see that the correlation between Y and $\widetilde{Y}$ is negative.

In our experiments, this variance reduction technique is quite powerful compared with Monte Carlo simulation without any variance reduction. With this method, if we simulate 5 million samples for each state variable by Monte Carlo, then the standard deviation of the swaption prices by Monte Carlo is less than 1×10^{-5}, with order $o(10^{-7})$. Without this technique, the standard error of swaption price will be more than 1×10^{-4}. In Schrager and Pelsser (2006), they simulate 500,000 samples, less number than our samples, for each state variable to price swaptions the same as ours.

4.3 Numerical Results Under Three-factor Gaussian Model

In this section, our attention will be paid to the performance of CDG approximation for swaption pricing under Three-factor Gaussian model, and the accuracy of a measure for its error. Collin-Dufresne and Goldstein (2002) only show an accurate case, but not analyze it in detail. Instead, we will provide more specific analysis and intensive insight.

We will start by experiment 4.1, which prices a $T_0 * (T_2 - T_0)$ receiver swaption at time $t = 0$, with maturity of swaption $T_0 = 1$ year, and tensor $\delta = 0.5$ year. Experiment results are displayed in four tables, from Table 4.1.1 to Table 4.1.4.

The parameters of swaption from (37) and (38) are given in Table 4.1.1, where x_i represents $x_i(t)$, the initial value of this factor. k is the swap rate. We set $t = 0$ so the price of swaption is current. For convenience, let all three factors to be mutually independent, so $\rho_{ij} = 0$ if $i \neq j$.

Table 4.1.2 compares the swaption price by CDG approximation, with that by Monte Carlo simulation. The Monte Carlo price is obtained using the exact Normal distribution of each factor $x_i(T_0)$ from (41). The number of simulation for each factor under each forward measure is all 5 million with Antithetic Variates, and we take values by Monte Carlo method to be the true values.

Table 4.1.3 illustrates the probabilities $P_t^{T_i}(CB(T_0) > K)$ under different forward measures T_1,..,T_n respectively, where the first row of this table indicates their forward measures.

At last, the scaled cumulants are showed in Table 4.1.4, with sc_i on the first row indicating the i-th scaled cumulant, and forward measures are displayed on the first column. The last column shows values of our measure of error l defined in (30), corresponded with probabilities under different forward measures.

Experiment 4.1

Consider a $T_0 * (T_2 - T_0)$ receiver swaption, with maturity of swaption $T_0 = 1$ year and tensor $\delta = 0.5$ year. The parameters are given below

α_1	α_2	α_3	σ_1	σ_2	σ_3	x_1	x_2	x_3	δ_0	k
1	0.7	0.5	0.06	0.05	0.05	0.02	0.02	0.01	0.03	0.05

Table 4.1.1 Parameters of One-Year Swaption under Hull-White model

Experiment Result:

Swaption Price by Monte Carlo	0.0208
Swaption Price by CDG Approximation	0.0207
Absolute Error	0.0001
Relative Error	0.59%

Table 4.1.2 Swaption Price Approximation

	T_0	T_1	T_2
Monte Carlo	0.5427	0.5539	0.5616
CDG	0.5415	0.5525	0.5603
Absolute Error	0.0012	0.0014	0.0013

Table 4.1.3 Probability Approximation under Different Forward Measures

	sc_3	sc_4	sc_5	sc_6	sc_7	Meaure
T_0	0.02384931	0.00151864	0.00011341	0.00000937	0.00000048	-10.82
T_1	0.02384938	0.00151865	0.00011341	0.00000937	0.00000059	-10.61
T_2	0.02384944	0.00151866	0.00011341	0.00000938	0.00000086	-10.23

Table 4.1.4 Scaled Cumulants under Different Forward Measures

Conclusion:

In experiment 4.1, CDG approximation has excellent results. The scaled cumulants, illustrated on Table 4.1.4, decay very fast, with measure of error less than -10. This leads to accurate probabilities approximation, displayed on Table 4.1.3, with absolute errors no more than 0.0014 under each forward measure. It is displayed on Table 4.1.2 that the absolute error of CDG approximation for swaption is 0.0001, and the relative error is 0.59% compared with swaption price by Monte Carlo.

Next, we will try to price a three-year swaption, extending the number of swap times from previous two to six, to see whether it is still possible to generate accurate result by CDG approximation.

Experiment 4.2

Consider an $T_0 * (T_6 - T_0)$ receiver swaption, with maturity of swaption $T_0 = 1$ year and tensor $\delta = 0.5$ year.

α_1	α_2	α_3	σ_1	σ_2	σ_3	x_1	x_2	x_3	δ_0	k
1	0.7	0.5	0.05	0.03	0.03	0.02	0.02	0.01	0.03	0.05

Table 4.2.1 Parameters of Five-Year Swaption under Hull-White model

Experiment Result:

Swaption Price by Monte Carlo	0.0405
Swaption Price by CDG Approximation	0.0402
Absolute Error	0.0003
Relative Error	0.67%

Table 4.2.2 Swaption Price Approximation

	T_0	T_1	T_2	T_3	T_4	T_5	T_6
Monte Carlo	0.7263	0.7327	0.7369	0.7398	0.7422	0.7436	0.7445
CDG	0.7227	0.7289	0.7332	0.7361	0.7382	0.7397	0.7408
Absolute Error	0.0035	0.0038	0.0037	0.0037	0.0040	0.0038	0.0037

Table 4.2.3 Probability Approximation under Different Forward Measures

	sc_3	sc_4	sc_5	sc_6	sc_7	Meaure
T_0	0.02737426	0.00200378	0.00017212	0.00001637	0.00000170	-9.68
T_1	0.02737447	0.00200381	0.00017212	0.00001636	0.00000173	-9.67
T_2	0.02737462	0.00200383	0.00017212	0.00001636	0.00000179	-9.63
T_3	0.02737473	0.00200384	0.00017212	0.00001637	0.00000161	-9.74
T_4	0.02737481	0.00200386	0.00017212	0.00001636	0.00000162	-9.73
T_5	0.02737486	0.00200386	0.00017213	0.00001637	0.00000175	-9.66
T_6	0.02737491	0.00200387	0.00017213	0.00001636	0.00000174	-9.67

Table 4.2.4 Scaled Cumulants under Different Forward Measures

Conclusion:

In experiment 4.2, the performance of CDG approximation is also excellent. The scaled cumulants, illustrated on Table 4.2.4, decay very fast, with measure of error on average about -9.7. This results in accurate probabilities approximation, displayed on Table 4.2.3, with absolute errors a little more than 0.0035 under each forward measure. It is displayed on Table 4.2.2 that the absolute error of CDG approximation for swaption is 0.0003, and the relative error is 0.67% compared with swaption price by Monte Carlo.

Actually, CDG approximation can be highly accurate no matter the number of swap times is small or large. Then, readers may be interested in how to choose right parameters to display excellent CDG approximation and we will discuss it next.

In the forward probability $P_t^T(CB(T_0) > K)$ that CDG method deal with, we have

$$CB(T_0) = \sum_{i=1}^{n} C_i B(T_0, T_i)$$

where $C_i = k\delta K, i = 2, 3, ...n-1; C_n = (k\delta + 1)K$. In real world, $k\delta K$ is usually much smaller than K, and thus C_n is much larger than other C_i, $i = 1, 2, ...n-1$. If n, number of swap times, is also small, we have

$$CB(T_0) \approx C_n B(T_0, T_n)$$

Besides, for the same reason, we have by swaption pricing formula (17)

$$
\begin{aligned}
Swn(t) &= \sum_{i=1}^{n} C_i B(t,T_i)\widetilde{P}_t^{T_i}(CB(T_0) > K) - KB(t,T_0)\widetilde{P}_t^{T_0}(CB(T_0) > K) \\
&\approx C_n B(t,T_i)\widetilde{P}_t^{T_n}(CB(T_0) > K) - KB(t,T_0)\widetilde{P}_t^{T_0}(CB(T_0) > K) \\
&\approx C_n B(t,T_i)\widetilde{P}_t^{T_n}(C_n B(T_0,T_n) > K) - KB(t,T_0)\widetilde{P}_t^{T_0}(C_n B(T_0,T_n) > K) \quad (43)
\end{aligned}
$$

Obviously, we only use (43) to analyze our experiments, but do not use it to price swaption, because (43) only exists when the times of swap and the swap rate is small.

We notice that $B(T_0, T_n)$ is lognormal distributed in a three-factor Gaussian model, and thus can get idea from experiments on Lognormal distributions on the third chapter that the volatility σ of $B(T_0, T_n)$, defined as the standard deviation of $logB(T_0, T_n)$ can be used to adjust the accuracy of CDG approximation according to Figure 3.1 also on the third chapter, and this adjustment turns to be quite effective in practice.

To be more specific, according to (40) and (41), the volatility σ keeps the same even its forward measure changes. For Experiment 4.1, $\sigma = 0.0479$, and for Experiment 4.2, $\sigma = 0.0558$. From Figure 3.1, the optimal σ is about 0.045, which is affected little by the other parameter μ, so volatilities σ in both experiments are very close to the optimal σ. In the next experiment, Experiment 4.3, we will try to rise volatility σ_2 , σ_3, so that σ will rise to 0.0856. Intuitively, it will lead to less accurate CDG approximation.

Experiment 4.3

Consider a $T_0 * (T_2 - T_0)$ receiver swaption, with maturity of swaption $T_0 = 1$ year and tensor $\delta = 0.5$ year. The parameters are given below

α_1	α_2	α_3	σ_1	σ_2	σ_3	x_1	x_2	x_3	δ_0	k
1	0.7	0.5	0.06	0.10	0.10	0.02	0.02	0.01	0.03	0.05

Table 4.3.1 Parameters of One-Year Swaption under Hull-White model

Experiment Result:

Swaption Price by Monte Carlo	0.0378
Swaption Price by CDG Approximation	0.0385
Absolute Error	0.0006
Relative Error	1.69%

Table 4.3.2 Swaption Price Approximation

	T_0	T_1	T_2
Monte Carlo	0.5465	0.5659	0.5799
CDG	0.5397	0.5587	0.5725
Absolute Error	0.0068	0.0072	0.0074

Table 4.3.3 Probability Approximation under Different Forward Measures

	sc_3	sc_4	sc_5	sc_6	sc_7	Meaure
T_0	0.04270354	0.00487562	0.00065360	0.00009706	0.00001548	-7.92
T_1	0.01599799	0.00487571	0.00065362	0.00009706	0.00001547	-7.92
T_2	0.01599732	0.00487579	0.00065364	0.00009706	0.00001548	-7.92

Table 4.3.4 Scaled Cumulants under Different Forward Measures

Conclusion

As we expected, Experiment 4.3 shows relatively less accurate CDG approximation. The scaled cumulants decay relatively slow with measure -7.92. The errors of forward probabilities are about 0.0070 on average, and the approximation of the swaption price demonstrates absolute error about 0.0006, and relative error around 1.69% displayed on Table 4.3.2.

Measure of Errors of Swaption Price

Our next task is to develop an accurate measure of errors of swaption price based on our previous measure l of errors of densities.

There is something in common observed in previous three experiments: the smaller value l is, the more accurate the CDG approximation is on forward probabilities. In Experiment 4.1, l is less than -10, and the absolute errors of probabilities are about 0.0013 illustrated in Table 4.1.3. In Experiment 4.2, l is about -9.5, with error around 0.0035. In Experiment 4.3, l is about -7.9, with errors more or less than 0.0070. This implies that our measure l is also accurate in cases where the underlying variable $CB(T_0)$ is the sum of several lognormals.

In addition, by analyzing swaption price formula (17) and its coefficient C_1,...,C_n, among which C_n is much larger than other C_i, $i = 1, 2, ...n-1$, we find that $\widetilde{P}_t^{T_n}(CB(T_0) > K)$ and $\widetilde{P}_t^{T_0}(CB(T_0) > K)$ have much more weight than other forward probabilities $\widetilde{P}_t^{T_i}(CB(T_0) > K)$, $i = 1, 2, ...n-1$ in determining the swaption price. So, in order to make CDG approximation accurate, we should make probabilities under both forward measures well approximated, and thus consider to choose L as the measure of swaption errors defined as follows

$$L = l_0 + l_n = log|\frac{sc_7^{T_0}}{sc_3^{T_0}}| + log|\frac{sc_7^{T_n}}{sc_3^{T_n}}| \tag{44}$$

where l_0 is our previous measure l under T_0 forward measure and $sc_i^{T_0}$ is the $i-th$ scaled cumulant also under T_0 forward measure. For instance, in Experiment 4.1, $L = -10.82 - 10.23 = -21.05$.

