Hedging Cost Analysis of Put Option with Applications to Variable Annuities

by

Panpan Wu

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Graduate Department of Statistics
University of Toronto

Abstract

Hedging Cost Analysis of Put Option with Applications to Variable Annuities

Panpan Wu

Doctor of Philosophy

Graduate Department of Statistics

University of Toronto

2013

Variable annuities (VA) are equity-linked annuity contracts which provide the opportunity for policy-holders to benefit from financial markets appreciation and at the mean time provide protection from the downside risks of the markets. They have been overshadowing traditional fixed annuities to become the leading form of protected investment worldwide. However, the embedded guarantees in VA can bring significant downside risks to the insurer and need to be hedged. Among different hedging strategies, move-based discrete hedging strategies are widely adopted in practice but the cost analysis for move-based discrete hedging strategies are mathematically complex. In this thesis, we first examine various move-based hedging strategies and show that a two-sided underlier-based hedging strategy is desirable for the return of premium guarantee. Then we assume a GBM model for the sub-account to develop a semi-analytic framework for the hedging cost analysis of this strategy and thereby propose a modified "Percentile Premium Principle", which imposes a significant "loading" on top of the regular charge to cover the costs arising from the discrete re-balances under the two-sided underlier-based strategy. We apply the modified "Percentile Premium Principle" to the pricing of various VA designs, including GMMB, annual ratchet VA, structured product based VA with both buffered and contingent protection. Finally, we advance the algorithm towards a more general model-GBM with regime switching-to allow a better representation of the VA sub-account.

Acknowledgements

Several people have influenced this thesis in various ways and it is my pleasure to express my thanks to them here.

I owe a particular debt to my supervisor Prof. X. Sheldon Lin. His guidance illuminated my way towards independent research. His encouragement led me go the distance.

Professor Andrei L. Badescu, Professor Samuel Broverman, Professor Radu V. Craiu, Professor Andrey Feuerverger, Professor Tak Kuen (Ken) Siu, and Professor Fang Yao carefully reviewed the draft version of my thesis and provided numerous constructive comments. Their patience and insights are warmly appreciated. Furthermore, Professor Broverman generously offered me many study materials during my preparation for the SOA preliminary exams, which has greatly accelerated my progress.

My special thank goes to Professor Yichun Chi, for his enthusiastic discussions with me on a wide range of academic topics and his share of personal experiences in conducting excellent research.

I thank Professor Sebastian Jaimungal. His solid mastery of financial modeling benefited me through his courses on derivative pricing theory.

I would also like to thank Professor Radu V. Craiu, Professor Radford Neal, Professor Jeremy Quastel, Professor Nancy Reid, Professor Jeffrey Rosenthal, Professor Balint Virag and Professor Fang Yao, for their interesting courses on various aspects of probability and statistics that prepare me well for the challenges in research.

The staff members in my department, especially Ms. Andrea Carter and Ms. Angela Fleury, have always been forward in lending me a hand. Their professionalism is highly regarded.

At last, I am dearly and forever grateful to my parents and grandparents, for they fill my heart with hope and love.

Contents

Chapter 1

Introduction

1.1 Background

Over the last decade, variable annuities (VA) have been overshadowing traditional fixed annuities to become the leading form of protected investment worldwide. In the US, the total sales peaked in 2007, just before the financial crisis, at \$182.2 billion dollars, and the net assets almost doubled in the past ten years from \$885.8 billion in 2001 to \$1,505.0 billion in 2010. See Insured Retirement Institute (2011) for more detail. In UK, where it has recently started to thrive, the sales figures are catching up, with 538.7 million pounds in 2007, 1,153.3 million in 2008, 1,045.4 million in 2009 and 444.7 million in the first half of 2010. In Japan, where growth has been even more spectacular, from almost no market in 2001 to more than \$100 billion in 2007.

Variable annuities offer a tax-deferred way of investing in the capital markets. For an additional fee, many VA come with guarantees minimum benefits paid out even if the underlying funds dries up. In this sense, VA provides its policyholders the opportunity to gain from financial markets appreciation as well as the protection from the downside risks of the markets.

According to a report by Oliver Wyman Limited (2007)[1], the popularity of these contracts is driven by the demographic changes taking place in most parts of the world. The over-50 population is getting larger, richer and more diverse in styles of life. They demand access to market appreciation in order to keep abreast with the rising cost of living, but also expect protection for their assets and well-being given increased uncertainty in the volatility of asset returns. With the various investment options and multiple forms of guarantees variable annuities could offer, they would be able to choose the best fit for their desired risk/return target.

Though assets reached an all-time high of $1.61 trillion in the first quarter of 2012, the variable annuity market has been sending adverse signals lately. Sales are down, several key players have exited the market (Indeed, companies such as Sun Life, The Hartford, Jackson National, ING and MetLife have dialed down their variable annuity exposure or pulled-out entirely. See AnnuityDigest (2012) for more details). Critics say variable annuities are too costly, the relatively high fees annuity investors pay can eat up a significant amount of money over the long term. Variable annuities typically charge 1.25 percent to 1.60 percent mortality and expense fees on top of the fund expense ratios. So instead of paying around 1% for all-in participant fees, the participant pays at least 2.25 to 2.6 percent for annuity products and often even more. Some insurers respond to these criticism by designing "Structured Product Based Variable Annuities", whose payouts are based on the price changes of a reference asset (an index, equity, interest rate, etc.) but subject to buffers against losses and caps on gains. More specifically, structured products underlying these new annuity contracts pass on losses below some threshold or "buffer" to investors and limit the gains of investors at some "cap" (See Deng et al. (2012) and references therein). These new features allow the provision of

[1]The report discusses the European retirement protection market but the reasons given in the report are applicable to other countries too.

cheaper products by the insurers to boost their sales.

1.2 Basic Features of Variable Annuities

Tax-Deferred Returns

Payments and returns of the variable annuities are tax-deferred. All dividends, interest and capital gains are automatically reinvested without incurring local, state or federal taxes. All earnings are taxed at ordinary income tax rates when withdrawn. Therefore, returns compound more quickly without the erosive effects of taxes, and the value of the annuity is allowed to grow at a faster pace.

Separate Account and Sub-Accounts

Variable annuities provide the choice of stock and bond portfolios called “sub-accounts” that are similar to mutual funds. By law, this account must be kept separate from the insurance company’s general account, so that all dividends, interest, gains and losses are separate and apart from the finances of the insurance company. Sub-accounts allow the owner to tailor an asset allocation model for specific investment objectives. The number of sub-account choices and the specific funds will depend on the individual variable annuity contract that the investor chooses.

Equity Participation with Minimum Guarantees

Variable annuities offer participation in an underlying index or fund or combination of funds through its sub-account, in conjunction with one or more guarantees. The level of participation is usually controlled by the participation rate, which is the percentage of the gain of the sub-account to be credited to the policyholder. According to Insured Retirement Institute (2011), the guaranteed minimum benefit in a VA contract generally fall into four categories

- Guaranteed Minimum Maturity Benefit (GMMB): the policyholder is guaranteed a certain monetary amount at the maturity of the contract. A simple GMMB might be a guaranteed return of premium if the underlying fund falls over the term of the insurance. The guarantee may be fixed or subject to regular or equity-dependent increases.

  The typical payoff of the GMMB at maturity T is

  $$(S_0(1+g)^T - S_T)_+,$$

  where S_0 is the initial value of the sub-account, S_T is the value of the sub-account at maturity and g is guaranteed rate of return. From a derivative's perspective, this payoff is simply that of a put option (a financial derivative that offers the buyer the right to sell an asset at a certain price, regardless of its spot price at maturity) written on the sub-account with strike price $K = S_0(1+g)^T$.

- Guaranteed Minimum Death Benefit (GMDB): a specific monetary sum is given to the policyholder upon his/her death. Like GMMB, the death benefit may simply be the original premium, or may increase at a fixed rate of interest. But more generous death benefits are not uncommon.

  The typical payoff of the GMDB is

  $$(S_0(1+g)^{T(x)} - S_{T(x)}))_+,$$

  where S_0 is the initial value of the sub-account, $S_{T(x)}$ is the value of the sub-account at maturity, g is guaranteed rate of return and $T(x)$ is the time of death of a policyholder whose current age is x.

  When an insurer sells a large number of VA contracts, the mortality risk can be diversified during the accumulation phase of the VA contracts, and as a result it can be treated deterministically. For this reason we ignore the mortality risk in the following calculation. However, the methods we develop in Chapter 4 can be easily extended to incorporate the mortality risk by the use of combination of exponential functions to approximate the mortality density.

- Guaranteed Minimum Accumulation Benefit (GMAB): the policyholder has the option to renew the contract at the end of the original term, at a new guarantee level appropriate to the maturity value of the maturing contract. It is a form of guaranteed lapse and reentry option.

- Guaranteed Minimum Withdraw Benefit (GMWB): the holder can withdraw guaranteed periodic amounts up to the value of the initial capital. The GMWB terminates once the initial capital has been withdrawn; any remaining funds in the sub-account are returned to the policyholder at maturity.

1.3 Hedging Variable Annuities

The minimum guarantees wrapped in VA contracts appear as financial options to insurers. That is, the holder has the right to sell an asset at a certain price regardless of the its market value at the time of sale. These options can bring potentially significant downside risks to the insurer and therefore needs to be hedged or reserved for regulatory and economic purposes. According to Blamont and Sagoo (2009), the relevant risk level for reserving is typically the 0.5% percentile over a one-year horizon and if assets are kept as a reserve against deterministic market shocks, solvency is only guaranteed at the confidence level considered. Only reserves equal to assets can ensure solvency with 100% confidence. On the other hand, one can hedge the tail risk and the advantage of hedging over reserving is that the capital allocated to hedges can be a lot smaller than reserving. See Table 4 in Blamont and Sagoo (2009) for a comparison.

Broadly speaking, there are four hedging approaches. The first is no hedging at all. For small VA blocks, running naked may be acceptable. However, for larger ones, it can be very risky due to high market volatility that has been exhibited in recent years. The second is to buy reinsurance or structured products from another financial institution. Such products may offer substantial protection to the insurer. But they can be expensive or even unavailable for at least two reasons: firstly, they are customized products designed particularly to meet the insurer's needs; secondly, re-insurers may be reluctant to offer coverage for the guaranteed variable annuities given the increased equity market risks of these products. The third alternative is static hedging. This strategy aims to offset the embedded option in the VA contract through buying a portfolio of options from the financial market. Since the payoff structure of the embedded option is sometimes too exotic to be decomposed as a combination of the payoff of options available in the market, basis risk can be significant. Moreover, VA contracts usually span over a long period of time, but options longer than 5 or 10 years in maturity are not available in the market for reasonable prices. The last strategy, which we will discuss in much detail here, dates back to 1970's when Black and Scholes, in their seminal paper Black and Scholes (1973) on option pricing theory, proposed a continuous hedging strategy known as Greek hedging. Under this strategy, one constructs a replicating portfolio to track the value of the derivatives through re-balancing that incurs transaction costs.

Vast research efforts have been devoted to the pricing/hedging of different VA products in this continuous setting. For instance, Windcliff et al. (2001) explore the valuation of segregated funds using an approach based on the numerical solution of a set of linear complementarity problemsMoller (2001) examines a portfolio of equity-linked life insurance contracts and determines risk minimizing hedging strategies within a discrete-time setup Milevsky and Posner (2001) obtain the no arbitrage and equilibrium valuation of a stochastic-maturity Titanic option whose payoff structure is in between European and American style but is triggered by death; Boyle et al. (2001) propose a method for valuing American options using a Monte Carlo simulation approach that can be used to price the reset feature found in some equity-linked insurance contracts; Forsyth et al. (2003) investigate the performance of several types of hedging strategies for segregated fund guarantees using stochastic simulation techniques; Lin and Tan (2003) propose an economic model that has the flexibility of modeling the underlying index fund as well as the interest rates, which is then applied to the pricing of equity-indexed annuities; Jaimungal and Young (2005) investigate the pricing problem for pure endowment contracts whose life contingent payment is linked to the performance of a tradable risky asset or index whose price is modeled by a finite variation Levy process; Milevsky and Salisbury (2006) develop a variety of methods for assessing the cost and value of the GMWB; Coleman et al. (2007) take into account the jump and volatility risks embedded in guarantees with a ratchet feature and evaluate relative performances of delta hedging and dynamic discrete risk minimization hedging strategies; Shah and Bertsimas (2008) price a life option with guaranteed withdrawal benefits using different asset pricing models, including those that allow the interest rates and the volatility of returns to be stochastic; Melnikov and Romanyuk (2008) use the efficient hedging methodology in order to optimally price and hedge equity-linked life insurance contracts whose payoff depends on the performance of multiple risky assets; Marquardt et al. (2008) propose a methodology for pricing GMDBs under a benchmark approach which does not require the existence of a risk neutral probability measure; Dai et al. (2008) develop a singular stochastic control model for pricing variable annuities with the GMWB; Bauer et al. (2008) introduce a model which permits a consistent and extensive analysis of all types of guarantees currently offered within variable annuity contracts; Wang (2009) derives quantile hedges for GMDB under various assumptions; Lin et al. (2009) consider the pricing problem of equity-linked annuities and variable annuities under a regime switching model when the dynamic of the market value of a reference asset is driven by a generalized geometric Brownian motion model with regime switching; Belanger et al. (2009) model the problem of GMDB with partial withdraw as an impulse control problem and give a method for computing the fair value of the associated insurance fee; Marshall et al. (2010) offer a simple but effective way for insurers to measure the value of the GMIB; Peng et al. (2010) consider the pricing of variable annuities with the Guaranteed Minimum Withdrawal Benefit (GMWB) under the Vasicek stochastic interest rate framework; Piscopo and Haberman (2011) introduce a theoretical model for the pricing and valuation of guaranteed lifelong withdrawal benefit (GLWB) options embedded in variable annuity contracts; Ng and Li (2011) develop a multivariate valuation framework for VA on mixed funds; Jaimungal et al. (2012) develop an efficient method for valuing path-dependent VA products through re-writing the problem in the form of an Asian styled claim and a dimensionally reduced PDE, whose results are then compared with an analytical closed form approximation; Chi and Lin (2012) study the flexible premium variable annuities (FPVA) that allow contributions during the accumulation phase.

Though perfectly in theory, continuous delta hedging can be costly and impractical for insurance companies. The reason lies in the fact that one has to re-balance his hedging portfolio continuously (or at least very frequently) in order to remain hedged. For large investment houses, this is not a problem since they have at hand plentiful financial products and are therefore able to maintain their hedging portfolio at low operational costs. Insurance companies, on the other hand, do not enjoy this advantage. As a result, they seek alternative hedging strategies which can offer the same protection with low costs. Discrete hedging is such a solution. Using this strategy, one constructs the same initial replicating portfolio as continuous hedging, but adjust it discretely in time. This difference gives rise to a non-self-financing replicating portfolio. The cost of discrete hedging consists of two parts, one is the cost of constructing the initial hedging portfolio, the other is the cost associated with each subsequent adjustment.

In practice, there are mainly two kinds of discrete hedging strategies, time-based and move-based respectively. The former hedges the option at equally spaced points in time. Boyle and Emanuel (1980) is among one of the first studies on the distribution of the local tracking error of time-based discrete hedging, which we would briefly summarize in Section 3.4. For the global tracking error, Bertsimas et al. (2000) (and references therein, which investigate the trade-off between hedging frequency and transaction costs.) derives the asymptotic distribution of the tracking error at each re-balancing point, as the number of re-balancing points tends to infinity; Hayashi and Mykland (2005) generalize the result of Bertsimas et al. (2000) to continuous Ito processes and also suggest a data-driven nonparametric hedging strategy for the case of unknown underlying dynamics; Angelini and Herzel (2009) compute the expected value and the variance of the error of a hedging strategy for a contingent claim when trading in discrete time, which are valid for any fixed number of trading dates (however, their methods are not applicable to the move-based hedging). Despite of its analytic tractability, the time-based strategy is a plain approximation to continuous hedging with no regard to the volatility risk. When the volatility is high, the value of the sub-account fluctuates remarkably over short periods, causing the necessity of frequent re-balancing of the hedging portfolio. A wiser choice for this situation is to use move-based hedging, which hedges whenever the value of the sub-account moves out of a prescribed region (See Table 1.3.1 and Table 1.3.2 for comparison of the time- and move-based discrete hedging strategies under the same volatility and hedging frequency. The cost distribution associated with the move-based hedging exhibits smaller variance and thinner right tail). Martellini and Priaulet (2002) provide a systematic empirical comparison of four different hedging strategies in the presence of transaction costs within a unified mean-variance framework. They conclude, among others, that the advantage of move-based methods over time-based methods increases with a decrease in the

	Time-Based	Two-sided Underlier-Based
mean	0.0185	0.0077
std	0.1991	0.1169
skewness	0.1004	-0.4511
kurtosis	6.4500	7.3026
90% quantile	0.2404	0.1344
95% quantile	0.3318	0.1847
97.5% quantile	0.4293	0.2378
99% quantile	0.5697	0.3116

Table 1.3.1: Moments and quantiles of cost distribution under the time-based and move-based hedging strategies for an at-the-money put option. The model parameters are $T = 3, S_0 = K = 50, r = 0.02, \mu = 0.1, \sigma = 0.1, d = 0$. For the time-based hedging, the number of re-balance is 100 for 3 years and for the two-sided underlier-based hedging, α is chosen to be 0.0168 so that the expected number of re-balance is also 100. The row "mean" refers to the mean cost of discrete re-balance. For comparison purpose, we calculate the cost of continuous hedging to be 2.0927 for this case.

drift of the underlying asset, and with an increase in the volatility of the underlying asset.

Unfortunately, cost estimation for move-based hedging is mathematically complex because it involves the stopping times of the value of the sub-account. One way to analyze the cost of move-based discrete hedging is through Monte Carlo simulation. Although straightforward by its nature, the Monte Carlo method has certain drawbacks. The path-dependency of the total hedging cost demands the generation of the whole trajectory of the underlier at each iteration, which is done by discretization. However, as pointed out in Glasserman (2003), this leads to bias in the estimates. For example, the hitting times have to be approximated by interpolation. Furthermore, with the discrete process the band is almost never hit exactly, overshoots are everywhere, which violates the hedging rule. The other class of methods is analytic approximation. For a certain type of move-based hedging, Dupire (2005) finds the limit of the end-of-period tracking error as the bandwidth goes to 0 and compares it with the time-based hedging to conclude that "Nothing beats the move based". Henrotte (1993) derives approximate expressions for expected transactions costs and the variance of the total cash flow from both time- and

	Time-Based	Two-sided Underlier-Based
mean	0.0063	0.0023
std	0.8289	0.5005
skewness	0.1140	-0.4296
kurtosis	4.6064	4.9739
90% quantile	0.9828	0.5736
95% quantile	1.3606	0.7747
97.5% quantile	1.7132	0.9680
99% quantile	2.2065	1.2049

Table 1.3.2: Moments and quantiles of cost distribution under the time-based and move-based hedging strategies for an at-the-money put option. The model parameters are $T = 3, S_0 = K = 50, r = 0.02, \mu = 0.1, \sigma = 0.3, d = 0$. For the time-based hedging, the number of re-balance is 100 for 3 years and for the two-sided underlier-based hedging, α is chosen to be 0.05 so that the expected number of re-balance is also 100. The row "mean" refers to the mean cost of discrete re-balance. For comparison purpose, we calculate the cost of continuous hedging to be 8.5598 for this case.

move- based strategies. Toft (1996) extends the work of Henrotte (1993) by showing how these expressions can be simplified and computed efficiently for general input parameters.

As a matter of fact, all the analytic results we mentioned above are asymptotic. Indeed, Dupire's and Henrotte's expressions are obtained in the limit as the bandwidth, the transactions costs and the time between rebalancing points, respectively, go to zero. These limits, however, are clearly unrealistic. Moreover, the mean-variance analysis conducted in Henrotte (1993) and Toft (1996) does not seem to provide very useful information on the two hedging strategies. When a discrete hedging is employed, the value of the option based on the risk-neutral valuation and the cost incurred from discrete hedging should be separated as the former is tradable while the latter is not, nor replicable. A more sensible approach would be, in our opinion, to consider only the latter and to compare their cost distributions for the same re-balancing frequency under the physical measure.

In this thesis, we examine various move-based hedging strategies and develop an alternative semi-analytic algorithm for hedging cost analysis of a desirable move-based hedging strategy. We also propose a modified "Percentile Premium Principle" for variable annuities to incorporate the significant cost associated with discrete hedging. Based on this modified scheme, the insurers implementing move-based discrete hedging in managing their VA risk exposures should charge a significant "loading" on top of the traditional fee.

Chapter 2

Option Pricing Basics

In this chapter, we review some basic knowledge of option pricing theory that are essential in understanding variable annuity hedging.

2.1 Dynamic Hedging of Derivatives

Derivatives are financial contracts which derive their value from some other assets (called the underlying asset). Suppose we sell a derivative written on the underlying asset S with payoff at maturity T $\varphi(S_T)$. We denote by S_t the value of the underlying at time t and by P_{tT} the time t price of a zero coupon bond maturing at T. In order to hedge the option, we construct a self-financing portfolio that replicates the option value at maturity. To start with, we assume that the portfolio is only restructured at finite time points $0 = T_0 < T_1 < \cdots < T_n = T$. Our portfolio contains $\Delta_i e^{-d(T_n - T_i)}$ units of underlying S in $[T_i, T_{i+1})$, where d is the dividend yield of the underlying asset.

At T_0, the initial value of our replicating portfolio is V_{T_0}, which consists of $\Delta_0 e^{-dT_n}$

units of underlying and therefore $(V_{T_0} - \Delta_0 e^{-dT_n} S_0)P_{0n}^{-1}$ units of bond

$$\begin{aligned}
V_{T_0} &= \Delta_0 e^{-dT_n} S_0 + ((V_{T_0} - \Delta_0 e^{-dT_n} S_0)P_{0n}^{-1})P_{0n} \\
&= \Delta_0 e^{-dT_n} S_0 + (V_{T_0} P_{0n}^{-1} - \Delta_0 F_0)P_{0n},
\end{aligned}$$

where $F_0 = e^{-dT_n} S_0 P_{0n}^{-1}$ is the time T_0 forward price of the underlying asset and $P_{in} = P_{T_i,T_n}$.

The value of the portfolio at time $T_1, T_2, \ldots, T_k$ are

$$\begin{aligned}
V_{T_1} &= \Delta_0 e^{-dT_n} S_1 e^{dT_1} + (V_{T_0} P_{0n}^{-1} - \Delta_0 F_0)P_{1n} \\
&= \Delta_1 e^{-d(T_n - T_1)} S_1 + (\Delta_0 - \Delta_1)e^{-d(T_n - T_1)} S_1 + (V_{T_0} P_{0n}^{-1} - \Delta_0 F_0)P_{1n} \\
&= [\Delta_1 F_1 + (\Delta_0 - \Delta_1)F_1 + V_{T_0} P_{0n}^{-1} - \Delta_0 F_0]P_{1n}, \\
V_{T_2} &= \Delta_1 e^{-d(T_n - T_1)} S_1 e^{d(T_2 - T_1)} + (V_{T_0} P_{0n}^{-1} - \Delta_0 F_0)P_{2n} \\
&= \Delta_2 e^{-d(T_n - T_2)} S_2 + (\Delta_1 - \Delta_2)e^{-d(T_n - T_2)} S_2 + (\Delta_0 - \Delta_1)e^{-d(T_n - T_1)} S_1 + (V_{T_0} P_{0n}^{-1} - \Delta_0 F_0)P_{2n} \\
&= [\Delta_2 F_2 + (\Delta_1 - \Delta_2)F_2 + (\Delta_0 - \Delta_1)F_1 + V_{T_0} P_{0n}^{-1} - \Delta_0 F_0]P_{2n}, \\
\cdots & \\
V_{T_k} &= [\Delta_k F_k + (\Delta_{k-1} - \Delta_k)F_k + \cdots + (\Delta_0 - \Delta_1)F_1 + V_{T_0} P_{0n}^{-1} - \Delta_0 F_0]P_{kn} \\
&= [\Delta_{k-1}(F_k - F_{k-1}) + \Delta_{k-2}(F_{k-1} - F_{k-2})) + \cdots + \Delta_0(F_1 - F_0) + V_{T_0} P_{0n}^{-1}]P_{kn}.
\end{aligned}$$

In the continuous time limit, we obtain

$$\frac{V_t}{P_{tT}} = \int_0^t \Delta dF + \frac{V_0}{P_{0T}}.$$

Define the forward value of the option by $U_t = \frac{V_t}{P_{tT}}$, then

$$U_t = \int_0^t \Delta dF + U_0, \tag{2.1.1}$$

and in particular

$$U_T = \int_0^T \Delta dF + U_0. \tag{2.1.2}$$

In general, the RHS of (2.1.2) depends on the whole path of F up to time T. However, on the LHS, $U_T = V_T = \varphi(S_T) = \varphi(F_T)$ depends only on F_T. So we have to find a strategy that gives path-independent value of the portfolio. In other words, our replicating portfolio U_t depends only on t and F_t, not on the values assumed by F before t.

To further analyze this strategy, we need to postulate the dynamics of F. We make the assumption that F follows a stochastic differential equation (SDE)

$$dF_t = \mu(t, F_t)dt + \sigma(t, F_t)dW_t, \tag{2.1.3}$$

where W_t is a Brownian motion. Then according to Ito's lemma

$$dU(t, F_t) = [\frac{\partial U}{\partial t} + \frac{1}{2}\sigma^2(t, F_t)\frac{\partial^2 U}{\partial F^2}]dt + \frac{\partial U}{\partial F}dF_t. \tag{2.1.4}$$

Comparing (2.1.4) with (2.1.1), we get the equations that a self-financing and path-independent portfolio should satisfy

$$\begin{cases} \frac{\partial U}{\partial F} = \Delta_t \\ \frac{\partial U}{\partial t} + \frac{1}{2}\sigma^2(t, F_t)\frac{\partial^2 U}{\partial F^2} = 0. \end{cases} \tag{2.1.5}$$

The partial differential equation (PDE) in the second line of (2.1.5) is of fundamental importance in derivative pricing theory.

2.2 The Black-Scholes Option Pricing Formula

Let us restrict to the simplest form of financial derivatives-European call and put options, whose payoff at maturity T are $\varphi(S_T) = (S_T - K)_+$ and $\varphi(S_T) = (K - S_T)_+$, respectively. In light of the no arbitrage assumption, if we could construct a self-financing and path-independent portfolio that replicates the option payoff at its maturity, then the price of the option at any time t before maturity should coincide with the time-t value of the replicating portfolio.

As noted in the previous section, the forward value U of any self-financing and path-independent portfolio satisfies

$$\frac{\partial U}{\partial t} + \frac{1}{2}\sigma^2(t, F_t)\frac{\partial^2 U}{\partial F^2} = 0.$$

At maturity T, we should have

$$U(T) = V(T) = \varphi(S_T) = \varphi(F_T).$$

Combining these two equations, we get

$$\begin{cases} \frac{\partial U}{\partial t} + \frac{1}{2}\sigma^2(t, F_t)\frac{\partial^2 U}{\partial F^2} = 0 \\ U(T, F_T) = \varphi(F_T). \end{cases} \tag{2.2.1}$$

As the PDE in the first line of (2.2.1) is of first order in time, the final condition in the second line of (2.2.1) is sufficient for computing $U(t, F_t)$ for all $t \in [0, T]$. It is worthwhile to point out that the above PDE does not depend on $\mu(t, F_t)$. So its solution, the option price, will not be affected by changes in $\mu(t, F_t)$.

Once we obtain $U(t, F_t)$, the hedging portfolio can be constructed in the following way:

At time $t(< T)$, we hold $e^{-d(T-t)}\frac{\partial U(t,F_t)}{\partial F_t}$ units of the underlying asset S and buy $V(t, F_t) - e^{-d(T-t)}\Delta_t S_t$ worth of zero-coupon bond P_{tT}. $e^{-d(T-t)}\frac{\partial U(t,F_t)}{\partial F_t}$ is called the Delta of the derivative at time t and this hedging strategy is thus named as delta hedging.

For a call option, (2.2.1) becomes

$$\begin{cases} \frac{\partial U}{\partial t} + \frac{1}{2}\sigma^2(t, F_t)\frac{\partial^2 U}{\partial F^2} = 0 \\ U(T, F_T) = (F_T - K)+ . \end{cases} \tag{2.2.2}$$

In their seminal paper on option pricing-Black and Scholes (1973), Black and Scholes assumed $\sigma(t, F_t) = \sigma F_t$ to obtain the celebrated Black-Scholes equation

$$\begin{cases} \frac{\partial U}{\partial t} + \frac{1}{2}\sigma^2 F_t^2 \frac{\partial^2 U}{\partial F^2} = 0 \\ U(t = T, F_T) = (F_t - K)+ . \end{cases} \tag{2.2.3}$$

Its solution, the Black-Scholes formula, is

$$\begin{aligned} U(t, F_t) &= FN(d_1) - KN(d_2), \\ d1 &= \frac{\log(\frac{F_t}{K}) + \frac{1}{2}\sigma^2(T-t)}{\sigma\sqrt{T-t}}, \\ d_2 &= d_1 - \sigma\sqrt{T-t}, \end{aligned} \tag{2.2.4}$$

So the price of an European call option at any $t < T$ is

$$C(t, F_t) = P_{tT}[FN(d_1) - KN(d_2)].$$

To find the price for the European put $P(t, F_t)$, we exploit the put-call parity

$$C(t, F_t) - P(t, F_t) = P_{tT}(F_t - K),$$

which is a consequence of the no arbitrage hypothesis and the relation $(S_t - K)_+ - (K - S_t)_+ = S_t - K$.

So the put price is

$$P(t, F_t) = P_{tT}[KN(-d_2) - F_tN(-d_1)]. \tag{2.2.5}$$

2.3 Option Greeks

Option Greeks are the mathematical derivatives of the option price with respect to the underlying factors to which its value is attached. They reflect the sensitivities of the option price to small changes in the underlying factors and therefore, can be used to assess risk exposure. For example, by looking at the values of different Greeks, a market-maker with a portfolio of options may be able to understand how the changes in stock prices, interest rates and volatility affect profit and loss.

According to McDonald (2009), the definitions of various option Greeks are

1. Delta (Δ) measures the option price change when the stock price increase by \$1;

2. Gamma (Γ) the change in Delta when the stock price increase by \$1;

3. Vega measures the change in option price when there is an increase in volatility of 1 percentage point;

4. Theta (θ) measures the change in option price when there is a decrease in the time to maturity of 1 day;

5. Rho (ρ) measures the change in option price when there is an increase in the interest rate of 1 percentage point;

6. Psi (Ψ) measures the change in option price when there is an increase in the continuous dividend yield of 1 percentage point.

According to the Black-Scholes formula, the Delta for call and put are

$$\Delta^C(t, F_t) = e^{-d(T-t)}N(d1).$$

and and its Delta is

$$\Delta^P(t, F_t) = -e^{-d(T-t)}N(-d1).$$

respectively.

2.4 The Martingale Approach for Derivatives Pricing

There exists an alternative way to calculate the price of a derivative. To start with, we state the following theorem that provides a probabilistic representation of solutions to certain parabolic PDEs.

Theorem 2.4.1 (Feynman-Kac). *Assume $g \in \mathcal{C}^{1,2}$ and $E[h(X_T)] < \infty$, the two problems below have the same solution*

1.

$$\begin{cases} \frac{\partial g}{\partial t} + \mu(t,x)\frac{\partial g}{\partial x} + \frac{1}{2}\sigma^2(t,x)\frac{\partial^2 g}{\partial x^2} - r(t,x)g = 0 \\ g(t = T, x) = h(x). \end{cases} \tag{2.4.1}$$

2.

$$g(t,x) = E\left[h(X_T)e^{-\int_t^T r(u,X(u))du}\right], \tag{2.4.2}$$

where

$$\begin{aligned} dX_u &= \mu(u, X_u)du + \sigma(u, X_u)dW_u, \\ X_t &= x. \end{aligned}$$

The proof of this theorem can be found in any standard textbook on stochastic analysis. See e.g. Lin (2006).

According to Theorem 2.4.1, the solution to (2.2.1) can be expressed as

$$U(t, F_t) = E[\varphi(F_T)|\mathcal{F}_t] = E[\varphi(S_T)|\mathcal{F}_t],$$

where the expectation is taken w.r.t. a measure under which W_t is a Brownian motion and $dF_t = \sigma F_t dW_t$. We call this measure the T-forward measure and denote it by Q^T. The price of the derivative is $V(t, S_t) = P_{tT} E^{Q^T}[\varphi(S_T)]$, or equivalently

$$\frac{V(t, S_t)}{P_{tT}} = E^{Q^T}[\frac{\varphi(S_T)}{P_{TT}}] = E^{Q^T}[\frac{V(T, S_T)}{P_{TT}}], \quad \forall t < T \tag{2.4.3}$$

From a probabilistic point of view, $\frac{V(t,S_t)}{P_{tT}}$ is a martingale w.r.t. Q^T.
In light of (2.4.3), if we take the zero-coupon bond maturing at T as the numeraire, the relative price $\frac{V(t,S_t)}{P_{tT}}$ of any derivative would be a martingale under the T-forward measure Q^T.

Let r_t be the short rate of interest. Define the money market account by $B_t = e^{\int_0^t r_t dt}$ and the risk neutral measure Q by

$$(\frac{dQ}{dQ^T})_t = \frac{B_t/B_0}{P_{tT}/P_{0T}}, \tag{2.4.4}$$

which is clearly a martingale under Q^T.

Applying the Bayes formula (Karatzas and Shreve (1991) Page 193 5.3 Lemma), we get

$$\begin{aligned}
B_t E_t^Q[\frac{V(T,S_T)}{B_T}] &= E_t^Q[\frac{V(T,S_T)}{B_T}B_t] \\
&= \frac{E_t^{Q^T}[\frac{V(T,S_T)}{B_T}B_t(\frac{dQ}{dQ^T})_T]}{E_t^{Q^T}[(\frac{dQ}{dQ^T})_T]} \\
&= \frac{E_t^{Q^T}[\frac{V(T,S_T)}{B_T}B_t\frac{B_T/B_0}{P_{TT}/P_{0T}}]}{\frac{B_t/B_0}{P_{tT}/P_{0T}}} \\
&= P_{tT}E_t^{Q^T}[\frac{V(T,S_T)}{P_{TT}}],
\end{aligned}$$

where for compactness, we used $E_t[\bullet]$ as a shorthand for $E[\bullet|\mathcal{F}_t]$.

Recalling (2.4.3), we have

$$\frac{V(t,S_t)}{B_t} = E_t^Q[\frac{V(T,S_T)}{B_T}]. \tag{2.4.5}$$

In other words, if we take the money market account B as the numeraire, the relative price $\frac{V(t,S_t)}{B_t}$ of any derivative would be a martingale under the risk neutral measure Q. In summary, the martingale approach represents the price of a derivative as the (conditional) expectation of its payoff under an appropriate martingale measure.

2.5 Geometric Brownian Motion

In the derivation of the Black-Scholes formula, we made the assumption that $\sigma(t,F_t) = \sigma F_t$. Under this assumption, the SDE that F_t satisfies is

$$\frac{dF_t}{F_t} = \sigma dW_t, \tag{2.5.1}$$

where W_t is a standard Brownian motion w.r.t. Q^T.

If we further assume the short rate is constant, i.e. $r_t \equiv r$, then the T-forward measure Q^T would be equivalent to the risk neutral measure Q. Indeed, with constant short rate of interest, $B_t = e^{rt}$, $P_{tT} = e^{-r(T-t)}$ and the Doleans-Dade exponential defined in (2.4.4) is just 1. In this special case, a Q^T Brownian motion is also a Q Brownian motion. So the dynamics of F remains unchanged from Q^T to Q. Moreover, there is a simple relation between F_t and S_t, when the short rate is constant

$$F_t = S_t e^{(r-d)(T-t)},$$

where d is the dividend yield of S.

So the SDE for S is

$$\begin{cases} \frac{dS_t}{S_t} = (r-d)dt + \sigma dW_t \\ S(t=0) = S_0 \end{cases} \tag{2.5.2}$$

with W_t a Brownian motion under the risk neutral measure.

The solution to (2.5.2) is the Geometric Brownian Motion (GBM) with drift $r-d$ and volatility σ

$$S_t = S_0 e^{(r-d-\frac{1}{2}\sigma^2)t+\sigma W_t}.$$

So we conclude that in the Black-Scholes framework, the prorogation of the underlying asset is assumed to follow a GBM.

What is the rationale in this assumption? In fact, GBM can be view as the continuous limit of a discrete price process, generated by a random walk in returns.

Suppose the price of the underlying asset evolves discretely in time. Over each small time interval of length Δt, it can go either up by a factor of $e^{v\Delta t+\sigma\sqrt{\Delta t}}$ or down by $e^{v\Delta t-\sigma\sqrt{\Delta t}}$, with equal probability. We call v the drift and σ the volatility. In a mathematical format, the price changes by a factor of $e^{v\Delta t+\sigma\sqrt{\Delta t}X}$ over each time interval and the changes over disjoint intervals are independent, where X is a random variable with $P(X=1) = P(X=-1) = \frac{1}{2}$. Suppose the price of the underlying asset starts from S_0 at time 0. Than at any fixed time $t = N\Delta t$, the price is $S_t = S_0 e^{vt+\sigma\sqrt{t}\frac{X_1+\cdots+X_N}{\sqrt{N}}}$, where

$X_1, \cdots, X_N$ are i.i.d. random variables with $P(X = 1) = P(X = -1) = \frac{1}{2}$. In other words, $X_1 + \cdots + X_N$ is a random walk. According to Donsker's Theorem (See Karatzas and Shreve (1991) 4.20 Theorem), $\sqrt{t}\frac{X_1+\cdots+X_N}{\sqrt{N}}$ converges weakly to a Brownian motion W_t (to apply Donsker's Theorem, replace Δt by $\frac{1}{n}$ and let $n \to \infty$.) and $S_t = S_0 e^{vt+\sigma W_t}$ is a GBM.

In practice, the parameters μ and σ in the GBM model can be estimated as follow. Suppose we have a sample of historical stock prices: $S_{t_0}, S_{t_0+\Delta t}, \cdots, S_{t_0+N\Delta t}$. Define $X_i = \log(\frac{S_{t_0+i\Delta t}}{S_{t_0+(i-1)\Delta t}}), \quad 1 \leq i \leq N$. Then assuming that the stock pays no dividend, a consistent estimator for σ is $\hat{\sigma} = \frac{s}{\sqrt{\Delta t}}$ and a consistent estimator for μ is $\hat{\mu} = \frac{\bar{x}}{\Delta t} + \frac{1}{2}\hat{\sigma}^2$, where $\bar{x}$ and s are the sample mean and standard deviation of $\{X_i\}_1^N$ respectively.

Chapter 3

The Move-Based Hedging Strategies

A variable annuity with guaranteed minimum maturity benefit assures the policyholder of a minimum guarantee on the balance of his/her sub-account at the time of maturity T. The guarantee level, denoted by G, typically ranges from 75% to 100% of the purchase payments and may also accrue compound interest up to an advanced age. Denote by $T(x)$ the future lifetime of a policyholder at age x, then from the insurer's point of view, the gross liability of the GMMB rider is given by the discounted payoff

$$e^{-rT}(G - S_T)_+ I_{\{T(x)>T\}}, \tag{3.0.1}$$

where r is the risk free interest rate and S_T the value of the VA sub-account at T.

When the insurer sells a large number of VA contracts, the mortality risk is diversifiable during the accumulation phase of the VA contracts, and as a result it can be treated deterministically. For this reason, we ignore the mortality risk in our analysis and the insurer's liability thus reduces to

$$e^{-rT}(G - S_T)_+. \tag{3.0.2}$$

This payoff, from a financial perspective, is exactly the same as that of an European put option. So in this chapter, we will first describe several move-based hedging strategies for the European put option and compare their relative performances through investigating the cost distributions. Then we derive three key densities for geometric Brownian motion that underpins the analysis of two-sided underlier-based hedging. Finally for completeness, we modify the main result of Boyle and Emanuel (1980) to provide a local and asymptotic analysis of the time-based hedging. As we mentioned in Chapter 1, the cost distribution of the time-based hedging has been extensively in several papers, among which Boyle and Emanuel (1980) is one of the earliest and most intuitive.

3.1 The Move-Based Hedging Strategies

In practice, there are three commonly used move-based hedging strategies for put options, the two-sided underlier based, the one-sided underlier-based and the two-sided Greek-based. We now give a brief description for each of them.

1. **The Two-Sided Underlier-Based Hedging**: This strategy imposes a band for the movement of the value of the sub-account S_t and requires re-balancing at the hitting time to the band. It is the move-based hedging strategy we investigate in this paper for the reasons given in the next section. Below is a detailed description of this strategy.

   Suppose that at time 0, we sell a put option with maturity T written on the VA sub-account whose value is P_0. To hedge, we buy Δ_0 units of the sub-account and invest M_0 in the money market account. Thus, Δ_0 is the delta of the put option at time 0 and M_0 is such that $\Delta_0 S_0 + M_0 = P_0$. A band $[S_0 e^{-\alpha}, S_0 e^{\alpha}]$ is set for the movement of the sub-account value. We will not re-balance our hedging portfolio until the value of the sub-account hits the band. At the hitting time (say τ), two actions take place:

(i) resetting the band: the band is reset to be $[S_\tau e^{-\alpha}, S_\tau e^{\alpha}]$, which centers around the new sub-account value;

(ii) re-balancing the hedging portfolio: we adjust the hedging portfolio to match the delta of the option at that time and infuse or take out cash such that the value of the hedging portfolio equals the value of the option. The cost incurred is thus the difference between the option value and the value of the hedging portfolio before re-balancing. This procedure is repeated whenever the value of the sub-account hits the band up to the time of maturity.

2. **The One-Sided Underlier-Based Hedging**: Some practitioners may think the re-balancing triggered by the upward move of the sub-account redundant because we only suffer from the downward move when shorting a put. As a result, they often employ a one-sided underlier-based hedging strategy, which is almost identical to the two-sided hedging strategy we mentioned above except that the upper barrier of the band is removed. More precisely, the initial band $[S_0 e^{-\alpha}, S_0 e^{\alpha}]$ in the two-sided underlier-based hedging is replaced by a single lower barrier $[S_0 e^{-\alpha}, +\infty)$ and at the hitting time (say τ) of this barrier, it is reset to be $[S_\tau e^{-\alpha}, +\infty)$.

3. **The Two-Sided Greek-based Hedging**: Under this strategy, we set a band for the delta of the option, Δ_t, instead of for the sub-account value S_t. Moreover, the band for Δ_t is not proportional to its current value but of the form $[\Delta_t - \alpha, \Delta_t + \alpha]$ (Since the theoretical value of the Δ of a put option always falls into $[-1, 0]$, we should cut off the part of the band that lies out of $[-1, 0]$. And when the value of the Δ comes very close to the boundary -1 and 0, the band should become one-sided. These boundary adjustments are very important because otherwise, the tail of the cost distribution would be much fatter, as seen in Table 3.2.1). The hedging portfolio is re-balanced if Δ_t hits the band. This is intuitively more direct than the

underlier-based hedging as what we effectively update during each re-balancing is the delta of the hedging portfolio.

3.2 A First Look at the Cost Distribution

In this section, we compare the cost distribution of the three hedging strategies introduced in the previous section via Monte Carlo simulation.

We assume the value of the underlying asset follow GBM

$$S_t = S_0 e^{(\mu - d - \frac{1}{2}\sigma^2)t + \sigma W_t}, \quad \forall t > 0, \tag{3.2.3}$$

where S_0 is the initial value of the sub-account, μ the mean return rate, d the management fee, σ the volatility, r the risk free interest rate, and W_t a standard Brownian motion under the real probability measure.

In detail, we generate 100,000 paths (with the time step of 0.0001) of the sub-account value, along each path we first identify the hitting times of the sub-account and then calculate the cost associated with each re-balancing (i.e. the present value of the difference between the then-current value of the put and the value of our hedging portfolio just before re-balancing, discounted at rate r), and finally sum them up to obtain the total cost for this path. This procedure produces a sample of size 100,000, which can be used to estimate important statistics and plot histograms of the true cost distribution. To make a fair comparison, we choose the bandwidth (level of the barrier) so that all the three strategies have roughly the same hedging frequency (around 100 re-balances in 3 years). Table 3.2.1 summarizes the simulation results.

It is obvious from Table 3.2.1 that the one-sided underlier-based hedging is the worst and should be discarded. The two-sided underlier-based hedging has a smaller standard

	Underlier-Based One-Sided	Underlier-Based Two-Sided	Greek-Based without Boundary Adjustments	Greek-Based with Boundary Adjustments
mean	3.5248	0.0023	0.0924	0.0365
std	13.1283	0.5005	0.8076	0.5474
skewness	2.8675	-0.4296	2.2852	0.1579
kurtosis	18.5249	4.9739	17.3893	3.3606
90% quantile	19.9861	0.5736	0.9120	0.7214
95% quantile	29.8750	0.7747	1.4118	0.9833
97.5% quantile	40.0551	0.9680	2.0590	1.1951
99% quantile	51.5242	1.2049	3.0543	1.4119

Table 3.2.1: Moments and quantiles of cost distribution under the 3 hedging strategies for an at-the-money put option. The model parameters are $T = 3, S_0 = K = 50, r = 0.02, \mu = 0.1, \sigma = 0.3, d = 0$. For the one-sided underlier-based hedging, $\alpha = 0.0015$; two-sided underlier-based hedging, $\alpha = 0.05$; Greek-based hedging with or without boundary adjustments, $\alpha = 0.045$. The row "mean" refers to the mean cost of discrete re-balance. For comparison purpose, we calculate the cost of continuous hedging to be 8.5598.

deviation than that of the Greek-based hedging but a larger kurtosis. Hence it is not clear at the moment which strategy is superior. However, a closer look at all the statistics will provide a clear picture that the two-sided underlier-based hedging is indeed superior: the kurtosis measures the overall heaviness of the tail and it does not differentiate the heaviness of the left tail from that of the right tail. While for an insurance company, the right tail is of central significance because it reflects the losses (the left tail, on the other hand, represents profits). The larger quantile values of the Greek-based hedging and its positive skewness, as opposed to the negative skewness under the two-sided underlier-based hedging, all suggest that the cost distribution under the Greek-based hedging has a longer right tail, while the cost distributions under the two-sided underlier-based hedging has a longer left tail.