Note that the difference between l_0, l_n and other measures l_i, $i = 1, 2...n-1$ are insignificant, from previous three experiments. Empirically, if l_0 and l_n are both small, other l_i should also be small. We expect the smaller value L is, the more accurate the CDG approximation on swaption price is, just like our previous measure l on densities. Clearly, measure L of errors of swaption price is accurate based on Experiment 4.1, 4.2 and 4.3, where Experiment 4.1 and Experiment 4.2 show much smaller errors in swaption price than Experiment 4.3, both absolute errors and relative errors. In Experiment 4.1, $L = -21.05$ with the absolute error of swaption price about 0.0001 and relative error 0.59%, while in Experiment 4.3, $L = -15.84$ with the absolute error around 0.0006 and relative error 1.69%.

In order to evaluate measure L more generally, we need to collect more data, including values of measure L and swaption price errors, both absolute errors and relative errors. We will price several swaptions with the same parameters as in Experiment 4.1 except the parameter σ_3 since different σ_3 will lead to different value of L . The data is illustrated in Table 4.4, where the second row demonstrates the Monte Carlo swaption price with the same number of simulations as in our pervious experiments. The absolute errors and relative errors are showed on the third row, with relative errors in the brackets. The values of measure L are on the last row.

σ_3	$\sigma_3 = 0.07$	$\sigma_3 = 0.10$	$\sigma_3 = 0.12$	$\sigma_3 = 0.15$
Swaption Price	0.0248	0.0321	0.0376	0.0464
Errors	0.0001 (0.57%)	0.0005 (1.60%)	0.0009 (2.43%)	0.0016 (3.44%)
Measure L	-19.20	-17.21	-16.06	-14.52

Table 4.4

Luckily, from Table 4.4, we also see that the smaller value measure L is, the smaller errors are. So far, we have obtained seven swaption price, including three from Experiment 4.1, 4,2, and 4.3. The relationship between values of measure L and errors of these swaption price is plotted in Figure 4.1 and Figure 4.2, with Figure 4.1 dealing with absolute errors, and Figure 4.2 with relative errors.

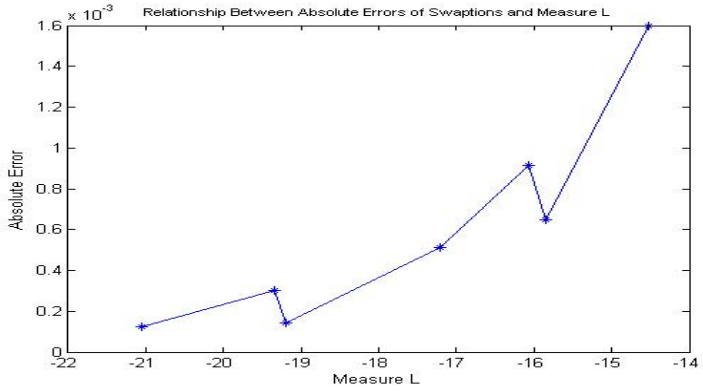


Figure 4.1 Relationship Between Absolute Errors of Swaptions and Measure L

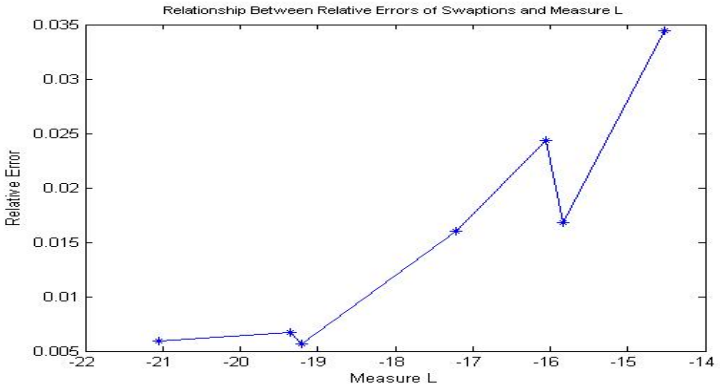


Figure 4.2 Relationship Between Relative Errors of Swaptions and Measure L

Conclusion:

From Figure 4.1 and Figure 4.2, we find that as what we expected, in general, the smaller value measure L is, the smaller the absolute errors and relative errors are although there is no guarantee smaller L will definitely result in smaller errors of both types. It implies our measure L is accurate in general. Besides, the trend can potentially be fitted by some smoothed functions with increasing slopes.

Furthermore, there is an interesting phenomenon that the trend of absolute errors coincides with that of relative errors almost perfectly. That's, if the absolute error of a swaption is smaller than that of another swaption, then its relative error is also smaller, see Figure 4.3. The direct application of this rule is that in the analysis of swaption price errors, absolute errors can be an estimator of relative errors, and vice versa.

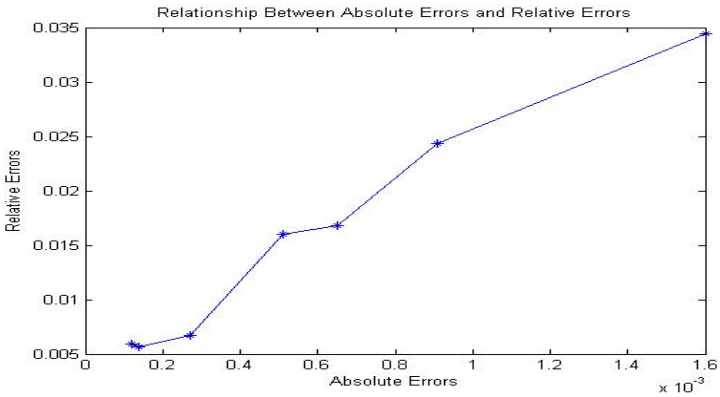


Figure 4.3 Relationship Between Absolute Errors and Relative Errors

In addition. we will analyze the reason behind this interesting phenomenon that absolute errors and relative errors coincide. Let Swn be the true swaption price, and swn be the CDG approximated price. So the absolute error is $|Swn - swn|$ and the relative error is

$\frac{|Swn-swn|}{Swn}$. In general, a small fluctuation of Swn in percentage would lead to large violation of $|Swn - swn|$ in percentage, because the absolute error is much smaller. Thus, the trend of $\frac{|Swn-swn|}{Swn}$ is primarily effected by that of $|Swn-swn|$. This explains why absolute errors coincide with relative errors in trend. For instance, in Experiment 4.1, $Swn = 0.0208$, $|Swn - swn| = 0.0001$, $\frac{|Swn-swn|}{Swn} = 0.59\%$, while in Experiment 4.3, $Swn = 0.0378$, $|Swn - swn| = 0.0006$, $\frac{|Swn-swn|}{Swn} = 1.69\%$.

Two Reasons Why Measure L Accurate Only In General

We will explore reasons why measure L is accurate only generally such that there is no guarantee smaller L would lead to smaller errors.

We can learn from any plot of errors of Lognormal distribution functions in Appendix C, that the place of the variable X on the axis can influence the errors within the same distribution function. In other words, at different places, errors can be quite different. Similarly, for a general distribution function, the place of the variable is also an influence. However, it is usually difficult to predict at which place the errors are small, especially for absolute errors.

The other reason is that swaption price is the combination of several forward probabilities, and accurate CDG approximation on probabilities does not necessary mean accurate swaption price although they usually do. However, in practice, the difference among forward probabilities in the same swaption is usually very small under Three-factor Gaussian model, and thus CDG method has similar approximation performance on them uniformly, see Experiment 4.1 where the relative errors of forward probabilities are all around 0.0012. This is the reason why accurate CDG approximation on probabilities can lead to accurate swaption price in general under Three-factor Gaussian model.

More research is suggested to analyze these two reasons intensively. If successful, then we may be able to work out a more accurate measure of errors of swaption price. In fact, the swaption price is a complicated system, affected by several factors. As far as I know, most relevant researches are concentrating on how to approximate accurate probabilities, and then rely on experiments to see whether the swaption price is accurate approximated. In addition, we propose to collect much more swaption price so we have more points that describe the relationship between errors and measure L. The trend may be fitted by some statistical regression models. Then, we can better assess the power of our measure L.

In practice, we suggest to set value of L less than -19, so absolute errors would be less than 0.0004, and relative errors less than 1%. Furthermore, our result on swaptions within Three-factor Gaussian model can be extended to cases within any Multi-factor Gaussian model, because a zero-coupon bond $B(t,T)$ within any of these models is still Lognormal distributed, for instance, Four-factor Gaussian model, refer to Langetieg (1980). Thus, our result is valuable and can be potentially widely applied.

4.4 CDG Approximation under CIR2++ Model

In this part, we will discuss CDG approximation under two-factor CIR model, in academics, called CIR2++ model, which is an extension of one-factor CIR model. Recall the stochastic

differential equation of one-factor CIR model for interest rate is $dr(u) = (a - br(u))du + \sigma\sqrt{r(u)}d\widetilde{w}(u)$. In CIR2++, its dynamics is similar.

$$dx_i(u) = (a_i - b_i x_i(u))du + \sigma_i\sqrt{x_i(u)}d\widetilde{w}_i(u)$$

$i = 1, 2$, and

$$r(u) = \delta + \sum_{i=1}^{2} x_i(u)$$

where the Brownian motions $\widetilde{w}_1(u)$ and $\widetilde{w}_2(u)$ are independent. a_i, b_i, σ_i for any $i = 1, 2$ and δ are all positive constants such that $2a_i > \sigma_i^2$.

The bond price $B(t, T)$ is obtained in different way with Three-factor Gaussian model, since it is difficult to acquire its price by the distribution of its CIR factors, which are non-central chi-square distributed. We have the following theorem.

Theorem 4.4.3. The price of at time t of a zero-coupon bond maturing at time T is

$$B(t, T) = e^{-A(T-t)-\sum_{i=1}^{2} C_i(T-t)x_i(t)}$$

where

$$C_i(\tau) = \frac{2(e^{\gamma_i\tau} - 1)}{(b_i + \gamma_i)(e^{\gamma_i\tau} - 1) + 2\gamma_i}$$

$$A(\tau) = -\delta\tau + \sum_{i=1}^{2}(\frac{2a_i}{\gamma_i - b_i}\tau - \frac{2a_i}{\sigma_i^2}log[\frac{(b_i + \gamma_i)(e^{\gamma_i\tau} - 1) + 2\gamma_i}{2\gamma_i}])$$

and we define $\gamma_i = \sqrt{b_i^2 + 2\sigma_i^2}$.

Proof. Assume the bond price can be written as

$$B(t, T) = f(t, x_1, x_2) = e^{-A(T-t)-C_1(T-t)x_1(t)-C_2(T-t)x_2(t)}$$

where $A(\tau)$ and $C(\tau)$ are deterministic functions and $A(0) = C(0) = 0$. It follows that

$$f_t(t, x_1(t), x_2(t)) = [A^{'}(T-t) + C_1^{'}(T-t)x_1(t) + C_2^{'}(T-t)x_2(t)]f(t, x_1(t), x_2(t))$$

$$f_{x_i}(t, x_1(t), x_2(t)) = -C_i(T-t)f(t, x_1(t), x_2(t))$$

$$f_{x_i x_i} = C_i^2(T-t)f(t, r(t))$$

For $i \in \{1, 2\}$, where $'$ denotes differentiation with respect to $\tau = T - t$.