This desirable feature provides us the rationale to further investigate the cost distribution under the two-sided underlier-based strategy (Another reason for the choice of the two-sided underlier-based strategy is the difficulty in finding the hitting time distribution for the delta of the put option. The nonlinearity of delta of the put option makes finding the hitting time distribution of the delta remain extremely challenging).

In the following, we conduct more simulation studies on the cost distribution of the

	$\mu = 0.05$ $\sigma = 0.2$	$\mu = 0.1$ $\sigma = 0.2$	$\mu = 0.15$ $\sigma = 0.2$	$\mu = 0.05$ $\sigma = 0.3$	$\mu = 0.1$ $\sigma = 0.3$	$\mu = 0.15$ $\sigma = 0.3$
mean	-0.0050	0.0333	0.1119	-0.0021	0.0049	0.0278
std	0.9947	0.9434	0.8541	1.0006	0.9927	0.9653
skewness	-0.6847	-0.7127	-0.7989	-0.5620	-0.5637	-0.6229
kurtosis	4.6133	4.8191	5.5205	4.7426	4.7170	5.0031

Table 3.2.2: Descriptive statistics for cost distribution. The common parameters are $S_0 = 50, K = 50, \alpha = 0.1, r = 0.02, T = 3, d = 0$. The row "mean" refers to the mean cost of discrete re-balance. For comparison purpose, we calculate the cost of continuous hedging to be 5.3183 when $\sigma = 0.2$ and 8.5598 when $\sigma = 0.3$.

two-sided underlier-based hedging by assigning 6 sets of parameter values to the GBM model (3.2.3). Descriptive statistics for these simulations are provided in Table 3.2.2. Histograms are displayed in Figure 7.1.1. These results suggest heavy and asymmetric tail behavior of the cost.

3.3 Three Density Functions for the Hitting Time Distribution of Geometric Brownian Motion

In this section, we derive three density functions that will prove to be very useful in the analysis of the two-sided underlier-based hedging.

We define the stopping time when the sub-account S starting from S_0 hits a two-sided band $[S_0 e^{-\alpha}, S_0 e^{\alpha}]$

$$\tau_{\alpha,-\alpha} = \inf\{t > 0 | S_t = S_0 e^{\alpha} \quad \text{or} \quad S_t = S_0 e^{-\alpha}\}, \tag{3.3.4}$$

and two auxiliary stopping times

$$\tau_{\alpha} = \inf\{t > 0 | S_t = S_0 e^{\alpha}, S_s > S_0 e^{-\alpha} (\forall 0 < s < t)\}, \tag{3.3.5}$$

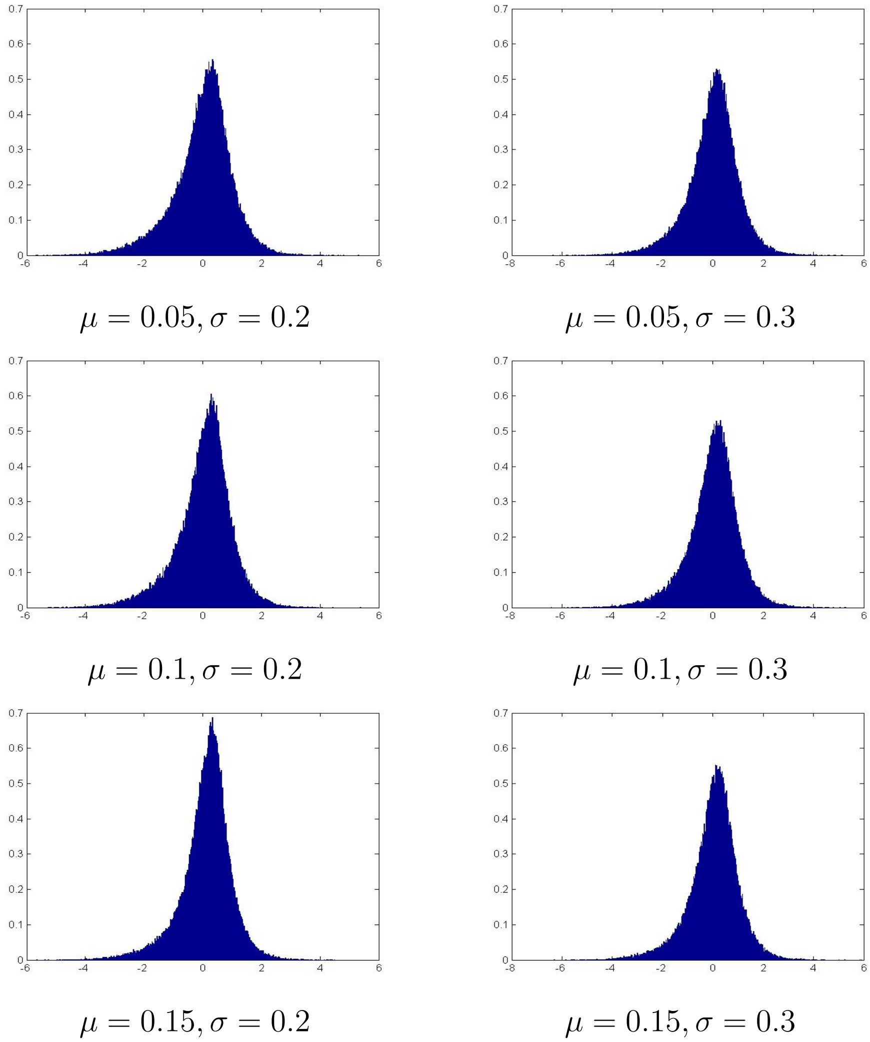


Figure 3.2.1: Histograms of cost distribution. The common parameters are $T = 3, r = 0.02, \alpha = 0.1, S_0 = 50, K = 50, d = 0$.

$$\tau_{-\alpha} = \inf\{t > 0 | S_t = S_0 e^{-\alpha}, S_s < S_0 e^{\alpha} (\forall 0 < s < t)\}. \tag{3.3.6}$$

In other words, τ_α is the first time that S hits $S_0 e^{\alpha}$ without hitting $S_0 e^{-\alpha}$ earlier and $\tau_{-\alpha}$ is the first time that S hits $S_0 e^{-\alpha}$ without hitting $S_0 e^{\alpha}$ earlier.

From the definitions, we can see: (1) $\tau_{\alpha,-\alpha} = \tau_\alpha \wedge \tau_{-\alpha}$; (2) $\tau_\alpha < \infty \Rightarrow \tau_{-\alpha} = \infty$ and $\tau_{-\alpha} < \infty \Rightarrow \tau_\alpha = \infty$.

We remark that under the two-sided underlier-based strategy, the bandwidth depends on the current value of the sub-account but the distribution of the three stopping times defined above does not. In fact, they are the exit times of $[-\alpha, \alpha]$ by a drifted Brownian motion $(\mu - d - \frac{1}{2}\sigma^2)t + \sigma W_t$, starting from 0.

Now we derive the densities of τ_α and $\tau_{-\alpha}$ in a slightly more general setting. Define $X(t) = vt + \sigma W_t$ and

$$
\begin{aligned}
T_{a,b} &= \inf\{t : X(t) = a \quad \text{or} \quad X(t) = b\}, \\
T_a &= \inf\{t : X(t) = a, X(s) < b(\forall 0 < s < t)\}, \\
T_b &= \inf\{t : X(t) = b, X(s) > a(\forall 0 < s < t)\},
\end{aligned}
$$

where $a < 0 < b$ and as before, $T_a(\text{or} \quad T_b) < \infty \Rightarrow T_b(\text{or} \quad T_a) = \infty$.
Following the approach of Lin (1998), we will first compute the Laplace transform of T_a and T_b using the Gerber-Shiu technique and then inverse the transform through rewriting them as a series of the Laplace transform of stable distribution. The Gerber-Shiu method makes use of the following exponential martingale

$$
Z_\lambda(t) = e^{\lambda X(t) - (\lambda v + \frac{1}{2}\lambda^2\sigma^2)t}. \tag{3.3.7}
$$

For any $u > 0$, we can find two values of λ s.t. $\lambda\mu + \frac{1}{2}\lambda^2\sigma^2 = u$,

$$
\lambda_1 = \frac{-v - \sqrt{v^2 + 2\sigma^2 u}}{\sigma^2} < 0, \quad \lambda_2 = \frac{-v + \sqrt{v^2 + 2\sigma^2 u}}{\sigma^2} > 0,
$$

and therefore obtain two martingales,

$$
\begin{aligned}
M_1(t) &= Z_{\lambda_1}(t) = e^{\lambda_1 X(t) - ut}, \\
M_2(t) &= Z_{\lambda_2}(t) = e^{\lambda_2 X(t) - ut}.
\end{aligned}
$$

Observe that for any fixed $\lambda_i \quad (i = 1, 2)$ and $u > 0$, $M_i(T_{a,b} \wedge t) < e^{\max\{|\lambda_i a|, |\lambda_i b|\}}$, so $M_i(T_{a,b} \wedge t)$ is a bounded martingale and we can apply the optional stopping theorem (See for example, Durrett (2010) Page 230) to get

$$E[M_i(T_{a,b})] = 1, \quad i = 1, 2. \tag{3.3.8}$$

More explicitly, for $i = 1$,

$$\begin{aligned} E[M_1(T_{a,b})] &= E[e^{\lambda_1 X(T_{a,b}) - uT_{a,b}}] \\ &= E[e^{\lambda_1 X(T_{a,b}) - uT_{a,b}} I_{\{T_{a,b}=T_a\}} + e^{\lambda_1 X(T_{a,b}) - uT_{a,b}} I_{\{T_{a,b}=T_b\}}] \\ &= E[e^{\lambda_1 a - uT_a} I_{\{T_{a,b}=T_a\}}] + E[e^{\lambda_1 b - uT_b} I_{\{T_{a,b}=T_b\}}]. \end{aligned} \tag{3.3.9}$$

But since $T_b < \infty \Rightarrow T_a = \infty$ and $u > 0$,

$$\begin{aligned} E[e^{\lambda_1 a - uT_a}] &= E[e^{\lambda_1 a - uT_a} I_{\{T_{a,b}=T_a\}} + e^{\lambda_1 a - uT_a} I_{\{T_{a,b}=T_b\}}] \\ &= E[e^{\lambda_1 a - uT_a} I_{\{T_{a,b}=T_a\}}]. \end{aligned}$$

(3.3.9) can be further simplified to

$$E[M_1(T_{a,b})] = E(e^{-uT_a})e^{a\lambda_1} + E(e^{-uT_b})e^{b\lambda_1} = 1, \tag{3.3.10}$$

and similarly for $i = 2$, we have

$$E[M_2(T_{a,b})] = E(e^{-uT_a})e^{a\lambda_2} + E(e^{-uT_b})e^{b\lambda_2} = 1. \tag{3.3.11}$$

From equation (3.3.10) and (3.3.10), we solve the Laplace transforms of T_a and T_b,

$$
\begin{aligned}
E(e^{-uT_a}) &= \frac{e^{b\lambda_2} - e^{b\lambda_1}}{e^{a\lambda_1+b\lambda_2} - e^{b\lambda_1+a\lambda_2}}, \\
E(e^{-uT_b}) &= \frac{e^{a\lambda_1} - e^{a\lambda_2}}{e^{a\lambda_1+b\lambda_2} - e^{b\lambda_1+a\lambda_2}}.
\end{aligned}
$$

Because

$$
\frac{1}{e^{a\lambda_1+b\lambda_2} - e^{b\lambda_1+a\lambda_2}} = e^{-a\lambda_1-b\lambda_2}\sum_{n=0}^{\infty} e^{-n(b-a)(\lambda_2-\lambda_1)},
$$

we have

$$
E(e^{-uT_a}) = \sum_{n=0}^{\infty} e^{-n(b-a)(\lambda_2-\lambda_1)-a\lambda_1} - \sum_{n=0}^{\infty} e^{-(n+1)(b-a)(\lambda_2-\lambda_1)-a\lambda_2}, \tag{3.3.12}
$$

$$
E(e^{-uT_b}) = \sum_{n=0}^{\infty} e^{-n(b-a)(\lambda_2-\lambda_1)-b\lambda_2} - \sum_{n=0}^{\infty} e^{-(n+1)(b-a)(\lambda_2-\lambda_1)-b\lambda_1}. \tag{3.3.13}
$$

In order to inverse the Laplace transforms (3.3.12) and (3.3.13), let $a_n = (1/\sigma)[2n(b-a)-a]$ and $b_n = (1/\sigma)[2n(b-a)+b]$ for $n \in \mathbf{Z}$, then the terms of the first series in (3.3.12) is

$$
\begin{aligned}
e^{-n(b-a)(\lambda_2-\lambda_1)-a\lambda_1} &= e^{\frac{av}{\sigma^2} - \frac{a_n v}{\sigma}\sqrt{1+\frac{2\sigma^2 u}{v^2}}} \\
&= e^{\frac{av}{\sigma^2} - \frac{a_n v}{\sigma}} e^{\frac{a_n v}{\sigma}[1-\sqrt{1+\frac{2\sigma^2 u}{v^2}}]}.
\end{aligned} \tag{3.3.14}
$$

Note that the Laplace transform of an Inverse Gaussian distribution with the shape parameter $\alpha > 0$, the scale parameter $\beta > 0$ and pdf

$$
f_{IG}(t) = \frac{\alpha}{\sqrt{2\pi\beta t^3}} e^{-\frac{1}{2\beta t}(\beta t-\alpha)^2}
$$

is

$$
\int_0^{\infty} e^{-ut} f_{IG}(t)dt = e^{\alpha[1-\sqrt{1+\frac{2u}{\beta}}]}. \tag{3.3.15}
$$

So by comparing (3.3.14) with (3.3.15), we conclude that inversion of the first series in (3.3.12) is

$$e^{\frac{av}{\sigma^2}-\frac{a_n v}{\sigma}}\frac{a_n}{\sqrt{2\pi t^3}}e^{\frac{(vt-a_n\sigma)^2}{2\sigma^2 t}} = e^{\frac{av}{\sigma^2}-\frac{1}{2}(\frac{v}{\sigma})^2 t}f(t;a_n),$$

where $f(t;a) = \frac{a}{\sqrt{2\pi t^3}}e^{-a^2/st}$ is the density of a one-sided stable distribution of index $\frac{1}{2}$. For the terms in the second series of (3.3.12), let $m = -n-1$, then

$$e^{-(n+1)(b-a)(\lambda_2-\lambda_1)-a\lambda_2} = e^{\frac{av}{\sigma^2}+\frac{a_m v}{\sigma}}e^{-\frac{a_m v}{\sigma}[1-\sqrt{1+\frac{2\sigma^2 u}{v^2}}]},$$

and its inversion is

$$-e^{\frac{av}{\sigma^2}+\frac{a_m v}{\sigma}}\frac{a_m}{\sqrt{2\pi t^3}}e^{\frac{(vt+a_m\sigma)^2}{2\sigma^2 t}} = -e^{\frac{av}{\sigma^2}-\frac{1}{2}(\frac{v}{\sigma})^2 t}f(t;a_m).$$

Finally, the inversion of (3.3.12), or the density of T_a, is

$$g_a(t) = e^{\frac{av}{\sigma^2}-\frac{1}{2}(\frac{v}{\sigma})^2 t}\sum_{n=-\infty}^{\infty}f(t;a_n). \tag{3.3.16}$$

A parallel argument lead to the density of T_b

$$g_a(t) = e^{\frac{bv}{\sigma^2}-\frac{1}{2}(\frac{v}{\sigma})^2 t}\sum_{n=-\infty}^{\infty}f(t;b_n). \tag{3.3.17}$$

The densities of τ_α and $\tau_{-\alpha}$, defined in (3.3.5) and (3.3.6) are given by $g_\alpha(t)$ and $g_{-\alpha}(t)$, with σ unchanged and v replaced by $\mu - d - \frac{1}{2}\sigma^2$.

Our next goal is to identify the distribution of the sub-accont at maturity T given it has not hit the band along the way. This distribution is characterized by the following probability density

$$P_{S_0}(S_T \in ds, \max_{0\le u\le T} S_u < S_0 e^{\alpha}, \min_{0\le u\le T} S_u > S_0 e^{-\alpha}), \tag{3.3.18}$$

which can be translated to a probability regarding only drifted Brownian motion

$$P_0(X_T \in dx, \max_{0 \le u \le T} X_u < \alpha, \min_{0 \le u \le T} X_u > -\alpha), \tag{3.3.19}$$

where $S_t = S_0 e^{X_t}, \forall t < T$ and $x = \log(\frac{s}{S_0})$.

When X_t is a standard Brownian motion, the above probability is given by the following theorem

Theorem 3.3.1 (Karatzas and Shreve (1991) 8.10 Proposition)**.** *Choose* $0 < x < a$. *Then for* $t > 0, 0 < y < a$*:*

$$P_x[W_t \in dy, T_0 \wedge T_a > t] = \sum_{n=-\infty}^{\infty} p_-(t; x, y + 2na)dy, \tag{3.3.20}$$

where W_t *is standard Brownian motion with* $X_0 = x$*,* T_0 *and* T_a *are the hitting time of* 0 *and* a *respectively,* $p_-(t; x, y) = p(t; x, y) - p(t; x, -y), p(t; x, y) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{(x-y)^2}{2t}}$.

We now generalize this theorem to the case of drifted Brownian motion $X_t = vt + \sigma W_t$. In spirit of the Girsanov Theorem (Karatzas and Shreve (1991) Page 191 5.1 Theorem), we introduce a new measure $\tilde{P}$ through

$$Z_t(W) = E[\frac{d\tilde{P}}{dP}|\mathcal{F}_t] = \exp\{-qW_t - \frac{1}{2}q^2 t\}, t < T$$

where $q = \frac{v}{\sigma}$. Then $\frac{dZ_t(W)}{Z_t(W)} = -qdW_t$ and therefore, the process $\tilde{W}_t$ defined by $\tilde{W}_t = W_t + qt$ is a standard Brownian motion under $\tilde{P}$ and $X_t = vt + \sigma W_t = \sigma \tilde{W}_t$.

According to Theorem 3.3.1,

$$\begin{aligned}
&\tilde{P}_{X_0=x}(X_T \in dy, \max_{0 \le u \le T} X_u < a, \min_{0 \le u \le T} X_u > 0) \\
&= \tilde{P}_{\tilde{W}_0=\frac{x}{\sigma}}(\tilde{W}_T \in d(\frac{y}{\sigma}), \max_{0 \le u \le T} \tilde{W}_u < \frac{a}{\sigma}, \min_{0 \le u \le T} \tilde{W}_u > 0) \\
&= \sum_{n=-\infty}^{\infty} p_-(t; \frac{x}{\sigma}, \frac{y}{\sigma} + 2n\frac{a}{\sigma})\frac{1}{\sigma}dy.
\end{aligned}$$

Applying the Bayes formula, we have, for $0 < x, y < a$,

$$\begin{aligned}
&P_{X_0=x}(X_T \in dy, \max_{0\le u\le T} X_u < a, \min_{0\le u\le T} X_u > 0) \\
&= E^{P}_{X_0=x}[I_{\{X_T\in dy, \max_{0\le u\le T} X_u<a, \min_{0\le u\le T} X_u>0\}}] \\
&= E^{\tilde{P}}_{X_0=x}[I_{\{X_T\in dy, \max_{0\le u\le T} X_u<a, \min_{0\le u\le T} X_u>0\}}(\frac{dP}{d\tilde{P}})_T] \\
&= E^{\tilde{P}}_{X_0=x}[I_{\{X_T\in dy, \max_{0\le u\le T} X_u<a, \min_{0\le u\le T} X_u>0\}}e^{\frac{q}{\sigma}X_T-\frac{1}{2}q^2T}] \\
&= E^{\tilde{P}}_{X_0=x}[I_{\{X_T\in dy, \max_{0\le u\le T} X_u<a, \min_{0\le u\le T} X_u>0\}}e^{\frac{q}{\sigma}y-\frac{1}{2}q^2T}] \\
&= e^{\frac{q}{\sigma}y-\frac{1}{2}q^2T}\sum_{n=-\infty}^{\infty} p_-(T;\frac{x}{\sigma},\frac{y}{\sigma}+2n\frac{a}{\sigma})\frac{1}{\sigma}dy \\
&= e^{\frac{v}{\sigma^2}y-\frac{1}{2}(\frac{v}{\sigma})^2T}\sum_{n=-\infty}^{\infty} p_-(T;\frac{x}{\sigma},\frac{y}{\sigma}+2n\frac{a}{\sigma})\frac{1}{\sigma}dy.
\end{aligned}$$

By the shift invariance property of the drifted Brownian motion, we have, for $x_1 < 0, x < x_2$,

$$\begin{aligned}
&P_{X_0=0}(X_T \in dx, \max_{0\le u\le T} X_u < x_2, \min_{0\le u\le T} X_u > x_1) \\
&= e^{\frac{v}{\sigma^2}(x-x_1)-\frac{1}{2}(\frac{v}{\sigma})^2T}\sum_{n=-\infty}^{\infty} p_-(T;\frac{-x_1}{\sigma},\frac{x-x_1}{\sigma}+2n\frac{x_2-x_1}{\sigma})\frac{1}{\sigma}dy \\
&= e^{\frac{v}{\sigma^2}(x-x_1)-\frac{1}{2}(\frac{v}{\sigma})^2T}\sum_{n=-\infty}^{\infty} \frac{1}{\sigma\sqrt{T}}\left[\phi(\frac{x+2n(x_2-x_1)}{\sigma\sqrt{T}}) - \phi(\frac{x+2n(x_2-x_1)-2x_1}{\sigma\sqrt{T}})\right],
\end{aligned} \tag{3.3.21}$$

Where $\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$. The above result also appears in He et al. (1998) (with a different proof).

3.4 An Asymptotic Analysis for the Time-Based Hedging

In this section, we modify the main result of Boyle and Emanuel (1980) for the case of the European put option to provide an asymptotic analysis for the cost distribution of the time-based hedging.

Under the time-based hedging strategy, we choose N equally spaced time points in $[0, T]$ and re-balance our replicating portfolio at these time points to match the then-current value of the put option. Denote by Δt the time elapse between two consecutive re-balances.