Then, the discounted bond price has differential

$$\begin{aligned} d(B(t,T)D(t)) &= -r(t)D(t)f(t,x_1,x_2)dt + D(t)df(t,x_1,x_2) \\ &= D[-rfdt + f_t dt + f_{x_1}dx_1 + f_{x_2}dx_2 \\ &\quad + \frac{1}{2}f_{x_1x_1}dx_1dx_1 + f_{x_1x_2}dx_1dx_2 + \frac{1}{2}f_{x_2x_2}dx_2dx_2] \\ &= D[-(\delta + x_1 + x_2)f + f_t + (a_1 - b_1x_1)f_{x_1} \\ &\quad + (a_2 - b_2x_2)f_{x_2} + \frac{1}{2}\sigma_1^2 x_1 f_{x_1x_1} + \frac{1}{2}\sigma_2^2 x_2 f_{x_2x_2}]dt \\ &\quad + D[\sigma_1\sqrt{x_1(t)}f_{x_1}d\widetilde{w}_1 + \sigma_2\sqrt{x_2(t)}f_{x_2}d\widetilde{w}_2] \end{aligned}$$

Because under the risk-neutral probability, the price of a discounted bond is a martingale, and thus the dt term of $d(B(t,T)D(t))$ should always be zero, then, we have

$$\begin{aligned} & -(\delta + x_1 + x_2)f + f_t + (a_1 - b_1x_1)f_{x_1} \\ & + (a_2 - b_2x_2)f_{x_2} + \frac{1}{2}\sigma_1^2 x_1 f_{x_1x_1} + \frac{1}{2}\sigma_2^2 x_2 f_{x_2x_2} \\ = \; & f[-(\delta + x_1 + x_2) + A^{'} + C_1^{'}x_1 + C_2^{'}x_2 - (a_1 - b_1x_1)C_1 \\ & -(a_2 - b_2x_2)C_2 + \frac{1}{2}\sigma_1^2 x_1 C_1^2 + \frac{1}{2}\sigma_2^2 x_2 C_2^2] \\ = \; & f[(-\delta + A^{'} - a_1C_1 - a_2C_2) + (-1 + C_1^{'} + b_1C_1 + \frac{1}{2}\sigma_1^2 C_1^2)x_1 \\ & + (-1 + C_2^{'} + b_2C_2 + \frac{1}{2}\sigma_2^2 C_2^2)x_2 \\ = \; & 0 \end{aligned} \tag{45}$$

As (42) must hold for all x_1 and y_1, the term $C_1^{'} + b_1C_1 + \frac{1}{2}\sigma_1^2 C_1^2$ multiplying x_1 must be zero. Similarly, the term $C_2^{'} + b_2C_2 + \frac{1}{2}\sigma_2^2 C_2^2$ multiplying x_2 must be zero, and the remaining term $-\delta + A^{'} - a_1C_1 - a_2C_2$ must also be zero. This leads to a system of three ordinary differential equations.

$$C_i^{'} = -b_iC_i - \frac{1}{2}\sigma_i^2 C_i^2 + 1 \tag{46}$$

$$A^{'} = \delta + a_1C_1 + a_2C_2 \tag{47}$$

where $i \in \{1,2\}$.The solutions to the above equations (46)-(47) with initial condition $A(0) = C_1(0) = C_2(0) = 0$ are

$$C_i(\tau) = \frac{2(e^{\gamma_i\tau} - 1)}{(b_i + \gamma_i)(e^{\gamma_i\tau} - 1) + 2\gamma_i}$$

$$A(\tau) = -\delta\tau + \sum_{i=1}^{2}\left(\frac{2a_i}{\gamma_i - b_i}\tau - \frac{2a_i}{\sigma_i^2}log\left[\frac{(b_i + \gamma_i)(e^{\gamma_i\tau} - 1) + 2\gamma_i}{2\gamma_i}\right]\right)$$

where we define $\gamma_i = \sqrt{b_i^2 + 2\sigma_i^2}$.

■

So far, we have obtained the exact formula of the bond price $B(t,T) = e^{-A(T-t)-\sum_{i=1}^{2} C_i(T-t)x_i(t)}$ under CIR2++ model. Note that the products of bond price with different maturities will take the form

$$B(T_0,T_1)B(T_0,T_2)...B(T_0,T_n) = e^{F_0 - F_1 x_1(T_0) - F_2 x_2(T_0)}$$

Our next step is to calculate the conditional expectation $\tilde{E_t^T}(e^{F_0 - F_1 x_1(T_0) - F_2 x_2(T_0)})$, where F_0, F_1, F_2 are all constants. We will only give formula and general procedure for it, For detailed explanation, refer to Duffie, Pan, and Singleton(2000), and Collin-Dufresne and Goldstein(2002).

The conditional expectation of $e^{F_0 - F_1 x_1(T_0) - F_2 x_2(T_0)}$ under T- forward measure can be written as

$$\tilde{E_t^T}(e^{F_0 - F_1 x_1(T_0) - F_2 x_2(T_0)}) = \frac{1}{B(t,T)} e^{M(T_0-t) - N_1(T_0-t)x_1(t) - N_2(T_0-t)x_2(t)} \tag{48}$$

where the functions $M(\tau)$, $N_1(\tau)$ and $N_2(\tau)$ satisfy the following Riccati equations

$$M^{'}(\tau) = -\delta - \sum_{i=1}^{2} a_i N_i(\tau)$$

$$N_i^{'}(\tau) = 1 - b_i N(\tau) - \frac{\sigma_i^2}{2} N_i^2(\tau)$$

with initial conditions $M(t) = F_0^*$ and $N_i(t) = F_i^*$, $i \in \{1,2\}$, where $F_0^* = F_0 - A(T - T_0)$, $F_i^* = F_i + C_i(T - T_0)$,

Solving the Riccati equations above, we have

$$M(\tau) = F_0^* - \delta\tau + \sum_{i=1}^{2}\left(\frac{2a_i}{\gamma_i - b_i}\tau - \frac{2a_i}{\sigma_i^2} log[\frac{(F_i^* - \lambda_{i,-})e^{\gamma_i\tau} - (F_i^* - \lambda_{i,+})}{2\gamma}\sigma_i^2]\right)$$

$$N_i(\tau) = \frac{F_i^*(\lambda_{i,+}e^{\gamma_i\tau} - \lambda_{i,-}) + \frac{2}{\sigma_i^2}(e^{\gamma_i\tau} - 1)}{F_i^*(e^{\gamma_i\tau} - 1) - (\lambda_{i,-}e^{\gamma_i\tau} - \lambda_{i,+})}$$

where $\lambda_+ = \frac{-b+\gamma}{\sigma^2}$ and $\lambda_- = \frac{-b-\gamma}{\sigma^2}$.

Formula (48) is used to calculate moments of a coupon bond under CIR2++ interest rate model. Under T forward measure, we have the differential equations of factors

Theorem 4.4.4. The processes $x_1(u)$, $x_2(u)$ under the T forward measure evolve according to

$$dx_i(u) = (a_i - (b_i + \sigma^2 C_i(T-u))x_i(u))du + \sigma_i\sqrt{x_i(t)}d\tilde{w_i}^T(u) \tag{49}$$

where $i \in \{1,2\}$, and $d\tilde{w_1}^T(u)d\tilde{w_2}^T(u) = d\tilde{w}_1(u)d\tilde{w}_2(u) = 0$. See Brigo and Mercurio (2006). This formula can be applied for Monte Carlo simulation.

■

Simulation

It is well known that the CIR process is non-central chi-square distributed, and there is an exact method to simulate its samples. The idea of this simulation method is to reduce the sampling of a non-central chi-square to sampling of an ordinary chi-square, including generating random variables from standard Normal distribution or Poisson distribution depending on the value of parameter $\frac{4a}{\sigma^2}$. For more detail, see Glasserman (2003).

However, this exact simulation method is quite slow in computation, according to Alfonsi (2005), in which several time discretization schemes of simulation for CIR model are analyzed together with this exact simulation method, and one of the conclusions is that it is more than ten times slower than any other time discretization schemes included. In financial world, usually, the number of samples by Monte Carlo required is very large, and thus we prefer time discretization schemes. In the following section, we will introduce Euler scheme, the simplest but popular time discretization scheme.

Euler Scheme

In mathematics and computational science, Euler scheme, named after Leonhard Euler, is a numerical procedure for solving SDE with a given initial value.

Consider a Ito stochastic differential equation

$$dX_t = f(X_t, t)dt + g(X_t, t)dW(t)$$

For a small time step Δt, we have

$$\begin{aligned} X_{t+\Delta t} &= X_t + \int_t^{t+\Delta t} f(X_s, s)ds + \int_t^{t+\Delta t} g(X_s, s)dW(s) \\ &\approx X_t + \int_t^{t+\Delta t} f(X_t, t)ds + \int_t^{t+\Delta t} g(X_t, t)dW(s) \\ &= X_t + f(X_t, t)\Delta t + g(X_t, t)(W_{t+\Delta t}(t) - W_t(t)) \end{aligned}$$

Or with $t = n\Delta t$, it is more common to write Euler Scheme as

$$X_{n+1} = X_n + f(X_n, n)\Delta t + g(X_n, n)\Delta W_n$$

This is called Euler scheme. The scheme is consistent with the Ito definition of an integral and can only be used for an Ito differential equations. It is well known in Finance, without specific explanation, all SDE are assumed to be Ito. Thus, Euler Scheme is quite useful for Monte Carlo simulation in Finance.

However, in the Euler scheme of CIR model, see (36)

$$r_{n+1} = r_n + (a - br_n)\Delta t + \sigma\sqrt{r_n}\Delta W_n \tag{50}$$

r_n can take negative values since the Brownian motion increment is not bounded from below, and thus $\sqrt{r_n}$ in (50) is not well-defined. To solve this problem, one can take Deelstra and Delbaen scheme, which replaces $\sqrt{r_n}$ by $\sqrt{r_n 1_{(r_n>0)}}$, see Deelstra and Delbaen (1998), or replace the whole (50) by $r_{n+1} = |(a - br_n)\Delta t + \sigma\sqrt{r_n}\Delta W_n|$, see Diop (2003). But a much better scheme is called $E(0)$ scheme in Alfonsi (2005), which is exactly the Milstein scheme

$$r_{n+1} = ((1 - \frac{b}{2}\Delta t)\sqrt{r_n} + \frac{\sigma\Delta W_n}{2(1 - \frac{b}{2}\Delta t)})^2 + (a - \frac{\sigma^2}{4})\Delta t \tag{51}$$

where $4a > \sigma^2$, then r can never be negative.

With respect of how to access the accuracy of these numerical schemes, we will introduce the definition of two types of order convergence.

Definition (strong order of convergence)

The strong order of convergence is j if there exists a positive constant K and a positive constant Δ such that for fixed $T = N\Delta t$

$$E(|X_T - X_N|) \leq K(\Delta t)^j$$

for all $0 < \Delta t < \Delta$. Where X_T is the true value of stochastic process at time T, and X_N is its approximated value.

Definition (weak order of convergence)

The strong order of convergence is j if there exists a positive constant K and a positive constant Δ such that for fixed $T = N\Delta t$

$$|E[h(X_T, t)] - E[h(X_N, t)]| \leq K(\Delta t)^j$$

for all $0 < \Delta t < \Delta$ and for all functions h with polynomial growth.

Both types of order convergence are used to measure the accuracy of numerical scheme, and it is obvious, the higher order of convergence, the better the numerical scheme is. In fact, the strong order of Euler scheme is 0.5, and its weak order is 1.0. On the other hand, Milstein scheme leads to more accurate solution with the same time step, with both strong order and weak order 1.0, but this scheme usually takes more time than Euler scheme. There is also a set of schemes called Taylor schemes which originates from Taylor expansion, with more accurate solution than that by Milstein scheme. In general, the higher order the scheme is, the more computation time will take with the same time step. So, there is a balance for it.

In Alfonsi (2005), by comparing several time discretization schemes for CIR process, including D-D (Deelstra-Delbaen scheme) and Diop (Diop scheme), it is concluded that $E(0)$ scheme is the best with respect to the order of convergence of both types and computational cost. To be more specific, $E(0)$ scheme dominates over D-D and Diop in speed of convergence of both type, but there is almost no difference in computation cost. So, we will use $E(0)$ scheme for CIR simulation. For schemes of higher order for CIR process, second and third order, see advanced research by Alfonsi (2010).

4.5 Numerical Results Under CIR2++ Model

In this section, our attention will be paid to the performance of CDG approximation for swaption pricing under CIR2++ model. Collin-Dufresne and Goldstein only show an accurate case, but not analyze it in detail, as what they do under Three-factor Gaussian model.