Boyle and Emanuel (1980) looks at the conditional distribution of the next re-balancing cost, given the current state of the underlying asset. Specifically, suppose we have just re-balanced the hedging portfolio at time $t = m\Delta t$ such that it comprises of $-N(-d_1)$ units of the underlying asset and the value of $p + N(-d_1)S$ in the money market account earning interest at risk-free rate r, where S is the current value of the underlier, $p = Xe^{-r\tau}N(-d_2) - SN(-d_1)$ is the current price of the put option, X is the strike price, $d_1 = \frac{\log(\frac{S}{X}) + (r + \frac{1}{2}\sigma^2)\tau}{\sigma\sqrt{\tau}}$ and $d_2 = d_1 - \sigma\sqrt{\tau}$. Suppose at time $t + \Delta t$, the value of the underlying asset moves to $S + \Delta S$ and option price changes to $p + \Delta p$. So the re-balance at $t + \Delta t$ bring us a cost

$$\begin{aligned} HR &= (p + \Delta p) - [-N(-d_1)(S + \Delta S) + Xe^{-r\tau}N(-d_2)(1 + r\Delta t)] + o(\Delta t^2) \\ &= \Delta p + \Delta S N(-d_1) - Xe^{-r\tau}r\Delta t N(-d_2) + o(\Delta t^2) \end{aligned} \tag{3.4.22}$$

To further simplify the above equation, assume the underlying asset follows GBM with volatility σ, then

$$\Delta p = \frac{\partial p}{\partial S}\Delta S + \frac{\partial p}{\partial t}\Delta t + \frac{1}{2}\frac{\partial^2 p}{\partial S^2}\sigma^2 S^2 u^2 \Delta t + o(\Delta t^{\frac{3}{2}}),$$

$$
\begin{aligned}
\frac{\partial p}{\partial S} &= -N(-d_1), \\
\frac{\partial p}{\partial t} &= Xe^{-r\tau}[rN(-d_2) - \frac{\phi(d_2)\sigma}{2\sqrt{\tau}}], \\
\frac{\partial^2 p}{\partial S^2} &= \frac{\phi(d_1)}{S\sigma\sqrt{\tau}}, \\
S\phi(d_1) &= Xe^{-r\tau}\phi(d_2).
\end{aligned}
$$

where u is a standard normal r.v. and $\phi(x)$ is the standard normal density.

Using the above identities and ignoring the high order terms of Δt, we can reduce (3.4.22) to

$$
\begin{aligned}
HR &= \frac{Xe^{-r\tau}}{2\sqrt{\tau}}\phi(d_1)(u^2 - 1)\Delta t, \\
&= \lambda y \Delta t.
\end{aligned} \tag{3.4.23}
$$

where $\lambda = \frac{Xe^{-r\tau}}{2\sqrt{\tau}}\phi(d_1)$ and $y = u^2 - 1$.

From (3.4.23) we see that the conditional expectation of the next re-balancing cost HR is 0 and the conditional standard deviation, skewness and kurtosis moments are $\sqrt{2}\lambda\Delta t$, $\frac{4}{\sqrt{2}}$, 15, respectively. So the conditional distribution is positively skewed with a very heavy tail. And its standard deviation exhibit an asymptotically linear relation with Δt. Moreover, HR is negative (positive) if and only if $|u| < 1$ ($|u| > 1$), which means our hedge will yield negative (positive) returns about 68% (34%) of the time.

Chapter 4

A Semi-Analytic Algorithm for the Cost Analysis of Put Option

In this chapter, we first develop semi-analytic algorithms to compute the expectation and the higher moments of the cost of the two-sided underlier-based hedging. Then using the methods of moments, we fit a parametric model for the cost distribution to its first 4 moment. Finally with the fitted distribution, we approximate the quantile of the total hedging cost, which is an key input in the implementation of the modified "Percentile Premium Principle" for variable annuities.

4.1 The Main Idea

Recall that in the two-sided underlier-based hedging strategy, each rebalance is triggered by the hit to a band. We denote by $\tau^{(i)}, (i = 1, 2, 3, ...)$ the timespan from the $(i-1)$-th hit to the i-th hit. Then by the strong Markov property and the stationarity of Brownian motion, $\{\tau^{(i)}\}_{i=1}^{\infty}$ are i.i.d. random variables with the same distribution as $\tau_{\alpha,-\alpha}$ defined in (3.3.4).

Let us consider the cost incurred at the first hit. If the first hit occurs before the maturity (i.e. $\tau^{(1)} < T$), we adjust our portfolio to match the option value at time $\tau^{(1)}$. Otherwise, the value of the sub-account never hit the band throughout the life of the option ($\tau^{(1)} \geq T$) and we simply close our position at maturity T. In other words, the first cost will occur at the truncated stopping time $\tau^{(1)} \wedge T$. If the first hit occurs before maturity($T > \tau^{(1)}$), then there would be a second cost, and the timespan from the first hit to the second hit is again a truncated stopping time $\tau^{(2)} \wedge (T - \tau^{(1)})$.

Now look at the two truncated stopping times

$$\tau^{(1)} \wedge T, \quad \tau^{(2)} \wedge (T - \tau^{(1)}).$$

If we replace the fixed maturity T by $\epsilon_\lambda^{(1)}$, an independent exponential r.v. with parameter λ, then by the strong Markov property and the stationarity of Brownian motion and the memoryless property of exponential distribution, the distribution of $\tau^{(2)} \wedge (\epsilon_\lambda^{(1)} - \tau^{(1)})$ will be exactly the same as that of $\tau^{(1)} \wedge \epsilon_\lambda^{(1)}$ conditional on $\mathcal{F}_{\tau^{(1)}}$ and $\epsilon_\lambda^{(1)} > \tau^{(1)}$. The use of random maturity for option pricing was pioneered in Carr (1998). This randomization allows us to derive a recursive formula for computing the cost at each truncated stopping time.

4.2 The Expected Cost

To keep the notation compact, we use $\bar{\tau}^{(i)}$ to denote the i-th truncated stopping time, i.e.

$$\bar{\tau}^{(i)} = \tau^{(i)} \wedge (\epsilon_\lambda^{(1)} - \tau^{(1)} - \cdots - \tau^{(i-1)}) = \tau^{(i)} \wedge \epsilon_\lambda^{(i)}.$$

Recall that the first cost is the difference between of the time-$\bar{\tau}^{(1)}$ value of the option and the hedging portfolio before re-balancing, discounted back to time 0 at the risk free interest rate:

$$e^{-r\bar{\tau}^{(1)}}\left[P_{\bar{\tau}^{(1)}}-\left(M_0e^{r\bar{\tau}^{(1)}}+\Delta_0S_{\bar{\tau}^{(1)}}\right)\right]. \tag{4.2.1}$$

With the techniques developed in Section 3.3, we are able to compute the expectation of (4.2.1). The essential steps are listed below.

First, we rewrite (4.2.1) as

$$\underbrace{I_{\epsilon_\lambda^{(1)}>\tau^{(1)}}\left\{Ke^{-r\epsilon_\lambda^{(1)}}\left[N(-\tilde{d}_2)-N(-d_2)\right]-S_{\tau^{(1)}}e^{-r\tau^{(1)}}e^{-d(\epsilon_\lambda^{(1)}-\tau_{(1)})}\left[N(-\tilde{d}_1)-N(-d_1)\right]\right\}}_{A_1} \tag{4.2.2}$$

$$+\underbrace{I_{\epsilon_\lambda^{(1)}\leq\tau^{(1)}}e^{-r\epsilon_\lambda^{(1)}}\left[(K-S_{\epsilon_\lambda^{(1)}})_+-KN(-d_2)+S_{\epsilon_\lambda^{(1)}}N(-d_1)\right]}_{A_2},$$

where

$$\begin{aligned}
d_1&=\frac{ln\frac{S_0}{K}+(r-d+\frac{1}{2}\sigma^2)\epsilon_\lambda^{(1)}}{\sigma\sqrt{\epsilon_\lambda^{(1)}}},\\
d_2&=d_1-\sigma\sqrt{\epsilon_\lambda^{(1)}},\\
\tilde{d}_1&=\frac{ln\frac{S_{\tau^{(1)}}}{K}+(r-d+\frac{1}{2}\sigma^2)(\epsilon_\lambda^{(1)}-\tau^{(1)})}{\sigma\sqrt{\epsilon_\lambda^{(1)}-\tau^{(1)}}},\\
\tilde{d}_2&=\tilde{d}_1-\sigma\sqrt{\epsilon_\lambda^{(1)}-\tau^{(1)}},
\end{aligned}$$

and abbreviate A_1 to $I_{\epsilon_\lambda^{(1)}>\tau^{(1)}}q(\epsilon_\lambda^{(1)},\tau^{(1)})$.

Then conditioning on $\epsilon_\lambda^{(1)}$, the conditional expectation of A_1 is

$$\begin{aligned}
&E_{\epsilon_\lambda^{(1)}}[I_{\epsilon_\lambda^{(1)}>\tau^{(1)}}q(\epsilon_\lambda^{(1)},\tau^{(1)})] \qquad (4.2.3)\\
&=E_{\epsilon_\lambda^{(1)}}[I_{\epsilon_\lambda^{(1)}>\tau^{(1)}}q(\epsilon_\lambda^{(1)},\tau^{(1)})I_{\tau^{(1)}=\tau_\alpha}]+E_{\epsilon_\lambda^{(1)}}[I_{\epsilon_\lambda^{(1)}>\tau^{(1)}}q(\epsilon_\lambda^{(1)},\tau^{(1)})I_{\tau^{(1)}=\tau_{-\alpha}}].
\end{aligned}$$

Because $\tau_\alpha = \infty$ when $\tau^{(1)} = \tau_{-\alpha}$, the first term on the RHS of (4.2.3) is

$$
\begin{aligned}
& E_{\epsilon_\lambda^{(1)}}[I_{\epsilon_\lambda^{(1)} > \tau_\alpha} q(\epsilon_\lambda^{(1)}, \tau_\alpha) I_{\tau^{(1)} = \tau_\alpha}] \\
&= E_{\epsilon_\lambda^{(1)}}[I_{\epsilon_\lambda^{(1)} > \tau_\alpha} q(\epsilon_\lambda^{(1)}, \tau_\alpha) I_{\tau^{(1)} = \tau_\alpha}] + E_{\epsilon_\lambda^{(1)}}[I_{\epsilon_\lambda^{(1)} > \tau_\alpha} q(\epsilon_\lambda^{(1)}, \tau_\alpha) I_{\tau^{(1)} = \tau_{-\alpha}}] \\
&= E_{\epsilon_\lambda^{(1)}}[I_{\epsilon_\lambda^{(1)} > \tau_\alpha} q(\epsilon_\lambda^{(1)}, \tau_\alpha)].
\end{aligned}
$$

Similarly, the second term on the RHS of (4.2.3) is:

$$
E_{\epsilon_\lambda^{(1)}}[I_{\epsilon_\lambda^{(1)} > \tau_{-\alpha}} q(\epsilon_\lambda^{(1)}, \tau_{-\alpha})].
$$

Conditioning on $\epsilon_\lambda^{(1)}$, the conditional expectation of A_2 is

$$
E_{\epsilon_\lambda^{(1)}} \left\{ I_{\epsilon_\lambda^{(1)} \leq \tau^{(1)}} e^{-r\epsilon_\lambda^{(1)}} \left[(K - S_{\epsilon_\lambda^{(1)}})_+ - KN(-d_2) + S_{\epsilon_\lambda^{(1)}} N(-d_1) \right] \right\}.
$$

To compute the above three conditional expectations, we need three key inputs: (1) the density of τ_α; (2) the density of $\tau_{-\alpha}$; (3) the conditional density of $X_t = (\mu - \frac{1}{2}\sigma^2)t + \sigma W_t$ given $\min_{0 \leq u \leq t} X_u > -\alpha, \max_{0 \leq u \leq t} X_u < \alpha$, all of which have been obtained in Section 3.3.

Hereafter, we denote the expectation of the first hitting cost (4.2.1) by $h(S_0)$.

The second cost, which incurs at $\tau^{(1)} + \bar{\tau}^{(2)}$, can be expressed in a similar way

$$
I_{\epsilon_\lambda^{(1)} > \tau^{(1)}} \; e^{-r(\tau^{(1)} + \bar{\tau}^{(2)})} \left[P_{\tau^{(1)} + \bar{\tau}^{(2)}} - \left(M_{\tau^{(1)}} e^{r\bar{\tau}^{(2)}} + \Delta_{\tau^{(1)}} S_{\tau^{(1)} + \bar{\tau}^{(2)}} \right) \right], \tag{4.2.4}
$$

The additional indicator function $I_{\epsilon_\lambda^{(1)} > \tau^{(1)}}$ in (4.2.4) reflects the fact that the second cost incurs if and only if the first hit occurs before the maturity.

To compute the expectation of (4.2.4), we first condition on $\mathcal{F}_{\tau^{(1)}}$ and $\epsilon_\lambda^{(1)} > \tau^{(1)}$. The inner expectation is

$$e^{-(r+\lambda)\tau^{(1)}} E_{\tau^{(1)}} \left[P_{\tau^{(1)}+\bar{\tau}^{(2)}} - \left(M_{\tau^{(1)}} e^{r\bar{\tau}^{(2)}} + \Delta_{\tau^{(1)}} S_{\tau^{(1)}+\bar{\tau}^{(2)}} \right) | \epsilon_\lambda^{(1)} > \tau^{(1)} \right], \tag{4.2.5}$$

where we have used the identity $E(XI_A) = E(X|A)P(A)$ for any random variable X and any event A.

Note that, by conditioning on $\mathcal{F}_{\tau^{(1)}}$ and $\epsilon_\lambda^{(1)} > \tau^{(1)}$, $\bar{\tau}^{(2)}$ has exactly the same distribution as $\bar{\tau}^{(1)}$. Thus, the expectation in (4.2.5) is $h(S_{\tau^{(1)}})$.

In general, the conditional expectation (conditioning on $\mathcal{F}_{\tau^{(1)}+\dots+\tau^{(n-1)}}$ and $\epsilon_\lambda^{(1)} > \tau^{(1)} + \cdots + \tau^{(n-1)}$) for the n-th hit is

$$h(S_{\tau^{(1)}+\tau^{(2)}+\dots+\tau^{(n-1)}}) e^{-(r+\lambda)(\tau^{(1)}+\tau^{(2)}+\dots+\tau^{(n-1)})}. \tag{4.2.6}$$

We now demonstrate the computation of the unconditional expectation of (4.2.6) by a simple example with $n = 3$.

When $n = 3$, the expectation becomes

$$E\left[h(S_{\tau^{(1)}+\tau^{(2)}}) e^{-(r+\lambda)(\tau^{(1)}+\tau^{(2)})} \right]. \tag{4.2.7}$$

From a computational point of view, since the function $h(\cdot)$ is computed numerically, we should try to avoid its frequent evaluation. This is achieved by observing that $S_{\tau^{(1)}+\tau^{(2)}}$ can only take 3 different values: $S_0e^{-2\alpha}$, S_0 and $S_0e^{2\alpha}$. So we can decompose the expectation into 3 parts and plug $h(\cdot)$ out. In particular,

$$\begin{aligned} E & \quad [h(S_{\tau^{(1)}+\tau^{(2)}}) e^{-(r+\lambda)(\tau^{(1)}+\tau^{(2)})}] \\ = & \quad h(S_0e^{2\alpha}) E\left[e^{-(r+\lambda)(\tau^{(1)}+\tau^{(2)})} I_{\{S_{\tau^{(1)}+\tau^{(2)}} = S_0 e^{2\alpha}\}} \right] \\ + & \quad h(S_0) E\left[e^{-(r+\lambda)(\tau^{(1)}+\tau^{(2)})} I_{\{S_{\tau^{(1)}+\tau^{(2)}} = S_0\}} \right] \\ + & \quad h(S_0e^{-2\alpha}) E\left[e^{-(r+\lambda)(\tau^{(1)}+\tau^{(2)})} I_{\{S_{\tau^{(1)}+\tau^{(2)}} = S_0 e^{-2\alpha}\}} \right]. \end{aligned}$$

The remaining expectation terms in the above expression can be written explicitly. In what follows, we will calculate $E\left[e^{-(r+\lambda)(\tau^{(1)}+\tau^{(2)})}I_{\{S_{\tau^{(1)}+\tau^{(2)}}=S_0\}}\right]$. The calculation of the other two expectations is almost identical.

$$
\begin{aligned}
E & \left[e^{-(r+\lambda)(\tau^{(1)}+\tau^{(2)})}I_{\{S_{\tau^{(1)}+\tau^{(2)}}=S_0\}}\right] \\
= & \ E\left[e^{-(r+\lambda)(\tau^{(1)}+\tau^{(2)})}I_{\{\tau^{(1)}=\tau_\alpha,\tau^{(2)}=\tilde{\tau}_{-\alpha}\}}\right]+E\left[e^{-(r+\lambda)(\tau^{(1)}+\tau^{(2)})}I_{\{\tau^{(1)}=\tau_{-\alpha},\tau^{(2)}=\tilde{\tau}_{\alpha}\}}\right] \\
= & \ E\left[e^{-(r+\lambda)\tau_\alpha}e^{-(r+\lambda)\tilde{\tau}_{-\alpha}}I_{\{\tau^{(1)}=\tau_\alpha\}}I_{\{\tau^{(2)}=\tilde{\tau}_{-\alpha}\}}\right]+E\left[e^{-(r+\lambda)\tau_{-\alpha}}e^{-(r+\lambda)\tilde{\tau}_{\alpha}}I_{\{\tau^{(1)}=\tau_{-\alpha}\}}I_{\{\tau^{(2)}=\tilde{\tau}_{\alpha}\}}\right] \\
= & \ E\left[e^{-(r+\lambda)\tau_\alpha}I_{\{\tau^{(1)}=\tau_\alpha\}}\right]E\left[e^{-(r+\lambda)\tilde{\tau}_{-\alpha}}I_{\{\tau^{(2)}=\tilde{\tau}_{-\alpha}\}}\right] \\
& \quad +E\left[e^{-(r+\lambda)\tau_{-\alpha}}I_{\{\tau^{(1)}=\tau_{-\alpha}\}}\right]E\left[e^{-(r+\lambda)\tilde{\tau}_{\alpha}}I_{\{\tau^{(2)}=\tilde{\tau}_{\alpha}\}}\right] \\
= & \ 2E\left[e^{-(r+\lambda)\tau_{-\alpha}}I_{\{\tau^{(1)}=\tau_{-\alpha}\}}\right]E\left[e^{-(r+\lambda)\tau_{\alpha}}I_{\{\tau^{(1)}=\tau_{\alpha}\}}\right],
\end{aligned}
$$

where τ_α, $\tau_{-\alpha}$ are defined in (3.3.5) and (3.3.6), respectively. $\tilde{\tau}_{\pm\alpha}$ is an independent copy of $\tau_{\pm\alpha}$.

By definition, $\tau_{-\alpha}=\infty$ $(\tau_\alpha=\infty)$ when $\tau_\alpha<\infty$ $(\tau_{-\alpha}<\infty)$. This property allows a further simplification of the above equation. In fact,

$$
\begin{aligned}
L_\alpha = E\left[e^{-(r+\lambda)\tau_\alpha}\right] &= E\left[e^{-(r+\lambda)\tau_\alpha}I_{\{\tau^{(1)}=\tau_\alpha\}}\right]+E\left[e^{-(r+\lambda)\tau_\alpha}I_{\{\tau^{(1)}=\tau_{-\alpha}\}}\right] \\
&= E\left[e^{-(r+\lambda)\tau_\alpha}I_{\{\tau^{(1)}=\tau_\alpha\}}\right],
\end{aligned}
$$

and similarly

$$L_{-\alpha}=E\left[e^{-(r+\lambda)\tau_{-\alpha}}\right]=E\left[e^{-(r+\lambda)\tau_{-\alpha}}I_{\{\tau^{(1)}=\tau_{-\alpha}\}}\right].$$

Hence,

$$E\left[e^{-(r+\lambda)(\tau^{(1)}+\tau^{(2)})}I_{\{S_{\tau^{(1)}+\tau^{(2)}}=S_0\}}\right]=2L_{-\alpha}L_\alpha.$$

Therefore, for the unconditional expectation (4.2.7) we have

$$E\left[h(S_{\tau^{(1)}+\tau^{(2)}})e^{-(r+\lambda)(\tau^{(1)}+\tau^{(2)})}\right]=h(S_0e^{2\alpha})L_\alpha^2+2h(S_0)L_\alpha L_{-\alpha}+h(S_0e^{-2\alpha})L_{-\alpha}^2. \quad (4.2.8)$$

In general, the expectation of the $(n+1)$-th cost is

$$E\left[h(S_{\tau^{(1)}+\tau^{(2)}+\ldots+\tau^{(n)}})e^{-(r+\lambda)(\tau^{(1)}+\tau^{(2)}+\ldots+\tau^{(n)})}\right] = \sum_{i=0}^{n} C_n^i h(S_0 e^{(2i-n)\alpha}) L_\alpha^i L_{-\alpha}^{n-i}. \quad (4.2.9)$$

The total expected cost is obtained by summing up all the individual cost.

In summary, we have derived the expected total cost for an exponential maturity. In order to find the expected total cost for a fixed maturity, the following connection is useful.

Let $g(T)$ be the total expected cost for an arbitrary fixed maturity T. Then the expectation under an exponential maturity with parameter λ is

$$E[g(\epsilon_\lambda^{(1)})] = \int_0^\infty g(T)\lambda e^{-\lambda T} dT.$$

Divide both sides by λ, we have

$$\frac{E[g(\epsilon_\lambda^{(1)})]}{\lambda} = \int_0^\infty g(T) e^{-\lambda T} dT. \quad (4.2.10)$$

Clearly, the right hand side of (4.2.10) is the Laplace transform of $g(T)$. So the total expected cost for fixed maturity can be retrieved from that for exponential maturities by numerically inverting the Laplace transform.

There exist many routines for the numerical inversion of the Laplace Transform. We find Hollenbeck (1998) and Brancik (2011) particularly useful. The former can return the value of the original function at several points in the time domain in a single run and the latter can invert a matrix of transformed functions simultaneously.

4.3 Higher Moments

Our next step towards the distribution of the total hedging cost, is to compute its higher moments with which we may estimate the cost distribution using a parametric distribution.

The algorithm introduced in Section 4.2 can be easily adapted to compute the second moment of each individual cost. To begin with, let us look at the square of the first cost

$$e^{-2r\bar{\tau}^{(1)}}\left[P_{\bar{\tau}^{(1)}}-\left(M_0 e^{r\bar{\tau}^{(1)}}+\Delta_0 S_{\bar{\tau}^{(1)}}\right)\right]^2. \tag{4.3.11}$$

Denote by $h^{(2)}(S_0)$ the expectation of (4.3.11). $h^{(2)}(\cdot)$ is evaluated through a slightly modified procedure for computing $h(\cdot)$: (4.3.11) can be written as $(A_1+A_2)^2$, where A_1 and A_2 appear in (4.2.2). Since A_1 and A_2 are mutually exclusive, i.e. $A_1A_2=0$, we have $(A_1+A_2)^2=A_1^2+A_2^2$. The expectation of A_1^2 and A_2^2 can be computed using numerical integration.