Experiments will be carried out to compare swaption price by CDG approximation and by Monte Carlo. The results are all illustrated in tables, with the same meaning as these under Three-factor Gaussian model.

The Monte Carlo price is obtained using $E(0)$ scheme with time discretization formula (51) for each factor $x_i(T_0)$ with differential equation (49). To reduce the time discretization bias, we choose a small time step: $dt = 1 * 10^{-3}$ and short maturity $T_0 = 1$. The number of simulations is all one million for each factor under each forward measure, so the order of error is $o(10^{-6})$. We price swaption in Experiment 4.4 by Monte Carlo two times and both of them showed the same accuracy up to the digit 10^{-5}. In Schrager and Pelsser (2006), they simulate 500,000 samples for CIR2++ model. In Singleton and Umantsev (2002), they use Euler scheme for CIR2++ model with time step $dt = 4*10^{-3}$ from 0 to maturity $T_0 = 5$, and Monte Carlo path is 100,000 without any variance reduction techniques. Clearly, our Monte Carlo price of swaption is more accurate. We consider the Monte Carlo value to be the true value.

Experiment 4.4.

Consider a $T_0 * (T_2 - T_0)$ receiver swaption, with maturity of swaption $T_0 = 1$ year and tensor $\delta = 0.5$ year. The parameters are given below

α_1	α_2	b_1	b_2	σ_1	σ_2	x_1	x_2	δ_0	k
0.01	0.02	0.3	0.5	0.1	0.1	0.04	0.05	0.01	0.1

Table 4.4.1 Parameters of One-Year Swaption under Hull-White model

Experiment Result:

Swaption Price by Monte Carlo	0.0123
Swaption Price by CDG Approximation	0.0127
Absolute Error	0.0004
Relative Error	3.25%

Table 4.4.2 Swaption Price Approximation

	T_0	T_1	T_2
Monte Carlo	0.6366	0.6431	0.6465
CDG	0.6385	0.6446	0.6490
Absolute Error	0.0019	0.0015	0.0025

Table 4.4.3 Probability Approximation under Different Forward Measures

	sc_3	sc_4	sc_5	sc_6	sc_7	Meaure
T_0	-0.07074529	0.00894448	-0.00087905	0.00003272	0.00002030	-8.16
T_1	-0.07082351	0.00897032	-0.00088476	0.00003496	0.00002040	-8.15
T_2	-0.07088556	0.00899081	-0.00088926	0.00003396	-0.00004137	-7.45

Table 4.4.4 Scaled Cumulants under Different Forward Measures

In this experiment, the parameters of two CIR factors $x_1(u)$ and $x_2(u)$ satisfy $2a \geq \sigma^2$ respectively. According to Feller condition, both factors will never hit zero. Then, obviously, $4a > \sigma^2$, so we can use (51) to simulate Monte Carlo samples.

The performance of CDG approximation is good: from Table 4.4.2, the absolute error of swaption is 0.0004 and the relative error is 3.25%, although less accurate than Experiment 4.1 and Experiment 4.2 under Three-factor Gaussian model. This is because the scaled cumulants decay less faster with measure $L = -15.61$ showed in Table 4.4.4. Unlike Three-factor Gaussian model, in practice, it is much difficult to find a case under CIR2++ with fast decay of scaled cumulants, and thus Collin-Dufresne and Goldstein implies it is more difficult to acquire accurate swaption price under CIR2++ model than under Three-factor Gaussian model.

We illustrate this experiment in order to show it is possible to obtain accurate swaption price by CDG approximation. Due to time restriction, and the fact that the simulation of CIR2++ model is rather time consuming, we cannot provide more experimental data. Since the scaled cumulants generally decay less faster under CIR2++ model than that under Three-factor Gaussian model, we can guess that it is less likely to acquire accurate swaption price under CIR2++ interest rate model. In order to assess the power of our measure L, we need to collect more swaption price, and plot the relationship between L and errors of swaption price like what we have done previously, to see whether the trend is similar to that under Three-factor Gaussian model. It is highly recommended to study that.

5 Other Methods for Swaption Pricing

In this chapter, we will introduce other swaption pricing methods under affine interest rate models that provide different insights into swaption other than CDG approximation. In some one-factor interest rate models, there are analytical swaption price, such as Vasicek model and CIR model. Actually, Jamshidian (1989) shows that under one certain condition, the swaption price can be converted into the sum of several options on zero-coupon bonds, and if there is an analytical solution for such an option, there is an analytical swaption price. My innovation is to derive these analytical solutions in a different form, but equivalent to these on literatures. However, in multi-factor models, there is no analytical solution, and thus swaption price has to be approximated. In this section, we will discuss the method by Singleton and Umantsev (2002) and two measures of errors proposed by myself that might be potentially accurate. However, both measures have not been realized yet, and they are my research recommendation. Besides, a trick is proposed by myself to price swaptions under multiple-factor Gaussian models by Singleton and Umantsev method.

5.1 Formulas for One-Factor Models

In this part, I will first introduce my derivation of the analytical solution for the swaption, then discuss the solution in existing literatures, which is in a different form, at last show they are equivalent.

Recall the receiver swaption pricing formula (17) in the second chapter

$$Swn(t) = \sum_{i=1}^{n} C_i B(t,T_i) P_t^{T_i}(CB(T_0) > K) - KB(t,T_0) P_t^{T_0}(CB(T_0) > K))$$

It is already known under one-factor interest rate model, we can write $CB(T_0)$ as

$$\begin{aligned} CB(T_0) &= \sum_{i=1}^{n} C_i B(T_0,T_i) \\ &= \sum_{i=1}^{n} C_i e^{-A(T_i-T_0)-C(T_i-T_0)r(T_0)} \end{aligned}$$

where C_i for any i are all positive. $A(\tau)$, $C(\tau)$ are deterministic functions.

Since there is only one random variable $r(T_0)$ in $CB(T_0)$, we can written $CB(T_0) = f(r(T_0))$. Assume for any $\tau > 0$

$$C(\tau) > 0$$

then $CB(T_0)$ is a decreasing function with respect to $r(T_0)$. Thus, there must exist a value r^*, such that $CB(T_0) > K$ only when $r(T_0) < r^*$. So, $P_t^T(CB(T_0) > K) = P_t^T(r(T_0) < r^*)$. Moreover, it is clear that r^* does not rely on $T-$forward measure and it satisfies

$$\sum_{i=1}^{n} C_i e^{-A(T_i-T_0)-C(T_i-T_0)r^*} = K \tag{52}$$

Therefore, the swaption formula (17) can be rewrite as

$$Swn(t) = \sum_{i=1}^{n} C_i B(t,T_i) P_t^{T_i}(r(T_0) < r^*) - KB(t,T_0) P_t^{T_0}(r(T_0) < r^*)) \tag{53}$$

Then, the only problem left behind is: how to calculate the value of r^*. The answer is to solve equation (52) numerically, and there are mainly two methods for it: Bisection method and Newton–Raphson method.

Bisection method

The bisection method is a root-finding method which repeatedly bisects an interval and then selects a subinterval in which a root must lie for further processing. It is a very simple and robust method, but it is also relatively slow. The following is the procedure for our problem.

Step 1. Set $xl = 0$, $xu = 0.2$, then $f(xl) - K > 0$, $f(xu) - K < 0$. Set er as the maximum error.

Step 2. Let $x = \frac{1}{2}(xl + xu)$. if $f(x) - K > 0$, $xl = x$. if $f(x) - K < 0$, $xu = x$.

Step 3. Repeat step 2 until $xu - xl < er$.

Step 4. Return $r^* = \frac{xu+xl}{2}$ as the solution of equation $f(x) - K = 0$.

where $CB(T_0) = f(r(T_0))$ as we discussed before.

Newton–Raphson method

Newton method is a fast and powerful iteration method for finding approximation to the root. However, it might fail in some situations. For our problem, Newton method always works since $f(r(T_0))$ satisfies all of its assumption includes $f^{'}(x) \neq 0$. The following is the procedure.

Step 1. Set starting point $x_0 = 0.1$. Set er as the maximum error.

Step 2. $x_{n+1} = x_n - \frac{f(x_n)-K}{f^{'}(x_n)}$ until $|f(x_{n+1}) - K| < er$

Step 3. Return $r^* = x_{n+1}$ as the solution of equation $f(x) - K = 0$.

where $CB(T_0) = f(r(T_0))$.

Having obtained the boundary value r^*, our next procedure is to calculate $P_t^T(r(T_0) < r^*)$, for each $T = T_0, T_1, ...T_n$ respectively. If the distribution of random variable $r(T_0)$ is known under forward measures, then $P_t^T(r(T_0) < r^*)$ has an analytical solution, and so does the swaption price (53).

In the general Hull-White model, where the interest rate is normal distributed under any forward measure, $P_t^T(r(T_0) < r^*)$ is known under this model. However, under this model, the assumption $C(\tau) > 0$ is not always satisfied. Thus there might be no analytical solution for the swaption price, and a general and efficient method for it is to construct a trinomial tree to approximate the evolution of the interest rate process, see Brigo and Mercurio (2006).

But if the interest rate follows a Vasicek model, one-factor Hull-White model with constant parameters, then $C(\tau) > 0$ for any τ. In this case, its swaption has a close-form solution.

In one-factor CIR model, the assumption $C(\tau) > 0$ is always satisfied and the interest rate is non-central chi-squared distributed under any forward measure. Thus, there is an analytical swaption price. In the following section, we will derive its swaption price under both Vasicek model and one-factor CIR model.

Analytical solution for swaption under Vasicek model

In this model, we have

$$dr(t) = (a - br(t))dt + \sigma d\widetilde{w}(t) \tag{54}$$

where parameters, a, b and volatility σ all constants.

Then, let $B(t,T) = f(t, r(t))$, it is easy to compute

$$d(D(t)B(t,T)) = Df(-r + A^{'} + C^{'}r - aC + bCr + \frac{1}{2}C^2\sigma^2)dt - DCf\sigma d\widetilde{w}(t)$$

where D, f, r, A, C are short for discount factor $D(t)$, $f(t, r(t))$, $r(t)$, $A(T-t)$, $C(T-t)$ respectively.

By Property 1.1 in the second chapter, we have

$$d\widetilde{w}^T(t) = \sigma C(T-t)dt + d\widetilde{w}(t)$$

Thus, (54) can be rewritten as

$$dr(t) = (a - br(t) - \frac{\sigma^2}{b}(1 - e^{-b(T-t)}))dt + \sigma d\tilde{w}^T(t) \tag{55}$$

Obviously, interest rate $r(t)$ in equation (55) still follows the general Hull-White model under T-forward measure, and thus its distribution is also normally distributed

$$r(T_0) \sim N(e^{-b(T_0-t)}r_0 + (1-e^{-b(T_0-t)})(\frac{a}{b} - \frac{\sigma^2}{b^2}) + \frac{\sigma^2}{2b^2}e^{-b(T-T_0)} - \frac{\sigma^2}{2b^2}e^{-b(T+T_0-2t)}, \frac{\sigma^2}{2b}(1-e^{-2b(T_0-t)})) \tag{56}$$

where $r_0 = r(t)$ is the initial condition mentioned above for differential equation of interest rate.

Therefore, under T forward measure

$$P_t^T(r(T_0) < r^*) = N(r^*, e^{-b(T_0-t)}r_0 + (1-e^{-b(T_0-t)})(\frac{a}{b} - \frac{\sigma^2}{b^2}) + \frac{\sigma^2}{2b^2}e^{-b(T-T_0)} - \frac{\sigma^2}{2b^2}e^{-b(T+T_0-2t)}, \frac{\sigma^2}{2b}(1-e^{-2b(T_0-t)}))$$

where $N(x, \mu, \sigma^2)$ is a Normal distribution at x with expectation μ and variance σ^2. This implies the exact solution for swaption is

$$
\begin{aligned}
Swn(t) &= \sum_{i=1}^{n} C_i B(t,T_i) N(r^*, e^{-b(T_0-t)} r_0 + (1-e^{-b(T_0-t)})(\frac{a}{b} - \frac{\sigma^2}{b^2}) + \frac{\sigma^2}{2b^2} e^{-b(T_i-T_0)} - \frac{\sigma^2}{2b^2} e^{-b(T_i+T_0-2t)}, \frac{\sigma^2}{2b}(1-e^{-2b(T_0-t)})) \\
&\quad -KB(t,T_0) N(r^*, e^{-b(T_0-t)} r_0 + (1-e^{-b(T_0-t)})(\frac{a}{b} - \frac{\sigma^2}{b^2}) + \frac{\sigma^2}{2b^2} - \frac{\sigma^2}{2b^2} e^{-2b(T_0-t)}, \frac{\sigma^2}{2b}(1-e^{-2b(T_0-t)}))
\end{aligned}
$$

Analytical solution for swaption under one-factor CIR model

Like Vasicek model, there is also an analytical solution of the swaption price, and its derivation is similar to that under Vasicek model.

Recall in CIR model

$$dr(t) = (a - br(t))dt + \sigma\sqrt{r(t)}d\tilde{w}(t) \tag{57}$$

where parameters, a, b and volatility σ all constants.

Then, we have

$$d(D(t)B(t,T)) = Df(-r + A' + C'r - aC + bCr + \frac{1}{2}C^2\sigma^2 r)dt - DCf\sigma\sqrt{r}d\tilde{w}(t)$$

Again by Property 1.1, under T forward measure

$$d\tilde{w}^T(t) = \sigma C(T-t)\sqrt{r(t)}dt + d\tilde{w}(t)$$

It implies that

$$dr(t) = (a - (b + \sigma^2 C(T-t))r(t))dt + \sigma\sqrt{r(t)}d\tilde{w}^T(t) \tag{58}$$

The interest rate $r(t)$ in (58) is also non-central chi-squared distributed, and

$$P_t^T(r(T_0) < r^*) = \chi^2(q(t,T_0,T)r^*, v, \delta(t,T_0,T))$$

where $\chi^2(x,v,\lambda)$ is the non-central chi-squared distribution with degree of freedom v and non-centrality parameter λ.

$$q(t,T_0,T) = 2(\frac{2h}{\sigma^2 exp(h(T_0-t)-1)} + \frac{b+h}{\sigma^2} + C(T-T_0))$$

$$\delta(t,T_0,T) = \frac{16r_0 e^{h(T_0-t)}h^2}{q(t,T_0,T)\sigma^4 exp(2h(T_0-t)-2)}$$

and

$$h = \sqrt{b^2 + 2\sigma^2}$$

See Brigo and Mercurio (2006). This implies the exact solution for swaption

$$Swn(t) = \sum_{i=1}^{n} C_i B(t,T_i)\chi^2(q(t,T_0,T_i)r^*, v, \delta(t,T_0,T_i)) - KB(t,T_0)\chi^2(q(t,T_0,T_0)r^*, v, \delta(t,T_0,T_0))$$

■

We have derived analytical formula for swaption under both Vasicek model and one-factor CIR model. As a conclusion, under assumption $C(\tau) > 0$ for any τ , we can derive swaption formula (53). If the distribution of the interest rate $r(T_0)$ is known under forward measure, then the swaption price has an analytical solution.

Actually, there is another formula for swaption price under one-factor interest rate models presented in literatures. This formula converts the swaption price into sum of several zero-coupon bonds. Suppose r^* is the best solution of the following equation

$$\sum_{i=1}^{n} C_i e^{-A(T_i-T_0)-C(T_i-T_0)r^*} = K$$

The payoff of a receiver swaption at time T_0 can be written as

$$[\sum_{i=1}^{n} C_i e^{-A(T_i-T_0)-C(T_i-T_0)r(T_0)} - \sum_{i=1}^{n} C_i e^{-A(T_i-T_0)-C(T_i-T_0)r^*}]^+$$

If for any τ

$$C(\tau) > 0$$

Then, according to Jamshidian decomposition, see Jamshidian(1989), the payoff can be rewritten as

$$\sum_{i=1}^{n} C_i [e^{-A(T_i-T_0)-C(T_i-T_0)r(T_0)} - e^{-A(T_i-T_0)-C(T_i-T_0)r^*}]^+$$

So the price of the swaption becomes equivalent to the value of a portfolio of call options on zero-coupon bonds. That's

$$Swn(t) = \sum_{i=1}^{n} C_i ZBC(t,T_0,T_i, e^{-A(T_i-T_0)-C(T_i-T_0)r^*}) \tag{59}$$

where $ZBC(t,T_0,T_i,K)$ is the price at time t of a call option on a zero-coupon bond $B(T_0,T_i)$ with strike K and option maturity T_0.

Comparing the procedures of deriving swaption formula (53) and (59), we observed that both swaption formulas are under the same assumption, $C(\tau) > 0$ for any τ, and r^* in both have the same meaning. In fact, (53) and (59) can be proved to be exactly the same formula, and the only difference lies in their forms.

Recall in section 2.3, (17) demonstrated the relationship between a coupon bond price and forward probabilities. We will rewrite it in short as

$$\begin{aligned} Swn(t) &= \widetilde{E_t}(\frac{D(T_0)}{D(t)}(CB(T_0) - K)^+) \\ &= \sum_{i=1}^{n} C_i B(t,T_i)\widetilde{P}_t^{T_i}(CB(T_0) > K) - KB(t,T_0)\widetilde{P}_t^{T_0}(CB(T_0) > K) \end{aligned} \quad (60)$$

where $CB(T_0) = \sum_{i=1}^{n} C_i B(T_0, T_i)$

Obviously, from the derivation of the formula above, it still holds even we change the value of its parameters C_i for $i = 1, 2, ...n$. Let $C_i = 0$, $i \neq j$, $C_j = 1$, and $K^* = e^{-A(T_j - T_0) - C(T_j - T_0)r^*}$, then the swaption price in (60) becomes the price of a zero coupon bond displayed as follows

$$\begin{aligned} ZBC(t, T_0, T_j, K^*) &= \widetilde{E_t}(\frac{D(T_0)}{D(t)}(B(T_0 - T_j) - K^*)^+) \\ &= B(t,T_j)\widetilde{P}_t^{T_j}(B(T_0,T_j) > K^*) - K^* B(t,T_0)\widetilde{P}_t^{T_0}(B(T_0,T_j) > K^*) \\ &= B(t,T_j)\widetilde{P}_t^{T_j}(r(T_0) < r^*) - K^* B(t,T_0)\widetilde{P}_t^{T_0}(r(T_0) < r^*) \end{aligned}$$

where the transformation on the last step uses $C(\tau) > 0$ by assumption. Then, it is straightforward to show that two formulas (53) and (59) are equivalent.

To be honest, I found formula (53) when I was trying to formulate one-factor swaption price. Later, I read (59) in literatures with the same assumption. Luckily, they differ some way in forms, but can be proved to be equivalent.

5.2 Methods for Multiple-Factor Models

Under one-factor interest rate models, there might be analytical solutions for swaptions, however, in multi-factor models, it is a completely different story. We quote from Collin-Dufresne and Goldstein (2002) that "Unfortunately, closed-form solutions for swaptions apparently do not exist for multiple-factor affine models. Even in the simplest of models, where it is assumed that future bond price are lognormally distributed, the future value of such a portfolio of bonds would be described by a probability density composed of a sum of lognomral, which has no known analytical solution. It seems unlikely that exact closed-form solutions would ever be found for swaption prices.". Thus, it is essential to find efficient algorithm to price such a swaption.

Literature Review

There are several methods to approximation method and we will take a quick review of literatures related.

Brace and Musiela (1995) obtain a formula in terms of multi-dimensional Gaussian integral for multi-factor Gaussian models. In this formula, the integral has to be computed numerically because the function to be integrated can only be obtained by solving solutions of equations numerically. This approach for swaption pricing is fast and efficient when the factor dimension is two. However, for dimension higher than two, it often becomes numerically very burdensome.

Munk (1999) extends result from Wei (1997) and shows that the price of a European swaption is approximately proportional to the price of a European option on a zero coupon bond with maturity equal to the stochastic duration of the coupon bond, where the stochastic duration is defined to be the time to maturity of a zero-coupon bond having the same relative volatility as the coupon bond.

There is another swaption approximation method proposed by Singleton and Umantsev (2002) which approximate the exercise boundary with a linear function of factors, including least squares. This reduces the exercise of probability to the form that of a caplet. However, it seems not to provide an efficient estimation of the magnitude the pricing error.

Besides, Tanaka (2005) provides a swaption pricing method very similar to CDG approximation. It also calculates cumulants of the underlying coupon bond, but use Gram-Charlier expansion instead for the density. He demonstrates that the third order cumulant approximation is optimal, while in CDG approximation, the order is seven. Unfortunately, there is still no estimation of the pricing error. Assefa (2007) extends the application of CDG approximation into swaption under multi-factor quadratic Gaussian model, and works out similar results.

Other research includes Schrager and Pelsser (2006) in which the approximation method is based on approximating affine dynamics for the forward swap rate under the swap measure. This approximation also reduces price of a swaption to that of a caplet and may lead to the analytical conditional characteristic function of the forward swap rate directly. In this case, it is convenient to compute its density using some Fast Fourier Transform technique, such as Carr and Madan (1999).

Singleton and Umantsev Model

In this section, we will discuss the model by Singleton and Umantsev (2002). Singleton's innovation comes from the observation that the boundary of factors looks linear. In fact, I also find similar linear property independently while analyzing the mathematical property of the coupon bond. Singleton's work is highly valuable, but he seems not to provide an effective measure to indicate how large the swaption error is, although his error analysis is intensive. I try to analyze the error in different way, and provide two potentially accurate measures of errors, which is my research recommendation.

Let us first review Singleton and Umantsev's model. In m-factor models, a bond price can be written as

$$B(T_0, T_i) = e^{-A(T_i - T_0) - \sum_{j=1}^{m} C_j (T_i - T_0) X_j(T_0)}$$

Thus

$$\begin{aligned} CB(T_0) &= \sum_{i=1}^{n} C_i B(T_0, T_i) \\ &= \sum_{i=1}^{n} C_i e^{-A(T_i - T_0)} e^{-\sum_{j=1}^{m} C_j (T_i - T_0) X_j(T_0)} \end{aligned} \tag{61}$$

Singleton approximates the boundary of inequality $CB(T_0) > K$, or

$$CB(T_0) = K \tag{62}$$

by

$$a_1 X_1(T_0) + a_2 X_2(T_0) + ... + a_m X_m(T_0) = b \tag{63}$$

Then, the calculation of probability $P_t^{T_i}(CB(T_0) > K)$ is converted to the computation of $P_t^{T_i}(a_1 X_1(T_0) + ... + a_m X_m(T_0) < b)$, which has closed form conditional characteristic function, see Duffie, Pan, and Singleton (2000), and thus this probability can be easily computed by some Fast Fourier Transformation methods.

The parameters a_1,...,a_m and b in (63) are determined by fitting linear equation (63) through 2^{m-1} points on the boundary by the method of least squares. Suppose $m = 2$ and the density of $X_2(T_0)$ is negligible outside of the interval $[xl, xp]$, then the two points required on the boundary are calculated by endpoints of this interval xl, xu and equation (62). For $m > 2$, one must find 2^{m-1} vertices of a cube of dimension $m - 1$, together with equation (62) to obtain these 2^{m-1} boundary points.

To be more specific, suppose we have endpoints x_1,...,x_{m-1} for their correspondent factors $X_1(T_0)$,...,$X_{m-1}(T_0)$, estimated based on their univariate probability distributions. Let x_m to be a unknown value of factor $X_m(T_0)$ such that $(x_1, ..., x_{m-1}, x_m)$ is a boundary point. Then, x_m is the solution of equation (62), or

$$\sum_{i=1}^{n} C_i e^{-A(T_i - T_0)} e^{-\sum_{j=1}^{m-1} C_j (T_i - T_0) x_j} e^{-C_m (T_i - T_0) x_m} = K \tag{64}$$

This equation can be solved numerically by Newton method.

Remark: one assumption that I think is necessary for Singleton and Umantsev Model is

$$C_j(\tau) > 0$$

for any $\tau > 0$ and any $j = 1, 2, ...m$.