Given $h^{(2)}(\cdot)$, we now move to the square of the second cost

$$\begin{aligned}
&I_{\epsilon_\lambda^{(1)}>\tau^{(1)}}e^{-2r(\tau^{(1)}+\bar{\tau}^{(2)})}\left[P_{\tau^{(1)}+\bar{\tau}^{(2)}}-\left(M_{\tau^{(1)}}e^{\bar{\tau}^{(2)}}+\Delta_{\tau^{(1)}}S_{\tau^{(1)}+\bar{\tau}^{(2)}}\right)\right]^2\\
&=e^{-2r\tau^{(1)}}I_{\epsilon_\lambda^{(1)}>\tau^{(1)}}e^{-2r\bar{\tau}^{(2)}}\left[P_{\tau^{(1)}+\bar{\tau}^{(2)}}-\left(M_{\tau^{(1)}}e^{\bar{\tau}^{(2)}}+\Delta_{\tau^{(1)}}S_{\tau^{(1)}+\bar{\tau}^{(2)}}\right)\right]^2.
\end{aligned}$$

Its expectation is computed by first conditioning on $\mathcal{F}_{\tau^{(1)}}$ and $\epsilon_\lambda^{(1)}>\tau^{(1)}$ to get the inner expectation

$$e^{-(2r+\lambda)\tau^{(1)}}E_{\tau^{(1)}}\left(e^{-2r\bar{\tau}^{(2)}}\left[P_{\tau^{(1)}+\bar{\tau}^{(2)}}-\left(M_{\tau^{(1)}}e^{\bar{\tau}^{(2)}}+\Delta_{\tau^{(1)}}S_{\tau^{(1)}+\bar{\tau}^{(2)}}\right)\right]^2|\epsilon_\lambda^{(1)}>\tau_{(1)}\right).$$

The expectation term above can be easily seen to be $h^{(2)}(S_{\tau^{(1)}})$. As a result, the conditional expectation of the square of the second cost is $e^{-(2r+\lambda)\tau^{(1)}}h^{(2)}(S_{\tau^{(1)}})$.

In general, the conditional expectation of the square of the $(n+1)$-th cost has the following form

$$e^{-(2r+\lambda)(\tau^{(1)}+\cdots+\tau^{(n)})}h^{(2)}(S_{\tau^{(1)}+\cdots+\tau^{(n)}}),$$

and its full expectation is given by

$$\sum_{i=0}^{i=n} C_n^i h^{(2)}(S_0 e^{(2i-n)\alpha})\tilde{L}_\alpha^i \tilde{L}_{-\alpha}^{n-i},$$

where $\tilde{L}_\alpha = E(e^{-(2r+\lambda)\tau_\alpha})$ and $\tilde{L}_{-\alpha} = E(e^{-(2r+\lambda)\tau_{-\alpha}})$.
In a nearly effortless manner, we could continue this way to calculate the 3rd and 4th moments of each individual costs. The details are omitted here.

Now suppose we are interested in the moments of the total cost and have at hand the first to the fourth moments of each individual cost. As a matter of fact, the full expansion of the second, third and fourth moments of the total cost involve not only the moments of individual costs, but also the interaction terms. The computation of the latter is difficult in general and computationally expensive. However, by assuming independence of the individual costs, we can circumvent this difficulty with the aid of multinomial expansion theorem (This is a reasonable assumption. Indeed, numerical experiments conducted in Section 4.4 indicate very weak correlations between individual costs). The moments of the total cost are then

$$\begin{aligned} E(C_1 + C_2 + \cdots + C_m)^n &= \sum_{k_1+k_2+\cdots+k_m=n} \frac{n!}{k_1!k_2!\cdots k_m!} E(\prod_{1\le t\le m} C_t^{k_t}) \\ &= \sum_{k_1+k_2+\cdots+k_m=n} \frac{n!}{k_1!k_2!\cdots k_m!} \prod_{1\le t\le m} E(C_t^{k_t}), \end{aligned} \tag{4.3.12}$$

where C_i is the discounted cost for the i-th hit and $n = 2, 3, 4$. Though straightforward, the multinomial expansion approach can be very inefficient numerically, because the number of terms in (4.3.12) increases dramatically with m. So we now introduce an alternative recursive method to compute the first 4 moments of the total cost under the independence assumption.

Let $C = C_1 + C_2 + \cdots + C_m$ be the total (discounted) cost and $M_i = E(C^i)$ be its i-th raw moment. The recursive method exploits the central moments. Suppose we have obtained the central moments of each individual cost. Under the independence assumption, the second central moment of the total cost (i.e. the variance) is

$$Var(C) = Var(\sum_{i=1}^{m} C_i) = \sum_{i=1}^{m} Var(C_i).$$

So $M_2 = E(C^2) = Var(C) + E^2(C) = \sum_{i=1}^{m} Var(C_i) + M_1^2$.

The third central moment of C is

$$E\left([C - E(C)]^3\right) = E\left[\left(\sum_{i=1}^{m}[C_i - E(C_i)]\right)^3\right] = \sum_{i=1}^{m} E\left([C_i - E(C_i)]^3\right)$$

and therefore

$$\begin{aligned} M_3 = E(C^3) &= E\left([C - E(C)]^3\right) + 3E(C^2)E(C) - 3E(C)^3 + E(C)^3 \\ &= \sum_{i=1}^{m} E\left([C_i - E(C_i)]^3\right) + 3M_2M_1 - 2M_1^3. \end{aligned}$$

The fourth central moment of C is

$$\begin{aligned} E\left([C - E(C)]^4\right) &= E\left[\left(\sum_{i=1}^{m}[C_i - E(C_i)]\right)^4\right] \\ &= \sum_{i=1}^{m} E\left([C_i - E(C_i)]^4\right) + \sum_{1\le i<j\le m} E\left([C_i - E(C_i)]^2\right) E\left([C_j - E(C_j)]^2\right) \end{aligned}$$

and

$$
\begin{aligned}
M_4 = E(C^4) &= E\left([C-E(C)]^4\right) + 4E(C^3)E(C) - 6E(C^2)E^2(C) + 4E^4(C) - E^4(C) \\
&= \sum_{i=1}^{m} E\left([C_i - E(C_i)]^4\right) + \sum_{1\le i<j\le m} E\left([C_i - E(C_i)]^2\right) E\left([C_j - E(C_j)]^2\right) \\
&\quad + 4M_3M_1 - 6M_2M_1^2 + 3M_1^4.
\end{aligned}
$$

Other generalizations of the algorithm include the analysis of sensitivities based on the derivatives of the total expected cost w.r.t. the model parameters. Since these are of minor importance in the current context, we will briefly present the methods in Section 4.4.

4.4 Numerical Examples

In this section, we first validate the the proposed semi-analytic algorithms by Monte Carlo and then apply it to analyze the impact of the model parameters on the total hedging cost. We will then examine the accuracy of approximating the higher moments and quantiles of the cost distribution by the independence assumption.

Table 4.4.1 compares the total expected cost estimated by the semi-analytic algorithm and Monte Carlo simulation, respectively. The results suggest that these two methods agree with each other reasonably well.

In Table 4.4.2, we compute the total expected cost for different combinations of model parameters. From this table, we can observe several interesting trends

1. The cost decreases as σ increases;

2. For small μ ($\mu = 0.05$), the cost decreases with maturity T. While for large μ ($\mu = 0.1, 0.15$), it increases with maturity T;

	Semi-Analytic	Monte Carlo: $\Delta t = 10^{-4}$	Monte Carlo: $\Delta t = 10^{-6}$
$T = 1$	0.1395	0.1491(0.0030)	0.1383(0.0029)
$T = 2$	0.1840	0.1794(0.0021)	0.1830(0.0026)
$T = 3$	0.1993	0.1937(0.0026)	0.1990(0.0023)
$T = 4$	0.2021	0.2060(0.0030)	0.2024(0.0021)
$T = 5$	0.1992	0.2012(0.0020)	0.1993(0.0019)

Table 4.4.1: Total expected cost. The common parameters are $S_0 = 50, K = 50, r = 0.02, \mu = 0.2, \sigma = 0.2, \alpha = 0.1, d = 0$. In the column "Monte Carlo: $\Delta t = 10^{-4}$", we generate 100000 paths with time step 10^{-4} in the Monte Carlo simulation and the number in the bracket is the standard error of the estimator; In the column "Monte Carlo: $\Delta t = 10^{-6}$", we generate 100000 paths with time step 10^{-6} in the Monte Carlo simulation and the number in the bracket is the standard error of the estimator. Note that the standard errors shown in the bracket is a measure of the statistical error, not of the discretization error. The discretization error can be seen from the changes in the estimates when we shorten the length of the time step in the path generation.

	$\sigma = 0.1$	$\sigma = 0.2$	$\sigma = 0.3$
$\mu = 0.05$	(0.0144,0.0134,0.0132)	(-0.0018,-0.0030,-0.0038)	(-0.0023,-0.0034,-0.0041)
$\mu = 0.1$	(0.1522,0.1952,0.2107)	(0.0202,0.0295,0.0355)	(0.0025,0.0036,0.0044)
$\mu = 0.15$	(0.3615,0.4132,0.4018)	(0.0697,0.0980,0.1125)	(0.0177,0.0253,0.0302)

Table 4.4.2: Total expected cost for different sets of parameters. The common parameters are $S_0 = 50, K = 50, r = 0.02, \alpha = 0.1, d = 0$. Each unit is a vector containing values for maturities $T = 1, T = 2, T = 3$, respectively.

3. The cost increases as μ increases.

To understand these trends, we begin with an analysis for the effect of the upward and downward movements of the sub-account on the cost. Suppose that we are now at a time point before maturity and have just rebalanced our hedging portfolio so that its value is equal to the option value. If the sub-account moves upward thereafter, both the option and the hedging portfolio will lose value. If the sub-account moves in the other direction, both gain in value. Whatever is the direction of the movement, the magnitudes of the changes in the value of the hedging portfolio are the same, since its position in the sub-account is fixed. However, due to the asymmetry of the option value, a downward movement in the sub-account will increase the option value more than that an upward movement can decrease. We will call the loss caused by an upward move of the sub-account the "up loss" and the "down loss" for the one caused by a downward move. From the above discussion, we know that the up loss should in general be smaller than the down loss.

With this intuition in mind, we can then explain the observed trends. As σ increases, the difference between the up an down losses is flattened because the asymmetry of the put option value is lessened. This means the up loss will increase while the down loss will decrease, and the down loss's decrease is slightly faster than the up loss's increase. If μ is small, the odds of the sub-account's moving up and down are roughly equal. So the overall effect of the flattening is the decrease in total loss. Moreover, a larger σ will dictate a higher hedging frequency. These two factors together explain trend 1.

The increase in T has the similar mitigation effect as the increase of σ on the asymmetry of option value. So when μ is small, we can see a steady decrease in cost caused by the increase of T in the first row of Table 4.4.2, which is captured by the first part of trend 2. However, when μ is large, the odds of the sub-account's moving up would be significantly higher than its moving down. Hence, although the magnitude of the up loss's increase is smaller than that of the down loss's decrease, the total loss would increase with T because the increased up loss gets a better chance to show up. These explain both trend 3 and the second part of trend 2. From another perspective, when the put is worthless, diligent hedging can be superfluous or even counterproductive.

So far we have analyzed the impact of μ, σ and T on the hedging cost, we now wish to investigate the effect of changes in S_0 and r. We will do this in a more straightforward way: computing the corresponding "Greeks" of the total expected cost. As we have mentioned in Section 2.3, Greeks are the derivatives of the options price (in the current context, the total expected cost) and reflect its sensitivity to certain underlying factors. For the derivative w.r.t. the initial sub-account value S_0, let DC_j^i be the i-th cost discounted to the j-th hitting time $\tau^{(j)}$ and let $h^S(S_0)$ be the derivative of the expectation of the first cost w.r.t. the sub-account value

$$h^S(S_0) = \frac{\partial}{\partial S_0} E\left\{DC_0^1\right\}.$$

Then, for the derivative of the second cost, we have

$$\begin{aligned}\frac{\partial}{\partial S_0} DC_0^2 &= I_{\epsilon_\lambda^{(1)} > \tau^{(1)}} e^{-r\tau^{(1)}} \frac{\partial}{\partial S_0} DC_1^2 \\ &= I_{\epsilon_\lambda^{(1)} > \tau^{(1)}} e^{-r\tau^{(1)}} \frac{\partial S_{\tau^{(1)}}}{\partial S_0} \frac{\partial}{\partial S_{\tau^{(1)}}} DC_1^2.\end{aligned}$$

Conditioning on $\mathcal{F}_{\tau^{(1)}}$ and $\epsilon_\lambda^{(1)} > \tau^{(1)}$, the inner expectation is

$$e^{-(r+\lambda)\tau^{(1)}} \frac{\partial S_{\tau^{(1)}}}{\partial S_0} h^S(S_{\tau^{(1)}}),$$

and the full expectation is

$$e^{-\alpha}h^S(Se^{-\alpha})L_{-\alpha} + e^{\alpha}h^S(Se^{\alpha})L_{\alpha},$$

where $L_{-\alpha}$ and L_{α} are defined as before.

In general, the expectation of the derivative the (n+1)-th cost w.r.t. the sub-account value is

$$\sum_{i=0}^{n} \binom{n}{i} e^{(2i-n)\alpha} h^S(S_0 e^{(2i-n)\alpha}) L_{\alpha}^{i} L_{-\alpha}^{n-i}.$$

For the derivative w.r.t. the risk free interest rate, let $h^r(S_0)$ be the derivative of the expectation of the first cost w.r.t. r:

$$h^S(S_0) = \frac{\partial}{\partial r} E\left\{DC_0^1\right\}.$$

Then the derivative of the second cost is

$$\begin{aligned}\frac{\partial}{\partial r}DC_0^1 &= \frac{\partial}{\partial r}(I_{\epsilon_\lambda^{(1)} > \tau^{(1)}} e^{-r\tau^{(1)}} DC_1^2) \\ &= I_{\epsilon_\lambda^{(1)} > \tau^{(1)}} e^{-r\tau^{(1)}} \frac{\partial}{\partial r} DC_1^2 + I_{\epsilon_\lambda^{(1)} > \tau^{(1)}} \tau^{(1)} e^{-r\tau^{(1)}} DC_1^2.\end{aligned}$$

Conditioning on $\mathcal{F}_{\tau^{(1)}}$ and $\epsilon_\lambda^{(1)} > \tau^{(1)}$, the inner expectation is

$$e^{-(r+\lambda)\tau^{(1)}} h^r(S_\tau^{(1)}) - \tau^{(1)} e^{-(r+\lambda)\tau^{(1)}} h(S_\tau^{(1)}).$$

This conditional expectation consists of two parts. The function $h^r(\cdot)$ and $h(\cdot)$ can be evaluated numerically. And using the indicator function, we can single them out from the full expectation. The rest are $E\left(e^{-(r+\lambda)\tau^{(1)}} I_{S_\tau^{(1)} = S_0 e^{k\alpha}}\right)$ and $E\left(\tau^{(1)} e^{-(r+\lambda)\tau^{(1)}} I_{S_\tau^{(1)} = S_0 e^{k\alpha}}\right)$. We have already obtained the analytical expression of the former. The latter can be

$S_0 = 30$	$S_0 = 40$	$S_0 = 50$	$S_0 = 60$	$S_0 = 70$
-0.0015	-0.0008	0.0038	0.0003	-0.0011

Table 4.4.3: Derivative of expected cost w.r.t. the sub-account. The common parameters are $K = 50, T = 1, r = 0.02, \mu = 0.05, \sigma = 0.2, \alpha = 0.1, d = 0$.

$T = 1$	$T = 2$	$T = 3$	$T = 4$	$T = 5$
-0.3164	-4842	-0.6002	-0.6879	-0.7569

Table 4.4.4: Derivative of expected cost w.r.t. interest rate. The common parameters are $S_0 = 50, K = 50, \mu = 0.05, \sigma = 0.2, \alpha = 0.1, d = 0$.

easily calculated by taking derivative of the former w.r.t. either r or λ.

In Table 4.4.3, we calculate the derivative of the total expected cost w.r.t. S_0. The derivatives are uniformly small, with slightly larger values at the strike. This indicates that the total hedging cost is not very sensitive to the initial sub-account value.

Table 4.4.4 contains the derivative of the total expected cost w.r.t. r. As expected, these derivatives are negative. Each re-balancing leads to either a loss or a profit and with r increased, the present values of both the loss and the profit are reduced. Since the total cost is the sum of all losses minus the sum of all profits, and over the course of hedging, loss are dominant in term of magnitude or frequency of occurrences (this is why we get many positive values for the total expected cost in Table 4.4.2), the total cost decreases. In addition, the impact of r increases with the maturity, as the discounting effect becomes more and more evident.

Now we come back to the question left in section 4.3: are the individual costs nearly independent? To find the answer, we used simulation to compare the true moments of the total cost with those obtained under the independence assumption. See Table 4.4.5 for the results. We comment that the second moment are very close to each other, while the differences in 3rd and 4th moments are slightly larger but still negligible. These

	$\mu = 0.05$ $\sigma = 0.2$	$\mu = 0.1$ $\sigma = 0.2$	$\mu = 0.15$ $\sigma = 0.2$	$\mu = 0.05$ $\sigma = 0.3$	$\mu = 0.1$ $\sigma = 0.3$	$\mu = 0.15$ $\sigma = 0.3$
moments-true:						
1st	-0.0050	0.0333	0.1119	-0.0021	0.0049	0.0278
2nd	0.9895	0.8912	0.7420	1.0012	0.9855	0.9326
3rd	-0.6886	-0.5095	-0.2514	-0.5692	-0.5369	-0.4827
4th	4.5301	3.7438	2.7696	4.7584	4.5705	4.2865
moments-independence:						
1st	-0.0050	0.0333	0.1119	-0.0021	0.0049	0.0278
2nd	0.9953	0.8936	0.7247	1.0100	0.9877	0.9325
3rd	-0.8625	-0.6557	-0.3079	-0.6723	-0.6254	-0.5124
4th	4.5217	3.6388	2.2792	4.1671	3.9535	3.4852

Table 4.4.5: Moments comparison. The common parameters are: $T = 3, S_0 = K = 50, r = 0.02, \alpha = 0.1, d = 0$.

suggests a very weak linear dependence between individual costs, which can therefore be sacrificed for computational efficiency. In Table 4.4.6-4.4.11, we approximate, with various distributions (single normal, Gumbel, mixture of two normals and Edgeworth expansion), the quantile of the the total cost. Because of the availability of its first 4 moments, we adopt the method of moments to determine the parameters in each of the candidate distributions and use the fitted distributions to approximate the true quantile (See Section 4.5 for the details).

We set the following criteria for the selection of the appropriate candidate distribution

1. Since the cost can be either positive or negative, the desirable distribution should be two-tailed;
2. The distribution should be flexible enough to capture the asymmetry and the heavy tail feature of the cost distribution;
3. Since we use the method of moments for fitting, the moments of the distribution should exist and be easy to calculate. Moreover, the expression of the moments should not be too exotic to be processed by a regular optimization routine. Finally,

	True	Normal	Gumbel	Mixture	Edgeworth
90%	1.0844	1.2698	1.0896	1.0734	1.1441
95%	1.4244	1.6362	1.2975	1.3687	1.3692
97.5%	1.7469	1.9447	1.4551	1.6571	1.5360
99%	2.1540	2.3091	1.6272	2.0721	3.1638

Table 4.4.6: Quantile comparison: $\mu = 0.05, \sigma = 0.2$. The common parameters are: $T = 3, S_0 = K = 50, r = 0.02, \alpha = 0.1, d = 0$.

	True	Normal	Gumbel	Mixture	Edgeworth
90%	1.0594	1.2423	1.0714	1.0407	1.1367
95%	1.3803	1.5874	1.2667	1.3180	1.3576
97.5%	1.6962	1.8824	1.4181	1.5945	1.5202
99%	2.1223	2.2280	1.5813	2.0721	2.0036

Table 4.4.7: Quantile comparison: $\mu = 0.1, \sigma = 0.2$. The common parameters are: $T = 3, S_0 = K = 50, r = 0.02, \alpha = 0.1, d = 0$.

the information of this distribution should concentrate in its first few moments (for example, the first two moments of the normal distribution contain all its information).

4. Among all the distributions satisfying the above conditions, we prefer choosing the simplest, in order to be as objective as possible.

The results suggest that the right tail of the true distribution is heavier than Gumbel and lighter than normal. In terms of one-sided quantile estimation, the normal mixture dominates all the others. In Figure 4.4.1, we also plot the real histogram against the densities of the fitted distribution. the normal mixture model provides the best overall fit to the cost distribution (though a slight deviation near the mode). In Figure 4.4.2, we visualize the tail behavior of these distributions using the Q-Q plot. From this plot, it is clearly seen that that the normal mixture model gives the best fit to the right tail of the cost distribution.

	True	Normal	Gumbel	Mixture	Edgeworth
90%	1.0228	1.2065	1.0517	1.0443	1.1263
95%	1.3232	1.5001	1.2136	1.2940	1.3436
97.5%	1.6169	1.7859	1.3656	1.5243	1.5027
99%	2.0258	2.0988	1.5133	1.8272	1.6456

Table 4.4.8: Quantile comparison: $\mu = 0.15, \sigma = 0.2$. The common parameters are: $T = 3, S_0 = K = 50, r = 0.02, \alpha = 0.1, d = 0$.

	True	Normal	Gumbel	Mixture	Edgeworth
90%	1.1182	1.2803	1.0989	1.1519	1.1704
95%	1.4803	1.6512	1.3100	1.4655	1.4273
97.5%	1.8360	1.9591	1.4666	1.7523	1.6310
99%	2.2933	2.3257	1.6397	2.1202	1.8422

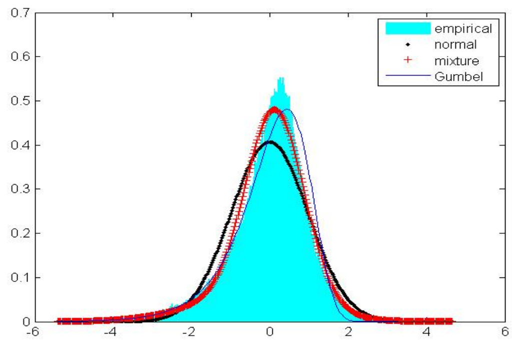
Table 4.4.9: Quantile comparison: $\mu = 0.05, \sigma = 0.3$. The common parameters are: $T = 3, S_0 = K = 50, r = 0.02, \alpha = 0.1, d = 0$.