Under this assumption, $CB(T_0)$ is a monotony decreasing function with respect to the factor $X_j(T_0)$ for any j, if other factors are all fixed. Thus, it guarantees there is a unique solution x_m in (64). On the other hand, if this assumption is not satisfied, it is possible that there is more than one solutions of the equation above, and thus the boundary might not be approximated linearly. Luckily, it is known that in both Gaussian model and CIR model of multiple factors, $C_j(\tau)$ is always positive. Although Singleton and Umantsev do not mention this assumption in their paper, but the interest rate model they use for experiments, CIR2++, satisfy this assumption.

So far, we have discussed the general idea of model by Singleton and Umantsev, in which the boundary points are from the endpoints of each interval or cube. Since these boundary points are on the domain where their joint density is low, and thus a potential better way to improve the accuracy of the linear boundary approximation is to collect boundary points on domain with high joint density. So, it is strongly recommended that more study should concentrate on this, as well as the optimal number of boundary points required.

Error Analysis

Singleton and Umantsev find the error of the linear boundary approximation can be expressed as a difference of two strictly positive terms each on the order of $\bar{r} \times T_n \times var(r)$, where $\bar{r}$ and $var(r)$ are the average value of variance of the underlying rate. As indicated in their paper, the error is greater for volatile rates and long swap or bond maturities. However, they do not provide a measure of errors to show how magnificent errors are.

I try to analyze error in different way and provide an indication of errors presented as follows.

It is clear that if the boundary, whose equation is $CB(T_0) = K$, is absolutely linear, then its intersection with any linear subspace should also be linear. In this case, for any fixed positive integers α, β, the solutions of the system of equations (65) and (66) compose a straight line. We call it the boundary intersection.

$$CB(T_0) = K \tag{65}$$

$$X_j(T_0) = v_j, j \neq \alpha, j \neq \beta \tag{66}$$

where $X_k(T_0) = v_k$ is on the domain on which the density of variable $X_k(T_0)$ is not negligible.

Let $x = X_\alpha(T_0)$, $y = X_\beta(T_0)$. Equations (65)-(66) can be combined to be

$$F(x, y) = \sum_{i=1}^{n} C_i' e^{-A(T_l - T_0)} e^{-C_\alpha(T_i - T_0)x - C_\beta(T_i - T_0)y} - K = 0 \tag{67}$$

where C_i' is defined to be

$$C_i' = C_i \Pi_{j \neq \alpha, \beta} e^{-C_j(T_i - T_0)v_j}$$

Since

$$F_x(x,y) + F_y(x,y)\frac{dy}{dx} = 0$$

we have

$$\begin{aligned}\frac{dy}{dx} &= -\frac{F_x(x,y)}{F_y(x,y)} \\ &= -\frac{\sum_{i=1}^{n} C_i' C_\beta (T_i - T_0) e^{-A(T_i-T_0)} e^{-C_\alpha(T_i-T_0)x - C_\beta(T_i-T_0)y}}{\sum_{i=1}^{n} C_i' C_\alpha (T_i - T_0) e^{-A(T_i-T_0)} e^{-C_\alpha(T_i-T_0)x - C_\beta(T_i-T_0)y}} = \theta(x,y) \qquad (68)\end{aligned}$$

If the boundary is linear, $\theta(x,y)$ should be a constant θ for any x and y, and the sufficient condition for this is

$$\frac{C_\beta(\tau)}{C_\alpha(\tau)} = \theta \qquad (69)$$

for any $\tau > 0$, where θ is a constant.

In financial world, x and y are generally very small, and thus $e^{-A(T_i-T_0)} e^{-C_\alpha(T_i-T_0)x - C_\beta(T_i-T_0)y}$ in (68) variates little, which makes $\frac{dy}{dx}$ almost a constant. Naturally, the boundary intersection, on which the joint density of factors is not negligible, can indicate whether boundary can be approximated linearly more or less, just like if we observe waves on the shore, we can estimate how strong the waves are in the sea. This idea also looks similar to the idea in Singleton and Umantsev's method, in which only endpoints are used to fit the whole real boundary by linear regression. To be specific, the more the boundary intersections resemble lines, the better the boundary is linearly approximated in a general sense, and then the more likely Singleton's method will be accurate.

When the number of factors is two, the boundary is just a curve, so we can plot the boundary directly use (68) to see whether it is approximately linear. It implies that our analysis may be highly effective on these two-factor cases. In the following part, I will provide two potential accurate measures of errors based on our analysis.

Measure ψ of Errors

In Singleton and Umantsev (2002), it is suggested that the user of their method can assess the quality of the resulting approximation by replacing one of the endpoints of each interval with the middle point to fit the boundary by linear model (63) to determine new parameters a_1,...,a_m and b, and then compute the swaption price. If the swaption price is substantially different from the original price, then this method is not accurate in this case. However, there is still no measure of errors.

Followed by the Singleton's idea that the linear boundary is fitted by the least square method, I proposed to collect more points on boundary and use the averaged "distances" of these points to the fitted linear boundary as a measure of error. Obviously, these points are on the boundary where the joint density of interest rate factors is not negligible.

To understand more formally, suppose $x^i = (x_1^i, x_2^i, ...x_m^i)$ is such a point, where m is the number of interest rate factors, $i = 1, 2, ...N$, and N is the number of points collected. Let lx^i to be the distance of this point to linear boundary. We define the measure ψ to be

$$\psi = \frac{1}{N}\sum_{i=1}^{N}(lx^i * w(x^i)) \tag{70}$$

where $w(x^i)$ is a certain weight function of the point x^i.

We expect that the more accurate the Singleton and Umantsev's method is, the smaller ψ is. Intuitively, this measure must combine with such a weight function that the higher density a point is, the more weight this point take, because each point on the boundary has different joint density, which will affect the accuracy of Singleton and Umantsev's method extensively. Then, our difficulty becomes how to find such a weight function to make our measure accurate. In order to realize it, we may try several weight function based on the density function until we find an accurate measure. and it is strongly recommended to study them. For the parameter N, intuitively, it is required that N should be sufficient large as to keep this measure accurate, however, large N will cost additional computation time, because for each point, we have to solve equation (64), and thus it is interesting to find an optimal value of it to make a balance between speed and accuracy.

Measure ς of Errors

Although measure ψ has the potential to be an accuracy measure, but from the knowledge of Statistics, we know that number of points N required may grow significantly with number of factors m, and it may become relatively slow in calculation. Thus, we are interested in searching for an another measure that is much faster in computation.

In pervious part of error analysis, I propose a idea that boundary intersections can be used to indicate whether boundary can be approximately linear or not. That's, the geometric structure of boundary, whether it is nearly linear, can be reflected by several curves called boundary intersections. We can take the curvature as a measure. To be more specific, for any $\alpha, \beta \in \{1, 2, ..m\}$, we have a boundary intersection which satisfies equation (65)-(66).

Let $x = X_\alpha(T_0)$, $y = X_\beta(T_0)$ and suppose x is negligible outside the interval $[vx_l, vx_u]$. The length of this boundary intersection within this interval can be determined by the following steps.

- Step 1. Let $x = vx_l$, and solve equation (67), we have $y = vy_l$, such that (vx_l, vy_l) is an endpoint of this boundary intersection.
- Step 2. Choose vx_1, which is near to the vx_l, and we can get another point, denoted by (vx_1, vy_1) , also on the boundary intersection by equation (68). That's

$$\begin{aligned} vy_1 - vy_l &= (vx_1 - vx_l)\frac{dy}{dx}|_{x=vx_l, y=vy_l} \\ &= (vx_1 - vx_l)\theta(vx_l, vy_l) \end{aligned}$$

so

$$vy_1 = (vx_1 - vx_l)\theta(vx_l, vy_l) + vy_l$$

- Step 3. Repeat procedures similar to Step 2, we can access points $(vx_1, vy_1)...(vx_u, vy_u)$ and the computation is very fast.

- Step 4. The length of the boundary intersection, denoted by $l_{\alpha,\beta}$ is calculated by these points

$$l_{\alpha,\beta} = \sum_i \sqrt{(vx_i - vx_{i-1})^2 + (vy_i - vy_{i-1})^2}$$

On the other hand, we have the intersection of linear function that is used to approximates boundary, which satisfies equations (63) and (66), with length denoted by $l'_{\alpha,\beta}$, where we only consider its length within interval $x \in [vx_l, vx_u]$.

Obviously, for a m factor interest model, we have $\binom{m}{2}$ number of $l_{\alpha,\beta}$ of that kind defined above. I propose to define the measure of errors ς to be

$$\varsigma = \frac{1}{\binom{m}{2}} \sum_{\alpha,\beta} \frac{l'_{\alpha,\beta}}{l_{\alpha,\beta}}$$

The advantage of this measure ς is that it is much faster than the previous measure ψ in computation, because it is not required to solve equation $CB(T_0) = K$ numerically, however, the accuracy of this measure still need to be verified by experimental data.

Intuitively, in a general sense, for measure ς , the smaller it is, the more "linear" the boundary is, and thus the more accurate the Singleton's method is. It is recommended to verify this idea by mathematical experiments, and conclude whether ς is an accurate measure. In order to realize it, we should compare this measure with errors of swaptions under the a certain affine model of interest rate but with different parameters to see whether ς grows with error increases. If it is true, then we can try to extend this measure for the general affine models to see what will happen.

Another natural advantage is that it may be quite effective when the interest rate follows two-factor models. One may think about improving this measure by putting the joint density of these curves into consideration similar to what we do with measure ψ, because it is of influence intuitively.

Our research for measures is worthwhile because if any of the above measures is proved to be accurate, then the theoretical gap in Singleton and Umantsev (2002) can be filled, and Singleton and Umantsev's method can become more applicable in practice.

5.3 Trick In Dealing with Multi-factor Gaussian Models

In this section, we will explore when the factor boundary is absolutely linear in multi-factor Gaussian models, since in this situation, this model will cause no error in swaption pricing. This leads to a prospective nice idea that re-calibrating parameters of Gaussian models to fit this condition if parameters from this re-calibration also fit market data well.

Before our discussion, we will introduce two theorems.

Theorem 5.1: The boundary of factors is linear if $\frac{C_\beta(\tau)}{C_\alpha(\tau)}$ is constant for any α, β and $\tau > 0$.

Proof: Let $C_i(\tau) = a_i f(\tau)$ for any $i = 1, 2, ...n$ where a_i are all constants. So

$$\begin{aligned} CB(T_0) &= \sum_{i=1}^{n} C_i e^{-A(T_i-T_0)} e^{-\sum_{j=1}^{m} C_j(T_i-T_0)X_j(T_0)} \\ &= \sum_{i=1}^{n} C_i e^{-A(T_i-T_0)} e^{-\sum_{j=1}^{m} a_j f(T_i-T_0)X_j(T_0)} \\ &= \sum_{i=1}^{n} C_i e^{-A(T_i-T_0)} e^{-f(T_i-T_0)\sum_{j=1}^{m} a_j X_j(T_0)} \\ &= \sum_{i=1}^{n} C_i e^{-A(T_i-T_0)} e^{-f(T_i-T_0)Y(T_0)} \end{aligned}$$

where $Y(T_0) = \sum_{j=1}^{m} a_j X_j(T_0)$.

Let b be the solution of the equation

$$\sum_{i=1}^{n} C_i e^{-A(T_i-T_0)} e^{-f(T_i-T_0)b} = K$$

Thus, the boundary equation $CB(T_0) = K$ is equivalent to

$$Y(T_0) = b$$

or

$$\sum_{j=1}^{m} a_j X_j(T_0) = b$$

This proves that the boundary of factors is linear.

■

Theorem 5.2: Under multi-Gaussian models, the swaption price has absolute linear boundary if $\alpha_i = \alpha_j$ for any i, j.

Proof: In multi-Gaussian models

$$C_i(\tau) = \frac{1 - e^{-\alpha_i \tau}}{\alpha_i}$$

If $\alpha_i = \alpha_j$ for any i, j, it is obvious that $C_i(\tau) = C_j(\tau)$. Thus, the boundary is linear by the previous theorem.

■

This theorem implies, under a m factor Gaussian model, if

$$\alpha_1 = \alpha_2 = ... = \alpha_m$$

then the swaption price in this model would have no error. The only source of error is from calibration of parameters of this Gaussian model by market data.