	True	Normal	Gumbel	Mixture	Edgeworth
90%	1.1132	1.2771	1.0973	1.1218	1.1713
95%	1.4721	1.6397	1.3023	1.4302	1.4289
97.5%	1.8235	1.9506	1.4620	1.7250	1.6333
99%	2.2759	2.3143	1.6338	2.1261	1.8452

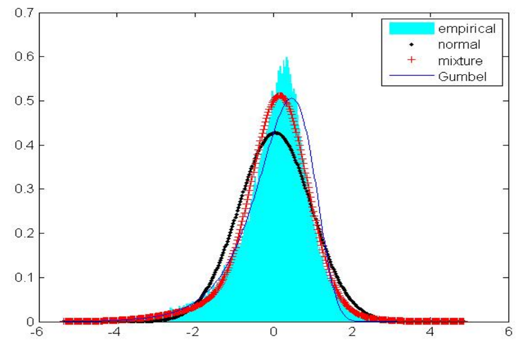
Table 4.4.10: Quantile comparison: $\mu = 0.1, \sigma = 0.3$. The common parameters are: $T = 3, S_0 = K = 50, r = 0.02, \alpha = 0.1, d = 0$.

	True	Normal	Gumbel	Mixture	Edgeworth
90%	1.1005	1.2649	1.0900	1.1268	1.1700
95%	1.4788	1.6156	1.2880	1.4264	1.4270
97.5%	1.7859	1.9198	1.4447	1.7044	1.6309
99%	2.2224	2.2735	1.6117	2.0678	1.8423

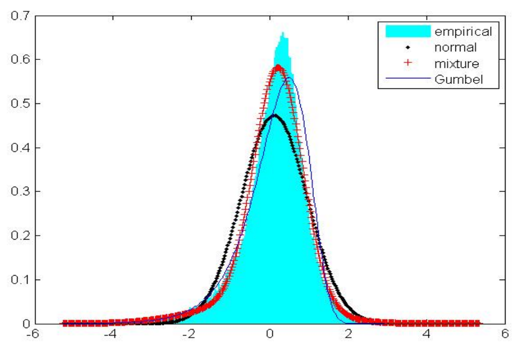
Table 4.4.11: Quantile comparison: $\mu = 0.15, \sigma = 0.3$. The common parameters are: $T = 3, S_0 = K = 50, r = 0.02, \alpha = 0.1, d = 0$.



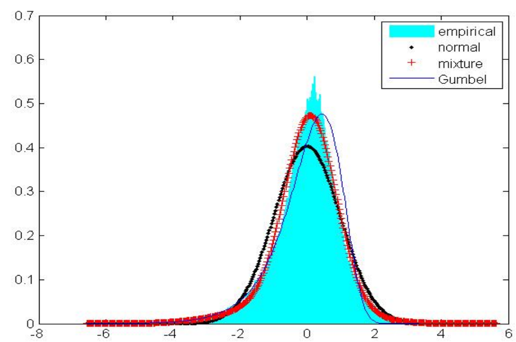
$\mu = 0.05, \sigma = 0.2$



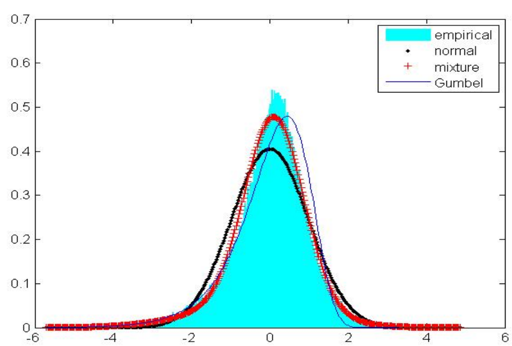
$\mu = 0.1, \sigma = 0.2$



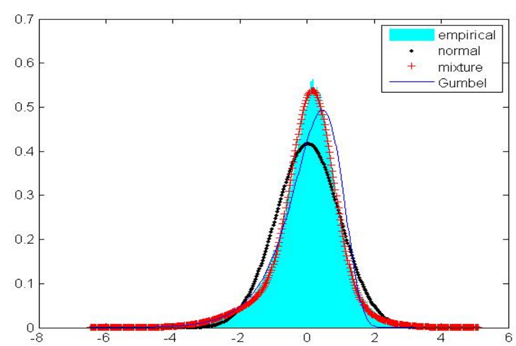
$\mu = 0.15, \sigma = 0.2$



$\mu = 0.05, \sigma = 0.3$



$\mu = 0.1, \sigma = 0.3$

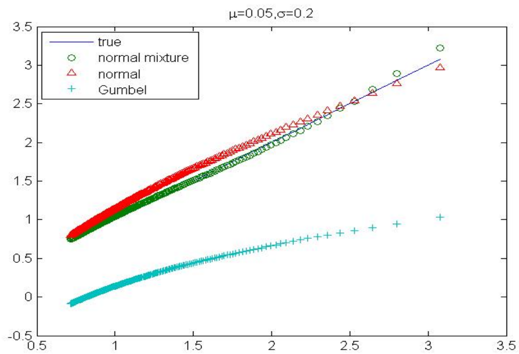


$\mu = 0.15, \sigma = 0.3$

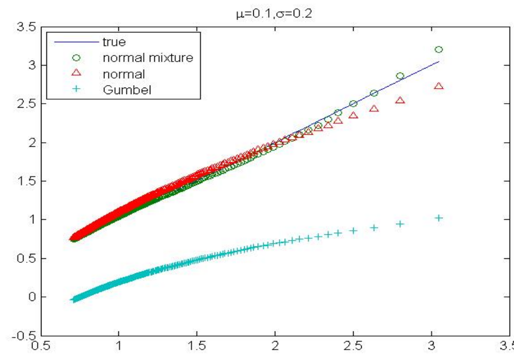
Figure 4.4.1: Density comparison. The common parameters are: $T = 3, S_0 = K = 50, r = 0.02, \alpha = 0.1, d = 0$.

4.5 Quantile Calculation

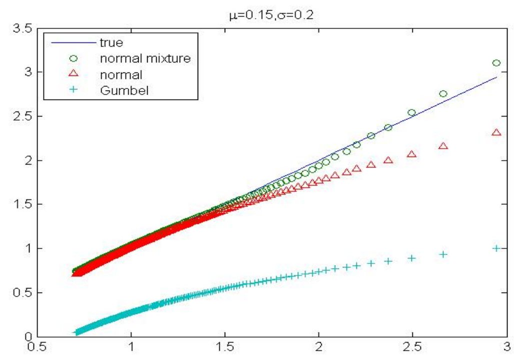
In this section, we review some statistical properties of the GEV distribution and the Edgeworth expansion that we use in Section 4.4 for the estimation of the quantiles of the hedging cost distribution.



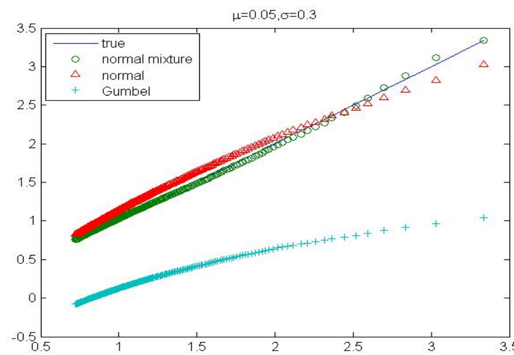
$\mu = 0.05, \sigma = 0.2$



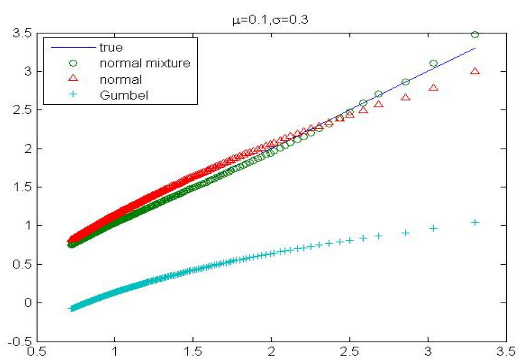
$\mu = 0.1, \sigma = 0.2$



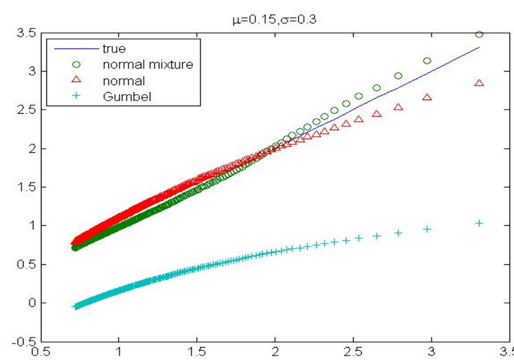
$\mu = 0.15, \sigma = 0.2$



$\mu = 0.05, \sigma = 0.3$



$\mu = 0.1, \sigma = 0.3$



$\mu = 0.15, \sigma = 0.3$

Figure 4.4.2: Quantile-Quantile plot. We plot the 80%-99.9% quantiles of the distributions. The X-axis is the true/empirical quantile of the cost distribution. The Y-axis is the quantile of the approximating distributions. The solid line is the true/empirical quantiles of the cost distribution estimated by Monte Carlo; the circle line is the quantiles of the mixture distribution of two normals; the triangle line is the quantils of the normal distribution and the plus sign line is the quantiles of the Gumbel distribution. The common parameters are: $T = 3, S_0 = K = 50, r = 0.02, \alpha = 0.1, d = 0$.

4.5.1 GEV Distribution

We consider this distribution class for its flexible tail behavior and compact parametrization. According to McNeil et al. (2005), the GEV distribution class has cdf $H_{\xi,\mu,\sigma}(x) := H_\xi((x-\mu)/\sigma)$, where

$$H_\xi(x) = \begin{cases} \exp(-(1+\xi x)^{-1/\xi}) & , \quad \xi \neq 0 \\ \exp(-e^{-x}) & , \quad \xi = 0 \end{cases} .$$

Here, ξ, μ and σ are the shape, location and scale parameter, respectively.

The distributions associated with $\xi > 0$ are called Frechet and these include well known fat tailed distributions such as the Pareto, Cauchy, Student-t and mixture distributions. If $\xi = 0$, the GEV distribution is the Gumbel class and includes the normal, exponential, gamma and lognormal distributions but only the lognormal distribution has a moderately heavy tail.
Finally, in the case where $\xi < 0$, the distribution class is Weibull. These are short tailed distributions with finite lower bounds and include distributions such as uniform and beta distributions.

From the histogram of the true cost distribution(Figure 4.4.1), we conclude that a short-tailed distribution($\xi < 0$) is obviously not suitable. Moreover, the third column in Table 4.4.6-4.4.11 suggest that the right tail of the true distribution is thinner than that of Normal, thus a heavy-tailed distribution($\xi > 0$) is also improper. So finally we are left with the Gumbel class($\xi = 0$).

The Gumbel distribution has two parameters: The location parameter μ and the scale parameter β(β is negative, the re-parametrization is made for the sake of consistency with the Matlab built-in functions for extreme value distribution).

Its cdf is $\exp(-e^{(x-\mu)/\beta})$, mean is $\mu+\gamma\beta$ and variance is $(\pi\beta)^2/6$, where $\gamma \approx 0.577215665$ is the Euler constant.

Now suppose we have calculated the first two moment of the cost distribution, denoted by m_1 and m_2, then using the method of moments, the estimate for μ and β can be obtained by solving

$$\begin{cases} \mu+\gamma\beta &= m_1 \\ (\pi\beta)^2/6 &= m_2 - m_1^2. \end{cases}$$

Once we get $\hat{\mu}$ and $\hat{\sigma}$, the α quantile of the fitted distribution is the root of

$$\exp(-e^{-(x-\hat{\mu})/\hat{\beta}}) = \alpha.$$

The results form the column "Gumbel" of Table 4.4.6-4.4.11.

4.5.2 Edgeworth Expansion

The Edgeworth Expansion provides a moment approximation to the CDF of a distribution. It gives an accuracy of $O(n^{-\frac{3}{2}})$, but can sometimes generate values that exceed the theoretical $[0,1]$ range.

In light of Cheah et al. (1993), if we denote by $\mu, \sigma^2, \mu_3, \mu_4$ the mean, variance, third and fourth central moment of the true distribution, then the corresponding cumulants are $\kappa_1 = \mu, \kappa_2 = \sigma^2, \kappa_3 = \mu_3, \kappa_4 = \mu_4 - 3\sigma^2$ and the cumulants of the normalized distribution are $\gamma_1 = 0, \gamma_2 = 1, \gamma_3 = \kappa_3/\sigma^3, \gamma_4 = \kappa_4/\sigma^4$.

The approximating cdf, up to the fourth cumulant, is

$$F_{\mathrm{E}}(z) = \Phi(z) + \phi(z)(-\frac{\gamma_3}{6}h_2(z) - \frac{3\gamma_4 h_3(z) + \gamma_3^2 h_5(z)}{72}),$$

where $\Phi(z)$ and $\phi(z)$ are the cdf and pdf of the standard normal distribution and $h_2(z) = z^2 - 1, h_3(z) = z^3 - 3z, h_5(z) = z^5 - 10z^3 + 15z$ are the Hermite polynomials.

Chapter 5

The Percentile Principle Premium for Variable Annuities

The semi-analytic algorithm developed in Chapter 4 allows us to analytically quantify the re-balancing cost of the move-based hedging. In this chapter, we introduce a modified "Percentile Premium Principle" for variable annuities, which is built upon this insightful quantification, to incorporate the significant discrete hedging cost.

The "Percentile Premium Principle", as an alternative to the "Expected Premium Principle", derives its utility from the information on the extreme losses to ensure that the probability of a loss on a contract will not exceed a risk threshold. In finance, the same consideration leads to the VaR (VaR is percentile in essence) based techniques for economical capital requirement, which delimits the amount of risk capital, assessed on a realistic basis, that a firm should possess to cover the risks that it is running or collecting as a going concern, such as market risk, credit risk, and operational risk.

Inspired by the spirit of the "Percentile Premium Principle", we modify it for the move-based hedging of variable annuities. In particular, we introduce a loading, in addition to the regular charge, to provide the insurer enough fund with high realistic confidence for the operation of discrete re-balancing. To see the modified "Percentile Premium Principle" in action, we start with the pricing of the Guaranteed Minimum Maturity Benefit to present the detailed procedures of its implementation. Then we apply it to more complex, path-dependent VA products.

5.1 Guaranteed Minimum Maturity Benefit

Consider a VA with guaranteed minimum maturity benefit (GMMB) whose payoff is $e^{-rT}(G - S_T)_+$.

Suppose at time 0, an annuitant starts with one unit of the VA sub-account, which is worth X_0. Let the maturity of this VA contract be T and over the time period $[0, T]$, the annuitant is guaranteed a minimum rate of return g. From the insurer's perspective, this guarantee lead to a time-T loss of

$$(X_0(1+g)^T - X_T)_+.$$

To compensate for the loss, the insurer charges a fee δ, proportional to the level of sub-account value over the life of the VA. We assume that the sub-account mimics a certain market index S_t that initially follows the same GBM

$$\begin{aligned} X_t &= X_0 e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma W_t}, \\ S_t &= S_0 e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma W_t}, \\ X_0 &= S_0. \end{aligned}$$

After the insurer's deducting the fee, the sub-account value falls to

$$X_t = X_0 e^{(\mu - \delta - \frac{1}{2}\sigma^2)t + \sigma W_t},$$

while the market index S_t remains unchanged.

The insurer's problem is to hedge a put option (option A) written on an untradable asset X_t, with payoff $(X_0(1+g)^T - X_T)_+$, using a tradable asset S_t. To tackle this problem, consider another put option (option B) with payoff $(X_0(1+g)^T - \bar{X}_T)_+$, where $\bar{X}_t$ is a tradable, dividend paying asset with dividend yield d and the following dynamic

$$\bar{X}_t = \bar{X}_0 e^{(\mu - d - \frac{1}{2}\sigma^2)t + \sigma W_t},$$
$$\bar{X}_0 = X_0.$$

Then according to Black-Scholes formula, the price of option B (or the cost of continuously hedging option B with $\bar{X}_t$) is

$$Ke^{-rT}N(-d_2) - S_0 e^{-\delta T}N(-d_1), \tag{5.1.1}$$

where

$$d_1 = \frac{log(\frac{S_0}{K}) + (r - \delta + \frac{1}{2}\sigma^2)}{\sigma\sqrt{T}},$$
$$d_2 = d_1 - \sigma\sqrt{T}.$$

Note that

(i) Option A and option B have the same payoff,

and

(ii) Using $\bar{X}_t$ to hedge is equivalent to using S_t if we reinvest all the dividends (indeed, if S_t pays dividends at rate g, then it becomes $\bar{X}_t$),

We conclude that the cost of continuously hedging option A with S_t is given by (5.1.1).

On the revenue side, the total amount of money the insurer accumulates over $[0, T]$

is

$$\int_0^T e^{-rt}\delta S_t dt.$$

Its present value is obtained by taking expectation under the risk neutral measure

$$E^Q(\int_0^T e^{-rt}\delta S_t dt) = \delta S_0 T. \tag{5.1.2}$$

Equating (5.1.1) and (5.1.2) gives us the fair value of δ for continuous hedging. The results are summarized in the lines "Regular Fee" of Table 5.1.1-5.1.5.

For discrete hedging, however, a surcharge (the loading) needs to be imposed to cover the cost arising from the non-self-financing re-balancing strategy. In the actuarial practice, the fair value of the loading is based on the "Expected Premium Principle"

$$\text{expected total re-balancing cost=expected revenue.} \tag{5.1.3}$$

However, due to the observed heavy tail of the total re-balancing cost, we would rather replace its expectation by its 95% quantile (of the real probability distribution) in (5.1.3), which leads to the following "Percentile Premium Principle"

$$\text{95\% quantile of the total re-balancing cost=expected revenue.} \tag{5.1.4}$$

As opposed to the traditional "Expected Premium Principle", the new valuation scheme (5.1.4) is more prudent in that it offers enough funds for the insurer to cover the discrete hedging cost not just on average, but for 95% of the time. The spirit of the "Percentile Premium Principle" also dictates the choice of the physical probability measure for its implementation, because we are now concerned with the extreme losses in the real world scenario. Consequently, we set the growth rate of the sub-account at $\mu - \delta$ in the algorithms proposed in Chapter 4 to compute the 95% quantile of the distribution of the

	$\mu = 0.05$ $\sigma = 0.2$	$\mu = 0.1$ $\sigma = 0.2$	$\mu = 0.15$ $\sigma = 0.2$	$\mu = 0.05$ $\sigma = 0.25$	$\mu = 0.1$ $\sigma = 0.25$	$\mu = 0.15$ $\sigma = 0.25$	$\mu = 0.05$ $\sigma = 0.3$	$\mu = 0.1$ $\sigma = 0.3$	$\mu = 0.15$ $\sigma = 0.3$
Regular Fee	0.0855	0.0855	0.0855	0.1083	0.1083	0.1083	0.1291	0.1291	0.1291
Loading	0.0145	0.0128	0.0120	0.0142	0.0139	0.0128	0.0148	0.0138	0.0135

Table 5.1.1: Management fee for discrete hedging. The common parameters are: $T = 2, S_0 = K = 50, r = 0.02, \alpha = 0.1, g = 0$.

	$\mu = 0.05$ $\sigma = 0.2$	$\mu = 0.1$ $\sigma = 0.2$	$\mu = 0.15$ $\sigma = 0.2$	$\mu = 0.05$ $\sigma = 0.25$	$\mu = 0.1$ $\sigma = 0.25$	$\mu = 0.15$ $\sigma = 0.25$	$\mu = 0.05$ $\sigma = 0.3$	$\mu = 0.1$ $\sigma = 0.3$	$\mu = 0.15$ $\sigma = 0.3$
Regular Fee	0.0617	0.0617	0.0617	0.0788	0.0788	0.0788	0.0945	0.0945	0.0945
Loading	0.0090	0.0083	0.0074	0.0092	0.0090	0.0080	0.0094	0.0089	0.0079

Table 5.1.2: Management fee for discrete hedging. The common parameters are: $T = 3, S_0 = K = 50, r = 0.02, \alpha = 0.1, g = 0$.

re-balancing cost and denote the value by $Q_{0.95}(\delta; \mu, \sigma, \alpha, r, S_0, K, T)$. On the revenue side, the present value of the total amount of money we accumulate over $[0, T]$ is

$$\int_0^T e^{-rt}\delta S_t dt.$$

Its expectation, under the physical measure, is

$$S_0\delta\frac{e^{(\mu-r-\delta)T} - 1}{\mu - r - \delta}.$$

The fair value of the loading δ can be obtained by solving

$$Q_{0.95}(\delta; \mu, \sigma, \alpha, r, S_0, K, T) = S_0\delta\frac{e^{(\mu-r-\delta)T} - 1}{\mu - r - \delta}. \tag{5.1.5}$$

Table 5.1.1-5.1.5 calculate the fair value of the loading in line with the "Percentile Premium Principle" (5.1.5). As expected, the regular fee-the cost of continuous hedging-is

	$\mu = 0.05$ $\sigma = 0.2$	$\mu = 0.1$ $\sigma = 0.2$	$\mu = 0.15$ $\sigma = 0.2$	$\mu = 0.05$ $\sigma = 0.25$	$\mu = 0.1$ $\sigma = 0.25$	$\mu = 0.15$ $\sigma = 0.25$	$\mu = 0.05$ $\sigma = 0.3$	$\mu = 0.1$ $\sigma = 0.3$	$\mu = 0.15$ $\sigma = 0.3$
Regular Fee	0.0484	0.0484	0.0484	0.0623	0.0623	0.0623	0.0750	0.0750	0.0750
Loading	0.0065	0.0060	0.0047	0.0068	0.0063	0.0053	0.0071	0.0062	0.0055

Table 5.1.3: Management fee for discrete hedging. The common parameters are: $T = 4, S_0 = K = 50, r = 0.02, \alpha = 0.1, g = 0$.

	$r = 0.01$	$r = 0.02$	$r = 0.03$	$r = 0.04$	$r = 0.05$
Regular Fee	0.1076	0.0945	0.0828	0.0725	0.0635
Loading	0.0091	0.0089	0.0087	0.0083	0.0080

Table 5.1.4: Management fee for discrete hedging. The common parameters are: $T = 3, S_0 = K = 50, \mu = 0.1, \sigma = 0.3, \alpha = 0.1, g = 0$.

	$S_0 = 40$	$S_0 = 50$	$S_0 = 60$	$S_0 = 70$
Regular Fee	0.2262	0.0945	0.0423	0.0213
Loading	0.0100	0.0089	0.0067	0.0053

Table 5.1.5: Management fee for discrete hedging. The common parameters are: $T = 3, K = 50, \mu = 0.1, \sigma = 0.3, r = 0.02, \alpha = 0.1, g = 0$.

independent of the real rate of return of the sub-account. The loading is the additional fee we need to charge when the re-balancing cost of discrete hedging is taken into account. The results suggest several interesting correlations between the cost and the model parameters.

1. The loading decreases with μ. Indeed, from Table 4.4.6-4.4.11, we observe that the 95% quantile of the discrete hedging cost distribution decreases with μ, so does the left hand side of (5.1.5). Because a higher value of μ dictates a higher level of the sub-account, to which the proportional management fee is linked, the value of the fee needs to decrease for the right hand side of (5.1.5) to match the left hand side. From another perspective, as μ decreases, the downside risk of the put option becomes more and more pronounced, the insurer thus needs to charge more for its risk exposure.

2. The loading increases with σ. The reason lies in two facts: (i) the 95% quantile of the discrete hedging cost distribution increase with σ, so does the left hand side of (5.1.5); (ii) The right hand side is independent of σ and is an (piecewise) increasing function of the loading δ.

3. The loading increases with r while decreases with T and S_0. To give an explanation, the following observations may be helpful: (i) The right hand side of (5.1.5) is a decreasing(increasing) function of r (T and S_0) for small $(\mu - r - \delta)T$(for $\delta < \mu - r$). (ii) The increase in r (T and S_0) has a more significant impact on the left (right) hand side of (5.1.5).

4. There is a positive correlation between the regular fee and the loading.

5.2 Annual Ratchet VA

In this section, we consider the pricing and hedging of a n-year Annual Ratchet VA. Under the ratchet contract design, the participation in the equity index is evaluated year by year. Each year the guaranteed payoff is stepped up by the greater of the floor rate g and the return of the sub-account.

Suppose the sub-account starts at S_0, then at time 1, the wealth of the policyholder is guaranteed to be the maximum of $S_0(1+g)$ and S_1, denote by S_1^*. The policyholder reinvests any extra amount to the sub-account so that the number of shares he holds for the next period is $\frac{S_1^*}{S_1}$. At time 2, the account value is the maximum of $S_1^*(1+g)$ and $\frac{S_1^*}{S_1}S_2$, denoted by $S_2^* \cdots$. This process of annual ratcheting continues to the end of year n, resulting in a total discounted cost for the insurer

$$\varphi_{ratchet,n} = e^{-r}[S_0(1+g)-S_1]_+ + e^{-2r}[S_1^*(1+g)-\frac{S_1^*}{S_1}S_2]_+ + \cdots + e^{-nr}[S_{n-1}^*(1+g)-\frac{S_{n-1}^*}{S_{n-1}}S_n]_+, \tag{5.2.6}$$

where

$$S_i^* = \max\{S_{i-1}^*(1+g), \frac{S_{i-1}^*}{S_{i-1}}S_i\} = S_{i-1}^* \max\{(1+g), \frac{S_i}{S_{i-1}}\} = S_{i-1}^* G_i, S_0^* = S_0. \tag{5.2.7}$$

Note that our formulation of the payoff for annual ratchet VA is somewhat uncommon (For common ones, see for example, Hardy (2004)). This is because we look at the cost to the insurer, instead of the payoff to the policyholder. Indeed, the annual ratcheting at the end of each year brings an instant cost to the insurer, as it has to boost the sub-account with money from its own pocket. However, the increase in the sub-account is not cashed out by the policyholder until the maturity of the contract. So from the policyholder's perspective, the annual increase is accumulated to the maturity, resulting in a payoff of a form similar to that in Hardy (2004). But for the insurer, the cost is the sum of each annual boosting, the present value of which is given by (5.2.6).

To price, we rewrite $\varphi_{ratchet,n}$ as

$$\varphi_{ratchet,n} = e^{-r} S_0[(1+g) - \frac{S_1}{S_0}]_+ + e^{-2r} S_1^*[(1+g) - \frac{S_2}{S_1}]_+ + \cdots + e^{-nr} S_{n-1}^*[(1+g) - \frac{S_n}{S_{n-1}}]_+.$$

For the i-th term on the RHS, we have

$$E^Q \left\{ e^{-ir} S_{i-1}^*[(1+g) - \frac{S_i}{S_{i-1}}]_+ \right\} = e^{-ir} E^Q(S_{i-1}^*) E^Q \left[(1+g) - \frac{S_i}{S_{i-1}} \right]_+ ,$$

where we use the fact that S_{i-1}^* is independent of $\frac{S_i}{S_{i-1}}$.

According to the Black-Scholes formula for a put option with strike $1+g$ and $S_0 = 0$,

$$E^Q \left\{ \left[(1+g) - \frac{S_i}{S_{i-1}} \right]_+ \right\} = (1+g)N(-d_2) - e^{r-d} N(-d_1),$$

where

$$d_1 = \frac{\log \frac{1}{1+g} + (r - d + \frac{1}{2}\sigma^2)}{\sigma},$$

$$d_2 = d_1 - \sigma.$$

By definition of S_i^*, we have

$$E^Q(S_i^*) = E^Q(S_{i-1}^*)E^Q[\max(1+g, \frac{S_i}{S_{i-1}})] = E^Q(S_{i-1}^*)E^Q\left\{\left[\frac{S_i}{S_{i-1}} - (1+g)\right]_+ + (1+g)\right\}.$$

And again, by virtue of the Black-Scholes formula, we have

$$A = E^Q\left\{\left[\frac{S_i}{S_{i-1}} - (1+g)\right]_+ + (1+g)\right\} = e^{r-d}N(d_1) - (1+g)N(d_2) + (1+g),$$

where d_1 and d_2 are defined as before.

Thus,

$$E^Q(S_i^*) = S_0A^i, \quad i \in \{1, 2, \cdots, n\}.$$

In summary, the price of the n-year annual ratchet VA, or the cost of continuously hedging it, is

$$P_{ratchet,n} = S_0[(1+g)N(-d_2) - e^{r-d}N(-d_1)]\sum_{i=1}^{n} e^{-ir}A^{i-1}. \tag{5.2.8}$$

The hedging strategy for the annual ratchet VA can be described as follow:

Suppose we sell a n-year annual ratchet VA at time 0 with payoff structure given by $\varphi_{ratchet,n}$. In period (0,1], we hedge $S_0^* = S_0$ units of a put option with payoff $[(1+g)-\frac{S_1}{S_0}]_+$ at time 1; In period (1,2], we hedge S_1^*(note that, this strategy is feasible since S_1^* is known at time 1) units of a put option with payoff $[(1+g) - \frac{S_2}{S_1}]_+$ at time 2;... In general, in period $(i-1, i]$ $(i \in \{1, 2, \cdots, n\})$, we hedge S_{i-1}^* units of a put option with payoff $[(1+g) - \frac{S_i}{S_{i-1}}]_+$ at time i.

It is worthwhile to point out that in the strategy we just described, the option we hedge in period $(i-1, i]$ does not depend on i, only the units of such an option differ from period to period. In fact, the option we hedge for each period is simply a vanilla put option with strike $K = 1+g$, time to maturity $T = 1$ and initial sub-account value $S_0 = 1$. We can use the semi-analytic algorithm to compute the moments of the total discounted (at time $i-1$) costs for discretely hedging such an option over period $(i-1, i]$. In particular, let C_{i-1} be the total discounted hedging cost for $(i-1, i]$ and M_k be C_{i-1}'s k-th moment. For a n-year annual ratchet VA, the total cost of discrete hedging is

$$TC = S_0 C_0 + e^{-r} S_1^* C_1 + \cdots + e^{-(n-1)r} S_{n-1}^* C_{n-1}. \tag{5.2.9}$$

Suppose we are interested in the k-th moment of the total cost (5.2.9), then by virtue of the multinomial theorem,

$$\begin{aligned}
&E(TC^k) \\
&= E\left[\sum_{m_1+m_2+\cdots+m_n=k} \frac{k!}{m_1!m_2!\cdots m_n!}(S_0^{m_1}C_0^{m_1})(e^{-m_2 r}S_1^{*m_2}C_1^{m_2})\cdots(e^{-(n-1)m_n r}S_{n-1}^{*m_n}C_{n-1}^{m_n})\right] \\
&= \sum_{m_1+m_2+\cdots+m_n=k} \frac{k!}{m_1!m_2!\cdots m_n!} e^{-(m_2+2m_3+\cdots+(n-1)m_n)r} E\left[(C_0^{m_1}C_1^{m_2}\cdots C_{n-1}^{m_n})(S_0^{*m_1}S_1^{*m_2}\cdots S_{n-1}^{*m_n})\right] \\
&= \sum_{m_1+m_2+\cdots+m_n=k} \frac{k!}{m_1!m_2!\cdots m_n!} e^{-(m_2+2m_3+\cdots+(n-1)m_n)r} E\left[(C_0^{m_1}C_1^{m_2}\cdots C_{n-1}^{m_n})\right] E\left[(S_0^{*m_1}S_1^{*m_2}\cdots S_{n-1}^{*m_n})\right].
\end{aligned} \tag{5.2.10}$$

Exploiting the independence of $C_0, C_1, \cdots, C_n$, we have

$$E\left[C_0^{m_1}C_1^{m_2}\cdots C_{n-1}^{m_n}\right] = M_{m_1}M_{m_2}\cdots M_{m_n}.$$

And by iterative conditioning, we have

$$E\left[(S_0^{*m_1}S_1^{*m_2}\cdots S_{n-1}^{*m_n})\right] = S_0^{m_1+m_2+\cdots+m_n} J(m_2+m_3+\cdots+m_n)J(m_3+\cdots+m_n)\cdots J(m_n),$$

where

$$\begin{aligned}
J(k) &= E(G_i^k) \\
&= (1+g)^k N(\frac{\log(1+g)-(\mu-d-\frac{1}{2}\sigma^2)}{\sigma}) \\
&\quad + e^{k(\mu-d-\frac{1}{2}\sigma^2)+\frac{1}{2}k^2\sigma^2}[1-N(\frac{\log(1+g)-(\mu-d-\frac{1}{2}\sigma^2)}{\sigma}-k\sigma)].
\end{aligned}$$

Now we turn to the calculation of the fair management fee δ for the annual ratchet VA (δ is constant over the term of the annual ratchet VA and does not vary from year to year). In light of the "Percentile Premium Principle", the fair management fee includes two parts, the regular fee and the loading. The first is to finance the cost of continuous hedging while the second covers the cost arising from discrete re-balance. For continuous hedging, the cost is given by $P_{ratchet,n}$ in (5.2.8), and the revenue collected by the insurer during the life of this contract is

$$\int_0^1 e^{-rt}\delta S_t dt + \int_1^2 e^{-rt}\delta \frac{S_1^*}{S_1} S_t dt + \cdots + \int_{n-1}^n e^{-rt}\delta \frac{S_{n-1}^*}{S_{n-1}} S_t dt.$$

Taking expectation under the risk neutral measure Q, we get the present value of the revenue

$$R_{cont} = \delta S_0 \sum_{i=1}^{n-1} e^{-ir} A^i. \tag{5.2.11}$$

The regular fee is obtained by equating $P_{ratchet,n}$ with R_{cont}. Interestingly, by comparing (5.2.8) with (5.2.11), we see that the regular fee does not depend on n, the number of ratcheting of the VA.

According to the "Percentile Premium Principle", the loading is the solution to

95% quantile of the total re-balancing cost=expected revenue under the physical measure.

The quantile can be approximated by our algorithm and the expected revenue is

$$\begin{aligned}
E^P \quad & [\int_0^1 e^{-rt}\delta S_t dt + \int_1^2 e^{-rt}\delta \frac{S_1^*}{S_1} S_t dt + \cdots + \int_{n-1}^n e^{-rt}\delta \frac{S_{n-1}^*}{S_{n-1}} S_t dt] \\
= \quad & \delta[\int_0^1 e^{-rt} S_0 e^{(\mu-d)t} dt + \int_1^2 e^{-rt} E^P(S_1^*) e^{(\mu-d)(t-1)} dt \qquad (5.2.12) \\
& + \cdots + \int_{n-1}^n e^{-rt} E^P(S_{n-1}^*) e^{(\mu-d)(t-n+1)} dt] \\
= \quad & \delta S_0 \frac{e^{\mu-d-r}-1}{\mu-d-r} \sum_{i=0}^{n-1} J^i(1) e^{-ir}. \qquad (5.2.13)
\end{aligned}$$

Tables 5.2.6 to 5.2.9 list the results generated by (5.2.10) and compare them with those computed by Monte Carlo simulation. We see that the 1st, 2nd and 4th moments are close to each other while the 3rd moments diverge a bit far. Figure 5.2.1 plots: 1) The histograms of the discrete hedging cost distribution generated by Monte Carlo simulation (with 10^6 iterations); 2) The density function of a mixture distribution of two normals, fitted to the first four moments of the cost distribution, which are estimated using simulation; 3) The density function of a mixture distribution of two normals, fitted to the first four moments of the cost distribution, which are estimated using the semi-analytic algorithm. Table 5.2.10 compares the quantiles of the true distribution and those given by the approximating distributions. Despite the relatively large difference in the estimation of the third moments, the two fitted densities are almost indistinguishable and therefore provide equally good estimations for the quantiles.

In Tables 5.2.11-5.2.14, we compute the regular fee and the loading for annual ratchet VAs with various model/contract specifications. Comparing the trend of these numbers with those for GMMB, we observe

the similarities:

1. Both the regular fee and the loading increase with σ;
2. There is a positive correlation between the regular fee and the loading,

and the differences:

1. The regular fee for the annual ratchet VA is independent of the length of the contract;
2. The loading of the annual ratchet VA increase with μ and g. As the level of ratcheting is determined by both the guaranteed growth rate g and the drift parameter of the sub-account μ, an increase in either of them will lead to a higher return and thus more charges.

	1st moment	2nd moment	3rd moment	4th moment
Monte Carlo	-0.0152	2.8349	-1.9119	29.7832
Semi-analytic	-0.0154	2.8496	-2.4526	32.4185

Table 5.2.6: Moments of the discrete hedging costs for a 2-year annual ratchet VA. The common parameters are $S_0 = 50, g = 0.05, r = 0.03, \mu = 0.1, \sigma = 0.3, d = 0.01, \alpha = 0.1$.

	1st moment	2nd moment	3rd moment	4th moment
Monte Carlo	-0.0249	5.1872	-3.2657	102.8921
Semi-analytic	-0.0245	5.2111	-5.4295	114.6941

Table 5.2.7: Moments of the discrete hedging costs for a 3-year annual ratchet VA. The common parameters are $S_0 = 50, g = 0.05, r = 0.03, \mu = 0.1, \sigma = 0.3, d = 0.01, \alpha = 0.1$.

	1st moment	2nd moment	3rd moment	4th moment
Monte Carlo	-0.0331	8.5345	-4.7317	306.3399
Semi-analytic	-0.0355	8.5571	-10.9250	341.7194

Table 5.2.8: Moments of the discrete hedging costs for a 4-year annual ratchet VA. The common parameters are $S_0 = 50, g = 0.05, r = 0.03, \mu = 0.1, \sigma = 0.3, d = 0.01, \alpha = 0.1$.

	1st moment	2nd moment	3rd moment	4th moment
Monte Carlo	-0.0434	13.2383	-6.2195	833.7747
Semi-analytic	-0.0483	13.2771	-21.0084	936.1415

Table 5.2.9: Moments of the discrete hedging costs for a 5-year annual ratchet VA. The common parameters are $S_0 = 50, g = 0.05, r = 0.03, \mu = 0.1, \sigma = 0.3, d = 0.01, \alpha = 0.1$.

	2-year	3-year	4-year	5-year
True:				
90%	1.9579	2.6876	3.4390	4.2513
95%	2.5176	3.4834	4.5072	5.6472
97.5%	3.0433	4.2298	5.5448	7.0177
99%	3.7097	5.2148	6.9347	8.9095
Monte Carlo:				
90%	1.9783	2.7232	3.3962	4.0850
95%	2.5472	3.5334	4.5751	5.8204
97.5%	3.0766	4.2868	5.7718	7.6024
99%	3.7597	5.2648	7.3191	9.7085
Semi-analytic:				
90%	1.9602	2.8039	3.3158	3.9393
95%	2.5171	3.5927	4.3534	5.3310
97.5%	3.0333	4.2760	5.4003	6.9088
99%	3.7097	5.0731	6.9347	9.2233

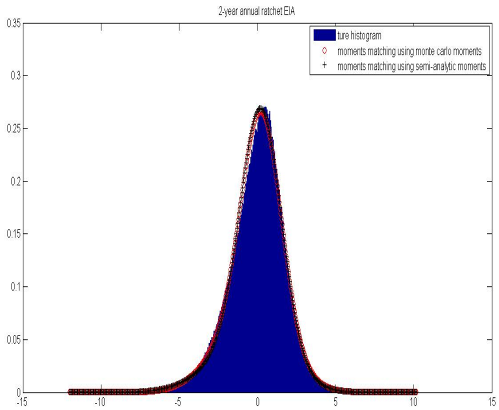
Table 5.2.10: Quantile comparison. The columns contain the quantiles of the cost distributions of hedging annual ratchet VAs with different terms. The first row computes the quantiles by simulation, the total number of iteration is 10^6. The second row uses a mixture model of two normals to match the first four moments of the cost distribution, obtained by simulation. The third row fits the mixture model with the moments given by the semi-analytic algorithm. The common parameters are $S_0 = 50, g = 0.05, r = 0.03, \mu = 0.1, \sigma = 0.3, d = 0.01, \alpha = 0.1$.

	n=2	n=3	n=4	n=5
regular fee	0.2930	0.2930	0.2930	0.2930
loading	0.0227	0.0193	0.0171	0.0158

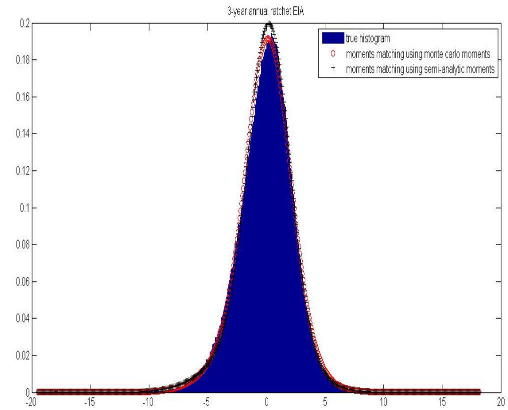
Table 5.2.11: Management fee for discrete hedging of annual ratchet VA. The common parameters are: $S_0 = 50, \mu = 0.1, \sigma = 0.3, r = 0.03, g = 0.05, \alpha = 0.1$.

	$\sigma = 0.2$	$\sigma = 0.3$	$\sigma = 0.4$	$\sigma = 0.5$
regular fee	0.2419	0.2930	0.3426	0.3889
loading	0.0153	0.0158	0.0163	0.0166

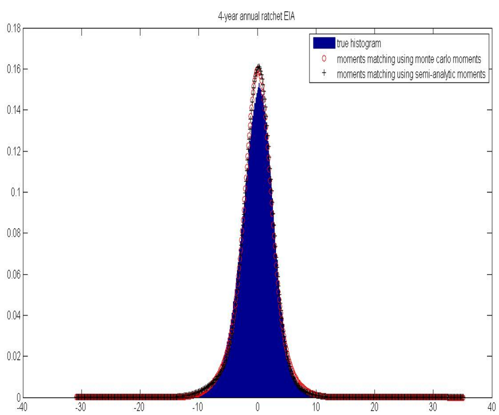
Table 5.2.12: Management fee for discrete hedging of annual ratchet VA. The common parameters are: $S_0 = 50, \mu = 0.1, n = 5, r = 0.03, g = 0.05, \alpha = 0.1$.



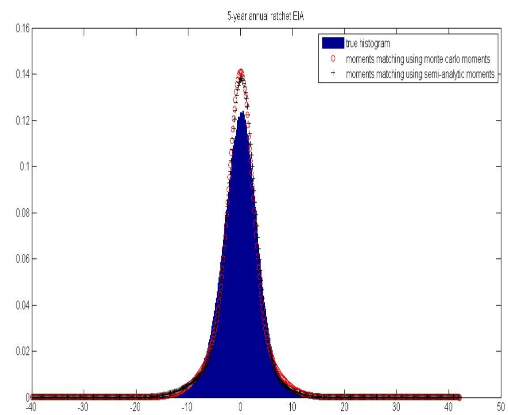
2-year



3-year



4-year



5-year

Figure 5.2.1: Density of the cost distribution of hedging annual ratchet VAs with different terms. The blue area represents the histogram obtained by simulation. The red circle line is the density of a mixture distribution of two normal, fitted to the simulated values of the first four moments of the cost distribution. The black plus sign line is the density of a mixture distribution of two normal, fitted to the values of the first four moments of the cost distribution given by the semi-analytic algorithm. The common parameters are $S_0 = 50, g = 0.05, r = 0.03, \mu = 0.1, \sigma = 0.3, d = 0.01, \alpha = 0.1$.

5.3 Structured Product Based VA

Recently, a new type of variable annuity with payoffs similar to those of structured products rather than those of a mutual fund has been introduced to the market. For example, AXA Equitable made available its first batch of Structured Capital Strategies, structured products inside a variable annuity on Oct. 4, 2010. This instrument allows investors to

	$g = 0.03$	$g = 0.05$	$g = 0.08$	$g = 0.10$
regular fee	0.2515	0.2930	0.3617	0.4097
loading	0.0157	0.0158	0.0159	0.0159

Table 5.2.13: Management fee for discrete hedging of annual ratchet VA. The common parameters are: $S_0 = 50, \mu = 0.1, \sigma = 0.3, n = 5, r = 0.03, \alpha = 0.1$.

	$\mu = 0.06$	$\mu = 0.08$	$\mu = 0.1$	$\mu = 0.12$
regular fee	0.2930	0.2930	0.2930	0.2930
loading	0.0157	0.0158	0.0158	0.0160

Table 5.2.14: Management fee for discrete hedging of annual ratchet VA. The common parameters are: $S_0 = 50, \sigma = 0.3, n = 5, r = 0.03, g = 0.05, \alpha = 0.1$.

select a reference asset (the S&P 500, Russell 2000, MSCI EAFE, gold or oil), a time frame (one, three or five years) and a certain level of downside protection (10%, 20% or 30%). See AXA Equitable (2012) for a detailed description.

In essence, structured products underlying these new annuity contracts partially absorbs losses below some threshold "buffer" and truncate the gains at some "cap". The limited protection from downside risks and limited participation to market growth may seems somewhat undesirable to investors, but on the issuer's side, there is a reason for marketing such products.

Structured notes have been around for years. In the past they have guaranteed full downside protection in exchange for limited upside participation. But with interest rates remaining low and volatility high by historical standards, fewer firms can afford to construct 100% principle protected products cheaply enough to attract buyers. Instead, most of today's versions offer only "buffered" or "contingent" protection. The former cushions the impact of losses to a certain extent while the latter covers losses only until the underlying asset falls to a prescribed level, after which, the protection is cancelled out, leaving the investors alone.

We discuss the cost of discretely hedging this new type of VA in this section.

5.3.1 Structured Capital Strategies

Let us draw a concrete example from Deng et al. (2012) for illustration.

Consider a structured product based variable annuity (spVA) that is linked to a certain index with cap c and buffer b. Suppose the index value is S_0 at the time of issue and S_T at the maturity T of the contract, then the payoff of the annuity is

$$\varphi(S_T) = \begin{cases} S_T + S_0 b & S_T \leq S_0(1-b) \\ S_0 & S_0(1-b) < S_T \leq S_0 \\ S_T & S_0 < S_T \leq S_0(1+c) \\ S_0(1+c) & S_0(1+c) < S_T \end{cases} \tag{5.3.14}$$

In other words, if the sub-account moves downward, the buffer would assume a fraction b of the losses and pass the rest, if any, to the investor; if it moves upward, the investor would earn the profits up to a cap rate of c.