Calibration means fitting the parameters of the model to the current market data. Usually, parameters are calibrated to fit the current yield curve, implied volatility of bonds or some derivatives, or other data that are available in the market. The general idea of model calibration is to find the optimal values of parameters such that this model is "closest" to the market data under a certain measure, for instance, l^2 measure. However, in this random world, the data from the real market can be interpreted as these randomly simulated from a certain "model", and thus the optimal parameters that fit the market best might not be the best to fit the trend illustrated by these market data.

My idea is that if parameters calibrated from the markets show that $a_1, ..., a_m$ are very close to each other, we can make $\alpha_1 = \alpha_2 = ... = \alpha_m$, then re-calibrate market data with a_i all fixed. If the interest rate model with parameters re-calibrated also fits these market data well, then adopt the new parameters, otherwise, use the original ones. This approach can be frequently seen in references in the field of Time Series.

When m is small, for instance $m = 3$, my approach might be quite useful. However, it is obvious that this approach becomes infeasible if m is large.

On the other hand, in terms of CIR models, the condition of linear boundary lead to a complicated equation, and a sufficient condition for linear boundary is that $\alpha_1 = \alpha_2 = ... = \alpha_m$ and $\sigma_1 = \sigma_2 = ... = \sigma_m$. Of course, this condition is too strong to make any sense in practice.

Recommendations

There are several of my recommendations in this thesis, and I will summarize part of them.

- Although our measure of errors of density functions is displayed to be accurate in experiments of Gamma and Lognormal distributions, it is still recommended to study whether this measure is accurate on other types of distributions, such as Beta distributions, to make a more general conclusion.
- One theoretical gap is that there may be no theoretical error bounds of CDG approximation on density functions. We recommend researches for that because they may provide another way to estimate errors other than by experiments, especially in a situation where the underlying distribution is complicated.
- We have plotted the relationship of measure L and swaption price errors under Three-factor Gaussian model. However, more swaption prices are suggested to be collected, so the trend can be better displayed. Some smoothed functions can be used to fit this trend.
- We have analyzed the influence of parameters on CDG approximation accuracy under Three-factor Gaussian model when swap rate k and number of time of swaps n are small. It is highly recommended to research on parameter influence when k and n are large.
- It is recommended to plot the relationship between measure L and swaption price errors under CIR2++ model, to see whether it coincides with that under Three-factor Gaussian model.

Appendix

Appendix A

Relationship Between Cumulants and Moments

$$
\begin{aligned}
c_1 &= \mu_1 \\
c_2 &= \mu_2 - \mu_1^2 \\
c_3 &= \mu_3 - 3\mu_1\mu_2 + 2\mu_1^3 \\
c_4 &= \mu_4 - 4\mu_1\mu_3 - 3\mu_2^2 + 12\mu_1^2\mu_2 - 6\mu_1^4 \\
c_5 &= \mu_5 - 5\mu_1\mu_4 - 10\mu_2\mu_3 + 20\mu_1^2\mu_3 + 30\mu_1\mu_2^2 - 60\mu_1^3\mu_2 + 24\mu_1^5 \\
c_6 &= \mu_6 - 6\mu_1\mu_5 - 15\mu_2\mu_4 + 30\mu_1^2\mu_4 - 10\mu_3^2 + 120\mu_1\mu_2\mu_3 - 120\mu_1^3\mu_3 \\
&\quad +30\mu_2^3 - 270\mu_1^2\mu_2^2 + 360\mu_1^4\mu_2 - 120\mu_1^6 \\
c_7 &= \mu_7 - 7\mu_1\mu_6 - 21\mu_2\mu_5 - 35\mu_3\mu_4 + 140\mu_1\mu_3^2 - 630\mu_1\mu_2^3 + 210\mu_1\mu_2\mu_4 - 1260\mu_1^2\mu_2\mu_3 \\
&\quad +42\mu_1^2\mu_5 + 2520\mu_1^3\mu_2^2 - 210\mu_1^3\mu_4 + 210\mu_2^2\mu_3 + 840\mu_1^4\mu_3 - 2520\mu_1^5\mu_2 + 720\mu_1^7
\end{aligned}
$$

where μ_i is the i-th moment, and c_i is the i-th cumulant. This above relationship applies to any distributions. See Gardiner(1983)

Appendix B

The relevant coefficients λ_i and γ_i are provided by Pierre Collin-Dufresne and Robert Goldstein as follows

$$
\begin{aligned}
\lambda_0 &= N(\frac{c_1 - K}{\sqrt{c_2}}) \\
\lambda_1 &= \frac{1}{\sqrt{2\pi c_2}} e^{-\frac{(K-c_1)^2}{2c_2}} c_2 \\
\lambda_2 &= c_2 N(\frac{c_1 - K}{\sqrt{c_2}}) + \frac{1}{\sqrt{2\pi c_2}} e^{-\frac{(K-c_1)^2}{2c_2}} c_2(K - c_1) \\
\lambda_3 &= \frac{1}{\sqrt{2\pi c_2}} e^{-\frac{(K-c_1)^2}{2c_2}} [c_2(K-c_1)^2 + 2c_2^2] \\
\lambda_4 &= 3c_2^2 N(\frac{c_1 - K}{\sqrt{c_2}}) + \frac{1}{\sqrt{2\pi c_2}} e^{-\frac{(K-c_1)^2}{2c_2}} [c_2(K-c_1)^3 + 3c_2^2(K-c_1)] \\
\lambda_5 &= \frac{1}{\sqrt{2\pi c_2}} e^{-\frac{(K-c_1)^2}{2c_2}} [c_2(K-c_1)^4 + 4c_2^2(K-c_1)^2 + 8c_2^3] \\
\lambda_6 &= 15c_2^3 N(\frac{K - c_1}{\sqrt{c_2}}) + \frac{1}{\sqrt{2\pi c_2}} e^{-\frac{(K-c_1)^2}{2c_2}} [c_2(K-c_1)^5 + 5c_2^2(K-c_1)^3 + 15c_2^3(K-c_1)] \\
\lambda_7 &= \frac{1}{\sqrt{2\pi c_2}} e^{-\frac{(K-c_1)^2}{2c_2}} [c_2(K-c_1)^6 + 6c_2^2(K-c_1)^4 + 24c_2^3(K-c_1)^2 + 48c_2^4]
\end{aligned}
$$

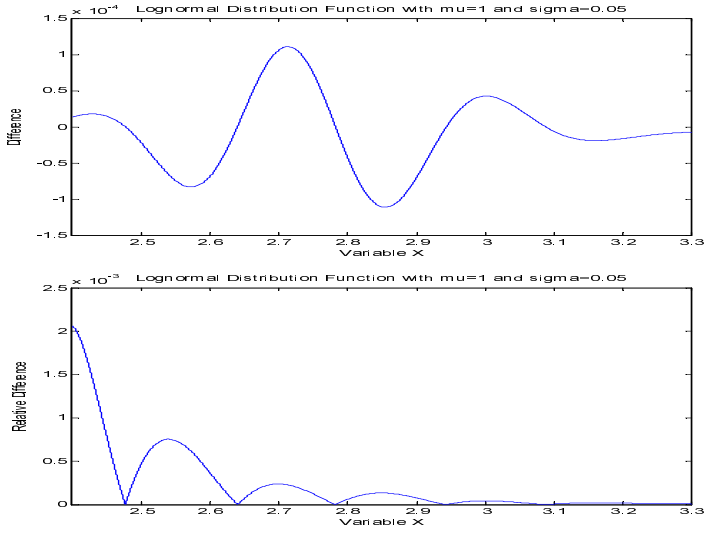
$$
\begin{aligned}
\gamma_0 &= 1+\frac{3}{c_2^2}(\frac{c_4}{4!})-\frac{15}{c_2^2}(\frac{c_6}{6!}+\frac{1}{2}\frac{c_3^2}{(3!)^2}) \\
\gamma_1 &= -\frac{3}{c_2^2}(\frac{c_3}{3!})+\frac{15}{c_2^3}\frac{c_5}{5!}-\frac{105}{c_2^4}(\frac{c_7}{7!}+\frac{c_3c_4}{(3!)(4!)}) \\
\gamma_2 &= -\frac{6}{c_2^3}(\frac{c_4}{4!})+\frac{45}{c_2^4}(\frac{c_6}{6!}+\frac{1}{2}\frac{c_3^2}{(3!)^2}) \\
\gamma_3 &= \frac{3}{c_2^2}(\frac{c_3}{3!})-\frac{10}{c_2^4}\frac{c_5}{5!}+\frac{105}{c_2^5}(\frac{c_7}{7!}+\frac{c_3c_4}{(3!)(4!)}) \\
\gamma_4 &= \frac{1}{c_2^4}(\frac{c_4}{4!})-\frac{15}{c_2^5}(\frac{c_6}{6!}+\frac{1}{2}\frac{c_3^2}{(3!)^2}) \\
\gamma_5 &= \frac{1}{c_2^5}(\frac{c_5}{5!})-\frac{21}{c_2^6}(\frac{c_7}{7!}+\frac{c_3c_4}{(3!)(4!)}) \\
\gamma_6 &= \frac{1}{c_2^6}(\frac{c_6}{6!}+\frac{1}{2}\frac{c_3^2}{(3!)^2}) \\
\gamma_7 &= \frac{1}{c_2^7}(\frac{c_7}{7!}+\frac{c_3c_4}{(3!)(4!)})
\end{aligned}
$$

where λ_i for any i is written in terms of normal distribution, for which there are excellent approximations without the need of numerical integration.

Appendix C

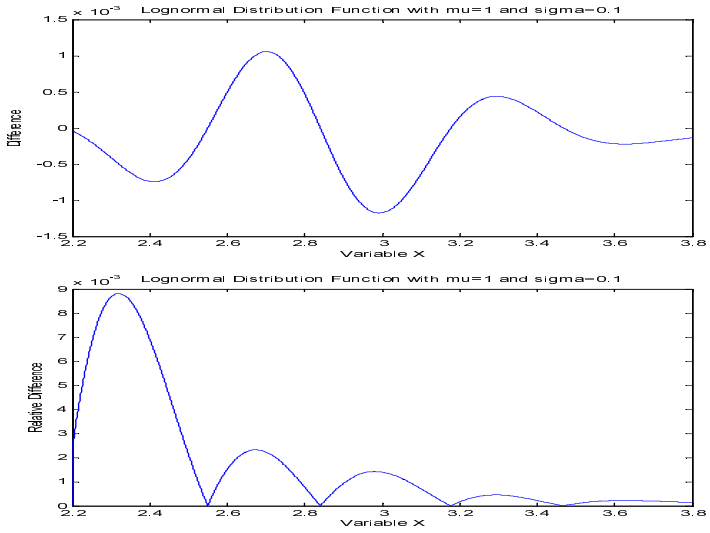
Errors of CDG approximation on Lognormal distribution functions with parameters given in section 3.4.

Illustration 1. $\mu = 1$, $\sigma = 0.05$



sc = [0.0250 0.0017 0.0001 0.0000 0.0000], $l = -9.9934$

Illustration 2. $\mu = 1$, $\sigma = 0.1$



sc = [0.0503 0.0068 0.0011 0.0002 0.0000], $l = -7.2652$

We illustrate these errors in order to show that the CDG approximation have the same approximation performance on distribution functions as on density functions.

Appendix D

Theorem: Scaled cumulants in a Gamma distribution will not be affected by its parameter β.

Proof: As have been discussed. In Gamma distribution $\Gamma(\alpha, \beta)$, the $i-th$ moment m_i is

$$\begin{aligned} m_i &= \frac{\alpha(\alpha+1)...(\alpha+i-1)}{\beta^i} \\ &= \frac{f(\alpha, i)}{\beta^i} \end{aligned}$$

where function $f(\alpha, i)$ is only determined by parameters α and i.

Interestingly, from Appendix A, we observe that the $i-th$ cumulant c_i can be written as

$$c_i = \frac{g(\alpha, i)}{\beta^i}$$

where function $g(\alpha, i)$ is composed of functions $f(\alpha, 1)$,...,$f(\alpha, i)$, but incorrelated with paramter β.

Thus, by definition of scaled cumulants, we have

$$sc_j = \frac{c_j}{j! c_2^{j/2}} = \frac{g(\alpha, j)}{j! g^{j/2}(\alpha, 2)}$$

where sc_j is the $j-th$ scaled cumulant of a Gamma distribution, $j \geq 3$.

It proves that the scaled cumulants in a Gamma distribution will not be affected by parameter β.

■

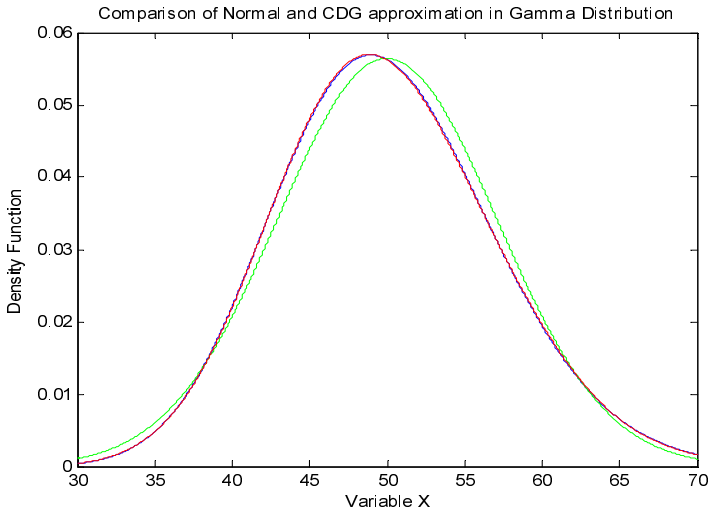
Appendix E

Comparison of CDG Approximation and Normal Approximation

As a special case of Edgeworth expansion, which is much more accurate than just approximating a random variable with a Normal distribution, CDG approximation also shares this nice property. In fact, approximating a random variable with a Normal distribution is equivalent to CDG approximation using only the first two cumulants, because for a Normal distribution, all its cumulants are zero except the first two cumulants as we have discussed.

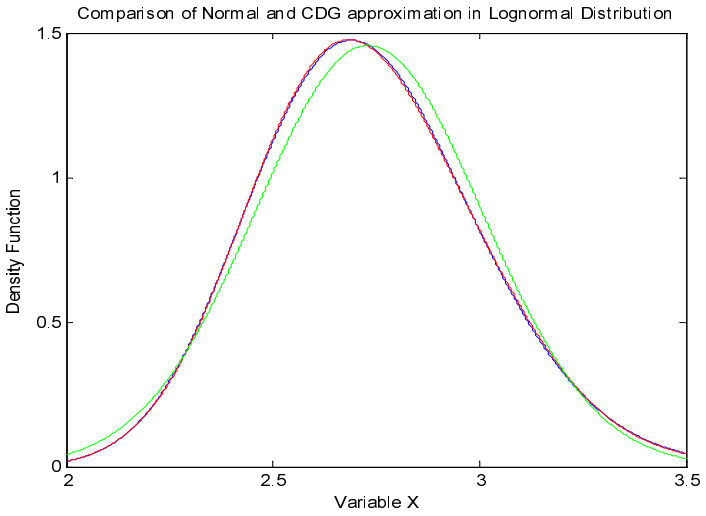
We will illustrate two examples, one from Gamma distribution, the other from Lognormal distribution, to compare their density approximation by Normal and by CDG approximation.

Gamma distribution with parameters $\alpha = 50$, $\beta = 1$



Note that: the blue curve is the true density function of this Gamma distribution, the red one is the CDG approximated density function, and the green is the Normal approximated density function. The same applies to the Lognormal distribution.

Lognormal distribution with parameters $\mu = 1$, $\sigma = 0.1$



Conclusion: It is obvious in both cases, CDG approximation is better than Normal approximation, even when the true density is very close to Normal.

Appendix F (Other Valuable Swaption Pricing Methods)

Three swaption pricing models other than CDG approximation are briefly introduced in this section, which are currently widely used in practice. They are Black-76, LIBOR market model and swap market model. These models are based on different assumptions from CDG approximation, at least not assume the interest rate to be affine. However, they can lead to analytical swaption price.

Definition (Forward swap rate). Forward swap rate is the value of swap rate that makes the initial value of the swap equal to zero. Let $R_n^N(t)$ be the t time forward swap rate of $T_n * (T_N - T_n)$ swap. We have its formula as follows

$$R_n^N(t) = \frac{B(t,T_n) - B(t,T_N)}{\sum_{i=n+1}^{N} \delta B(t,T_i)}$$

Black-76 Formula

The Black-76 formula resembles the famous Black-Scholes formula and quote prices in terms of the implied Black volatilities. It has been used by the market for a long time due to simple

formula and easy calibration, however, it does not have any explicit coherent underlying model.

The Black-76 formula for a $T_n * (T_N - T_n)$ payer swaption $Swn(t)$ with swap rate k is as follows

$$Swn(t) = S_n^N(t)(R_n^N(t)N(d_1) - kN(d_2))$$

where

$$d_1 = \frac{1}{\sigma\sqrt{T_n - t}}(In(\frac{R_n^N(t)}{k}) + \frac{1}{2}\sigma^2(T_n - t))$$

$$d_2 = d_1 - \sigma\sqrt{T_n - t}$$

$$S_n^N(t) = \sum_{i=n+1}^{N} \delta B(t, T_i)$$

and the constant σ is known as the Black volatility.

Swap Market Model

Swap market model was introduced by Brace, Gatarek, and Musiela in 1997. It bases on the assumption that forward swap rate is log-normal distributed.

Assume for each positive integer pairs (l, m) such that $n \leq l < m \leq N$

$$dR_l^m(t) = R_l^m(t)\sigma_{l,m}(t)dW_l^m(t)$$

where W_l^m is Brownian motion under Q_l^m measure, that is, using $S_l^m(t)$ as the numeraire.

Then, the $T_n * (T_N - T_n)$ payer swaption price is

$$Swn(t) = S_n^N(t)(R_n^N(t)N(d_1) - kN(d_2))$$

where

$$d_1 = \frac{1}{M_{n,N}}(In(\frac{R_n^N(t)}{k}) + \frac{1}{2}M_{n,N}^2)$$

$$d_2 = d_1 - M_{n,N}$$

$$M_{n,N}^2 = \int_t^{T_n} ||\sigma_{n,N}(s)||^2 ds$$

Remark: $||.||$ is the L^2 norm. It is clear that formula by swap market model is similar to one by Black-76 formula, while the major difference is the volatility $\sigma_{n,N}(s)$.

LIBOR Market Model

LIBOR market model assumes LIBOR forward rate has a log-normal distribution. That's

$$dL(t, L_{i-1}) = L(t, L_{i-1})\sigma_i(t)dW^i(t)$$

for any $n \leq i \leq N$, where W^i is Brownian motion under T_i forward measure.

Then, it is possible to price LIBOR derivatives in closed form solutions. LIBOR market model is powerful in pricing caps, floors and Bermudan swaptions. However, for swaptions in our case, the solution may have no easy form, and we will skip it.

The assumption of the LIBOR market model originates from the fact that $L(t, L_{i-1})$ is a martingale under T_i forward measure. The swap market model is not consistent with LIBOR Market model in general, because under the assumption of swap market model, the LIBOR rate will not be lognormal distributed under forward measure. But empirical evidence seems to reject the swap market model in favor of the LIBOR market model. See De Jong, J.Driessen and A.Pelsser (2000). For detailed description of these three models, refer to Bjork (2003).

References

[1] Alfonsi, A. (2005), On the discretization schemes for the CIR (and Bessel squared) processes, Monter Carlo Method and Application, pp 355-384.

[2] Alfonsi, A. (2010), High order discretization schemes for the CIR process: Application to affine term structure and Heston models, Mathematics of Computation, pp 209-237.

[3] Assefa, S. (2007), Calibration and Pricing in a Multi-factor Quadratic Gaussian Model. Working paper, University of Technology Sydney.

[4] Bhayyacharya and Ghosh (1978), On the validity of the formal Edgeworth expansion, Annals of Statistics, Vol 39, No 6, pp 435-451.

[5] Bickel, Götze and Zwet (1986), The Edgeworth Expansion For U-Statistics of Degree Two, The Annual of Statistics, Vol 14, No 4, pp 1463-1484.

[6] Bjork, T. (2003), Arbitrage Theory in Continuous Time, Oxford University Press.

[7] Blinnikov and Moessner (1998), Expansions for Nearly Gaussian Distributions, Astronomy & Astrophysics, pp 193-205.

[8] Bloznelis and Götze (2000), An Edgeworth Expansion for Finite-Population U-Statistics, Bernoulli, Vol 6, No 4, pp 729-760.

[9] Brace, A. and Musiela, M. (1995), A Multi-Factor Gauss Markov Implementation of Heath, Jarrow and Morton, Mathematical Finance, pp 259-283.

[10] Brigo, D. and Mercurio, F. (2006), Interest Rate Models – Theory and Practice, Springer.

[11] Carr and Madan (1999), Option Valuation Using the Fast Fourier Transform, Computational Finance, pp 61-73.

[12] Collin-Dufresne, P. and Goldstein, R. (2002), Pricing Swaption Within an Affine Framework, The Journal of Derivatives, pp 9-26.

[13] Cox, Ingersoll and Ross (1985), A Theory of the Term Structure of Interest Rates, Econometrica, pp 385–407.

[14] Deelstra, G. and Delbaen, F. (1998), Convergence of Discretized Stochastic (Interest Rate) Processes with Stochastic Drift Term, Appl. Stochastic Models Data Anal. 14, pp. 77-84.

[15] De Jong. Driessen, J. and Pelsser, A. (2000), LIBOR and Swap Market Models for Pricing Interest Rate Derivatives: An Empirical Analysis. Working paper, Tilburg University.

[16] Duffie. Pan. and Singleton. (2000), Transform Analysis and Option Pricing for Affine Jump-Diffusions, Econometirica.

[17] Feller, W. (1971), An Introduction to Probability Theory and Its Application, Volume Two, John Wiley and Sons.

[18] Gardiner (1983), Handbook of Stochastic Methods, Springer-Verlag.

[19] Glasserman (2003), Monte Carlo Methods in Financial Engineering, Springer.

[20] Hall, P. (1992), The bootstrap and Edgeworth Expansion, Springer-Verlag.

[21] Jamshidian, F. (1989), An Exact Bond Option Formula, The Journal of Finance, Vol 44, No 1, pp 205-209.

[22] Jarrow and Rudd (1982), Approximate Option Valuation for Arbitrary Stochastic Processes, Journal of Financial Economics, Vol 10, No 3, pp 347-369.

[23] Ju (2001), Pricing Asian and Basket Options Via Taylor Expansion of the Underlying Volatility, Working paper, University of Maryland.

[24] Langetieg (1980), A Multivariate Model of the Term Structure, Journal of Finance, Vol 35, pp 71-97.

[25] Munk, C. (1999), Stochastic Duration and Fast Coupon Bond Option Pricing in Multi-Factor Models, Review of Derivatives Research, pp 157-181.

[26] Musiela, M. and Rutkowshi, M. (2004), Martingale Method in Financial Modelling, Springer.

[27] Remmert, R. (1989), Theory of Complex Functions, Springer-Verlag.

[28] Schrager and Pelsser (2006), Pricing Swaptions And Coupon Bond Options In Affine Terms Structure Models, Mathematical Finance, Vol. 16, No. 4, pp 673–694.

[29] Shreve, S. (2004), Stochastic Calculus for Finance, Springer.

[30] Singleton and Umantsev (2002), Pricing Coupon-Bond Options and Swaptions in Affine Terms Structure Models. Mathematical Finance, pp 427-446.

[31] Tanaka, K. (2005), A Simple Formulation of Swaption Prices by Gram-Charlier Expansion. Working paper, Kyoto University.

[32] Turnbull and Wakeman (1991), A Quick Algorithm for Pricing European Average Options, Journal of Financial and Quantitative Analysis, pp 377-389.

[33] Wei (1997), A Simple Approach to Bond Option Pricing, Journal of Futures Markets, pp 131–160.