The payoff feature of spVA, however, is quite different from those of the traditional VAs. As we have noted in the case of GMMB, the guarantees wrapped in the traditional VAs can be viewed as a purchased put. The spVA, in contrast, renders the investors a nearly opposite position-a sold put. The payoff diagram of $\varphi(S_T)$ in Figure 5.3.2 demonstrates its similarity to a sold put. In fact, $\varphi(S_T)$ can be decomposed to a combination of: a zero-coupon bond with face value $S_0(1+c)$, a sold put option with strike $S_0(1-b)$, a sold put option with strike $S_0(1+c)$ and a purchased put with strike S_0.

We now turn to the cost analysis for discretely hedging a spVA with payoff $\varphi(S_T)$. As usual, we assume the reference index follows a GBM. We also ignore the zero-coupon component in the payoff, for it can be hedged statically. The hedging target thus becomes

$$\varphi(S_T) = (S_0 - S_T)_+ - (S_0(1-b) - S_T)_+ - (S_0(1+c) - S_T)_+. \tag{5.3.15}$$

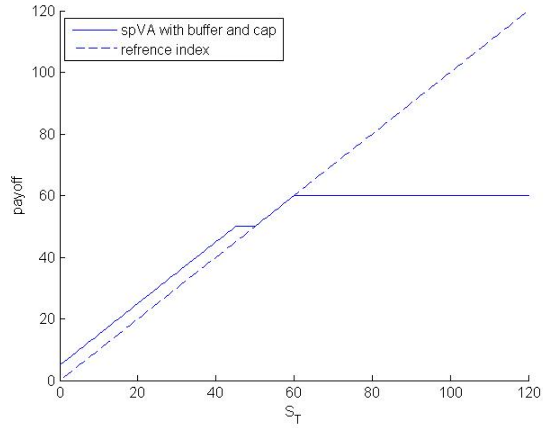


Figure 5.3.2: Payoff Diagram of spVA

Note that (5.3.15) is always non-positive, for one of the sold put (the one with strike $S_0(1+c)$) has a higher strike and thus worths more than the purchased put.

Though the algorithm we developed in Chapter 4 is designed for European put option, it can be generalized to the case of any other path-independent options in a natural way. Specifically, we randomize the maturity by an independent exponential random variable $\epsilon_\lambda^{(1)}$ to obtain a recursive formula for each re-balancing cost. The first cost has an almost identical form to that in (4.2.1)

$$e^{-r(\tau_1 \wedge \epsilon_1^\lambda)} \left[P^{sp}_{\tau_1 \wedge \epsilon_1^\lambda} - \left(M_0^{sp} e^{r(\tau_1 \wedge \epsilon_1^\lambda)} + \Delta_0^{sp} S_{\tau_1 \wedge \epsilon_1^\lambda} \right) \right], \tag{5.3.16}$$

where P_t^{sp} and Δ_t^{sp} are the time-t price and Delta of the spVA at time t (these are simply the linear combination of the prices and Deltas of the put options in $\varphi(S_T)$), $M_t^{sp} = P_t^{sp} - \Delta_t^{sp}$, τ_1 is the hitting time of the band and ϵ_1^λ is the exponential maturity of the contract.

Denote the expectation of (5.3.16) as $h^{sp}(S_0)$, then conditioning on $\mathcal{F}_{\tau^{(1)}+\cdots+\tau^{(n-1)}}$ and

$\epsilon_\lambda^{(1)} > \tau^{(1)} + \cdots + \tau^{(n-1)}$, the conditional expectation of the $(n+1)$-th cost is

$$h^{sp}(S_{\tau^{(1)}+\tau^{(2)}+\ldots+\tau^{(n-1)}})e^{-(r+\lambda)(\tau^{(1)}+\tau^{(2)}+\ldots+\tau^{(n-1)})}, \tag{5.3.17}$$

which is almost the same as (4.2.6) except that h is replaced by h^{sp}.

Therefore, the unconditional expectation is

$$\sum_{i=0}^{n} C_n^i h^{sp}(S_0 e^{(2i-n)\alpha}) L_\alpha^i L_{-\alpha}^{n-i},$$

where $L_{\pm\alpha} = E\left[e^{-(r+\lambda)\tau_{\pm\alpha}}\right]$.

The total expected cost is the sum of individuals'.

Recall that for put option, a good approximation to the higher moments of the discrete hedging cost can be achieved by assuming the independence of the individual costs. Since the payoff of the spVA resembles that of the short put, we have a strong argument to make the same assumption. As it turns out in Table 5.3.15, this gives us good results once again.

With the availability of the first four moments, we fit a mixture of two normals model to the cost distribution. The fitted density is plotted in Figure 5.3.3 and the fitted quantiles are presented in Table 5.3.16.

Finally, we apply the "Percentile Premium Principle" to calculate the fair value of the fee for spVA. As before, the fee breaks into two parts, with the regular fee used to cover the cost of continuous hedging and the additional loading for the discrete re-balancing. In a frictionless market, the continuous hedging cost of the spVA is given by its arbitrary free price. And since the payoff function $\varphi(S_T)$ of the spVA can be decomposed as a linear combination of three vanilla puts with strikes $S_0, S_0(1+c)$ and $S_0(1-b)$, the price of the spVA is equal to

$$P(S_0, S_0, r, \delta, \sigma, T) - P(S_0, S_0(1-b), r, \delta, \sigma, T) - P(S_0, S_0(1+c), r, \delta, \sigma, T), \quad (5.3.18)$$

where $P(S_0, K, r, \delta, \sigma, T)$ is the Black-Scholes price of a European put with the current value of the underlier being S_0, strike K, risk free interest rate r, dividend yield d and time to maturity T. At rate δ, the present value of the insurer's revenue over $[0, T]$ is

$$E^Q(\int_0^T e^{-rt}\delta S_t) = \delta S_0 T. \quad (5.3.19)$$

The fair value for the regular fee is obtained by equating (5.3.18) with (5.3.19).

For the fair value of the loading, we set the 95% quantile of the distribution of the total re-balancing cost-$Q_0.95(\delta, \mu, \sigma, r, \alpha, b, c, S_0, T)$ be equal to the expected revenue under the real probability measure

$$E^P(\int_0^T e^{-rt}\delta S_t dt) = S_0\delta\frac{e^{(\mu-r-\delta)T} - 1}{\mu - r - \delta}.$$

The results are summarized in Table 5.3.17 to 5.3.20, where the "regular fee with bond" is obtained by retaining the bond component (a zero-coupon bond with face value $S_0(1+c)$) in the spVA while the "regular fee w/o bond" drops the bond, leaving only the option part, which has payoff and price. From these tables, several trends can be observed

1. The regular fee for the option part of the spVA is negative. This confirms our earlier assertion that the spVA essentially render the beneficiary a sold put, which means he/she should receive , rather than pay, premiums. Moreover, the value of the sold put increase as the cap level c increases and the buffer b decreases. So the premiums he/she receives (absolute values of the "regular fee w/o bond") increase with c and decrease with b;

2. In contrary to the VA products we have seen before, there is no certain codependency between the “regular fee” and the “loading”. As b increases, the “regular fee with bond” increases while the absolute value of the “regular fee without bond” and the loading decreases (Indeed, as b increases, the flat part in Figure 5.3.2 becomes wider and therefore the move-based hedging does a better job, all fees drop accordingly); As c increases, the regular fee with bond, the absolute value of the “regular fee without bond” and the loading all increase; As σ increases, both the regular fee with bond and the loading decrease .However, the absolute value of the “regular fee w/o bond” increase with σ since the option the beneficiary shorts becomes more expensive. And the increased option value in turn drags down the value of the spVA, given the bond part is insensitive to the volatility.);

3. The regular fee is independent with μ while the loading decreases with it. This is intuitively clear, in that the payoff of spVA is capped at c and thus unaffected by the variability of the underlying asset at its high levels. The impacts on the loading of μ and the cap are similar.

5.3.2 Structured Notes with Contingent Protection

In this section, we consider the discrete hedging problem for structured notes with contingent protection. As a concrete example, let us look at one of UBS AG’s Return Optimization Securities with Contingent Protection whose payoff is summarized in Kim and Levisohn (2010):

Investors get 100% principal protection as long as the Standard & Poor’s 500-stock index hasn’t fallen more than 30% at the end of the product’s three-year term. If the index falls more than 30%, investors suffer all the losses. If the markets fall by less than 30%,

	$b = 0.1$ $c = 0.2$	$b = 0.1$ $c = 0.3$	$b = 0.1$ $c = 0.4$	$b = 0.2$ $c = 0.2$	$b = 0.2$ $c = 0.3$	$b = 0.2$ $c = 0.4$
moments-true:						
1st	-0.0008	0.0124	0.0207	-0.0057	0.0088	0.0158
2nd	1.0548	1.2422	1.3945	0.9683	1.2147	1.3870
3rd	0.4699	0.7058	1.0425	0.2860	0.5486	0.8014
4th	4.7803	7.0909	10.0097	3.7790	6.3520	8.8566
moments-independence:						
1st	-0.0008	0.0124	0.0207	-0.0057	0.0088	0.0158
2nd	1.0659	1.2408	1.3978	0.9695	1.2074	1.3917
3rd	0.6095	0.8713	1.1919	0.4104	0.6873	0.9427
4th	4.5487	6.5165	8.8230	3.9483	6.4214	8.7648

Table 5.3.15: Moments comparison. b is the buffer level and c the cap level. The common parameters are: $T = 3, S_0 = 50, \mu = 0.1, \sigma = 0.3, r = 0.03, \alpha = 0.1, d = 0.02$.

	$b = 0.1$ $c = 0.2$	$b = 0.1$ $c = 0.3$	$b = 0.1$ $c = 0.4$	$b = 0.2$ $c = 0.2$	$b = 0.2$ $c = 0.3$	$b = 0.2$ $c = 0.4$
True:						
90%	1.2933	1.3886	1.4309	1.2454	1.3919	1.4579
95%	1.8262	1.9748	2.1145	1.7270	1.9617	2.0614
97.5%	2.3087	2.5893	2.7819	2.1809	2.4999	2.6998
99%	2.9439	3.2554	3.6068	2.7154	3.1521	3.5260
Semi-analytic:						
90%	1.2829	1.3837	1.4204	1.2028	1.3430	1.4159
95%	1.8599	2.0337	2.0740	1.7394	1.9617	2.0181
97.5%	2.4071	2.6483	2.8064	2.2541	2.5639	2.6999
99%	3.0286	3.3357	3.6804	2.8439	3.2501	3.5673

Table 5.3.16: Quantile comparison. The columns contain the quantiles of the cost distributions of hedging spVA with different buffer and cap levels. The first row computes the quantiles by simulation, the total number of iteration is 10^6. The second row uses a mixture model of two normals to match the first four moments of the cost distribution, obtained by the semi-analytic algorithm. The common parameters are $S_0 = 50, r = 0.03, \mu = 0.1, \sigma = 0.3, d = 0.02, T = 3, \alpha = 0.1$.

	$b = 0.1$ $c = 0.2$	$b = 0.1$ $c = 0.3$	$b = 0.1$ $c = 0.4$	$b = 0.2$ $c = 0.2$	$b = 0.2$ $c = 0.3$	$b = 0.2$ $c = 0.4$
regular fee with bond	0.2028	0.2037	0.2044	0.2209	0.2217	0.2223
regular fee w/o bond	-0.0538	-0.0664	-0.0792	-0.0468	-0.0603	-0.0738
loading	0.0111	0.0122	0.0131	0.0106	0.0117	0.0128

Table 5.3.17: Management fee for spVA. b is the buffer level and c the cap level. The common parameters are: $T = 3, S_0 = 50, \mu = 0.1, \sigma = 0.3, r = 0.03, \alpha = 0.1$. The "regular fee w/o bond" is the fee charged for continuously hedging the options part (excluding the bond in the payoff) of the spVA.

	$b = 0.1$ $c = 0.2$	$b = 0.1$ $c = 0.3$	$b = 0.1$ $c = 0.4$	$b = 0.2$ $c = 0.2$	$b = 0.2$ $c = 0.3$	$b = 0.2$ $c = 0.4$
regular fee with bond	0.1958	0.1976	0.1990	0.2139	0.2155	0.2168
regular fee w/o bond	-0.0704	-0.0836	-0.0969	-0.0615	-0.0755	-0.0895
loading	0.0109	0.0118	0.0127	0.0105	0.0117	0.0121

Table 5.3.18: Management fee for spVA. b is the buffer level and c the cap level. The common parameters are: $T = 3, S_0 = 50, \mu = 0.1, \sigma = 0.4, r = 0.03, \alpha = 0.1$.

	$b = 0.1$ $c = 0.2$	$b = 0.1$ $c = 0.3$	$b = 0.1$ $c = 0.4$	$b = 0.2$ $c = 0.2$	$b = 0.2$ $c = 0.3$	$b = 0.2$ $c = 0.4$
regular fee with bond	0.2028	0.2037	0.2044	0.2209	0.2217	0.2223
regular fee w/o bond	-0.0538	-0.0664	-0.0792	-0.0468	-0.0603	-0.0738
loading	0.0092	0.0105	0.0118	0.0085	0.0103	0.0112

Table 5.3.19: Management fee for spVA. b is the buffer level and c the cap level. The common parameters are: $T = 3, S_0 = 50, \mu = 0.2, \sigma = 0.3, r = 0.03, \alpha = 0.1$.

	$b = 0.1$ $c = 0.2$	$b = 0.1$ $c = 0.3$	$b = 0.1$ $c = 0.4$	$b = 0.2$ $c = 0.2$	$b = 0.2$ $c = 0.3$	$b = 0.2$ $c = 0.4$
regular fee with bond	0.1958	0.1976	0.1990	0.2139	0.2155	0.2168
regular fee w/o bond	-0.0704	-0.0836	-0.0969	-0.0615	-0.0755	-0.0895
loading	0.0095	0.0104	0.0115	0.0086	0.0098	0.0110

Table 5.3.20: Management fee for spVA. b is the buffer level and c the cap level. The common parameters are: $T = 3, S_0 = 50, \mu = 0.2, \sigma = 0.4, r = 0.03, \alpha = 0.1$.

investors get back their principal at the end of product's term. If the index rises (the end-of-period value), investors earn 1.5 times the upside, up to a cap of 58.6%, which they get if the index is up 39%. Fees, also called the "underwriting discount", are 2.5%. The path-dependent part in this product is exactly the same as that of a down-and-out put (DOP) option, which is put that goes out of existence if the asset price falls to hit a barrier before maturity.

As a counterpart to the down-and-out put, a down-and-in put comes into existence if the asset price falls to hit a barrier before maturity. At a first glance, one may allege that insurers should market the down-and-in put rather than the down-and-out, for the former provides well-timed protection and should therefore be favored by policyholders over the latter. However, wise insurers would reject this proposal, for the same argument. The reason lies in the fact that the policyholders and the insurers have opposite interests. A well-timed protection for the investor means an ill-timed disaster for the insurer. Insurers issuing down-and-in put would face non-diversifiable risks in adverse market conditions, when every policyholder would claim for rescue. So the down-and-out put is indeed a prudent choice and we now move on to its hedging.

Suppose we sell a down-and-out put written on the sub-account with strike price K and barrier $B(B < K)$. If the sub-account value never hit the barrier, the payoff of the DOP at maturity will be the same as that of a standard put. Otherwise, the DOP is knocked out at the moment of hitting and we are free thereafter.

The price of DOP (with $K > B$) at time $t < T$ given that it has not been knocked out is (See e.g. Hull (2011))

$$
\begin{aligned}
&P^{DOP}(S_t, K, B, r, \sigma, T-t, d) \\
&\quad = Ke^{-r(T-t)}N(-d_2) - S_t e^{-d(T-t)}N(-d_1) + S_t N(-x_1)e^{-d(T-t)}
\end{aligned}
$$

$$
\begin{aligned}
&- Ke^{-r(T-t)}N(-x_1+\sigma\sqrt{T-t}) - S_t e^{-d(T-t)}(\frac{B}{S_t})^{2\lambda}[N(y)-N(y_1)] \\
&+ Ke^{-r(T-t)}(\frac{B}{S_t})^{2\lambda-2}[N(y-\sigma\sqrt{T-t}) - N(y_1-\sigma\sqrt{T-t})].
\end{aligned}
$$

where S_t is the sub-account value at time t, K is the strike price, B is the barrier level, r is the risk free interest rate, σ is the volatility, T is the maturity, d is the dividend yield and

$$
\begin{aligned}
\lambda =& \frac{r-d+\frac{1}{2}\sigma^2}{\sigma^2}, \\
x_1 =& \frac{\log\frac{S_t}{B}}{\sigma\sqrt{T-t}} + \lambda\sigma\sqrt{T-t}, \\
y_1 =& \frac{\log\frac{B}{S_t}}{\sigma\sqrt{T-t}} + \lambda\sigma\sqrt{T-t}, \\
y =& \frac{\log\frac{B^2}{KS_t}}{\sigma\sqrt{T-t}} + \lambda\sigma\sqrt{T-t}.
\end{aligned}
$$

The Delta of this option is given by:

$$
\begin{aligned}
&\Delta^{DOP}(S_t, K, B, r, \sigma, T-t, d) \\
&= -N(-d_1)e^{-d(T-t)} + N(-x_1)e^{-d(T-t)} - e^{-d(T-t)}\frac{\phi(x_1)}{\sigma\sqrt{T-t}} \\
&+ Ke^{-r(T-t)}\frac{\phi(x_1-\sigma\sqrt{T-t})}{S_t\sigma\sqrt{T-t}} - e^{-d(T-t)}B^{2\lambda}(1-2\lambda){S_t}^{-2\lambda}[N(y)-N(y_1)] \\
&- e^{-d(T-t)}B^{2\lambda}{S_t}^{1-2\lambda}\frac{\phi(y_1)-\phi(y)}{S_t\sigma\sqrt{T-t}} \\
&+ Ke^{-r(T-t)}B^{2\lambda-2}(2-2\lambda){S_t}^{1-2\lambda}[N(y-\sigma\sqrt{T-t}) - N(y_1-\sigma\sqrt{T-t})] \\
&+ Ke^{-r(T-t)}B^{2\lambda-2}{S_t}^{2-2\lambda}\frac{\phi(y_1-\sigma\sqrt{T-t}) - \phi(y-\sigma\sqrt{T-t})}{S_t\sigma\sqrt{T-t}}.
\end{aligned}
$$

where $\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$.

Suppose the sub-account value starts at S_0. As a matter of fact, DOP is a path-dependent option and the barrier level B can take any value smaller than K. Consequently in the hitting time problem, we have to deal with a new barrier B in addition to the three barriers $S = S_0 e^{\alpha}, S = S_0 e^{-\alpha}, t = T$ for the standard put. This could be troublesome in general. To circumvent this difficulty, we let B coincide with one of the old barriers, i.e. $B = S_0 e^{-m\alpha}$ for some positive integer m. For a general B, we may find m s.t. $B_- = S_0 e^{-(m-1)\alpha} < B < S_0 e^{-m\alpha} = B_+$, compute the hedging cost for B_- and B_+ and approximate the cost for B by interpolation.

Consider the expectation of the discounted cost at the first truncated stopping time $\tau^{(1)} \wedge \epsilon_\lambda^{(1)}$. Because $B = S_0 e^{-m\alpha}$, the DOP will not be knocked out during $[0, \tau^{(1)} \wedge \epsilon_\lambda^{(1)})$, so the expectation of the discounted cost at the first truncated stopping time $\tau^{(1)} \wedge \epsilon_\lambda^{(1)}$ will have the same form as that of the standard put, except the price P, delta Δ and money market account M are replaced by those of DOP. We denote this value by $h^{DOP}(S_0)$.

The discounted value for the cost of the $(n+1)$-th hit is

$$
\begin{aligned}
&I_{\{\epsilon_\lambda^{(1)} > \xi_n\}} I_{\{S_0 > B\}} I_{\{S_{\tau^{(1)}} > B\}} \cdots I_{\{S_{\xi_n} > B\}} e^{-r(\xi_n + \tau^{(n+1)} \wedge \epsilon_\lambda^{(n+1)})} \\
&\underbrace{\left[P^{DOP}_{\xi_n + \tau^{(n+1)} \wedge \epsilon_\lambda^{(n+1)}} - \left(M^{DOP}_{\xi_n} e^{r(\tau^{(n+1)} \wedge \epsilon_\lambda^{(n+1)})} + \Delta^{DOP}_{\xi_n} e^{d(\tau^{(n+1)} \wedge \epsilon_\lambda^{(n+1)})} S_{\xi_n + \tau^{(n+1)} \wedge \epsilon_\lambda^{(n+1)}} \right) \right]}_{D_{n+1}},
\end{aligned}
\tag{5.3.20}
$$

where $\xi_n = \tau^{(1)} + \cdots + \tau^{(n)}$.

The first indicator in (5.3.20) says that we have not reached the maturity after the n-th hit and the next $(n+1)$ indicators say that the DOP has not been knocked out by time $\tau^{(1)} + \cdots + \tau^{(n)}$ (this is another advantage we get by restricting $B = S_0 e^{-m\alpha}$. In general, we need to check the whole path to see whether the barrier is hit. But when $B = S_0 e^{-m\alpha}$, we only need to check this condition at each hitting time of the moving band). With these two conditions satisfied, we will continue to hedge the cost at the $(n+1)$-th hit and the form of the cost at the $(n+1)$-th hit is exactly the same as that of the first.

To compute its expectation, we condition on $\mathcal{F}_{\tau^{(1)} + \cdots + \tau^{(n)}}$ and $\epsilon_\lambda^{(1)} > \tau^{(1)} + \cdots + \tau^{(n)}\}$.

The conditional expectation is

$$
\begin{aligned}
&e^{-(r+\lambda)(\tau^{(1)}+\cdots+\tau^{(n)})} I_{\{S_0>B\}} I_{\{S_{\tau^{(1)}}>B\}} \cdots I_{\{S_{\tau^{(1)}+\cdots+\tau^{(n)}}>B\}} \\
&E\left[e^{-r(\tau^{(n+1)}\wedge\epsilon_\lambda^{(n+1)})} D_{n+1} | \mathcal{F}_{\tau^{(1)}+\cdots+\tau^{(n)}}, \epsilon_\lambda^{(1)} > \tau^{(1)} + \cdots + \tau^{(n)}\right] \\
&\quad = e^{-(r+\lambda)(\tau^{(1)}+\cdots+\tau^{(n)})} I_{\{S_0>B\}} I_{\{S_{\tau^{(1)}}>B\}} \cdots I_{\{S_{\tau^{(1)}+\cdots+\tau^{(n)}}>B\}} h^{DOP}(S_{\tau^{(1)}+\cdots+\tau^{(n)}}).
\end{aligned}
\tag{5.3.21}
$$

For the unconditional expectation, first note that when $n < m$,

$$
I_{\{S_0>B\}} I_{\{S_{\tau^{(1)}}>B\}} \cdots I_{\{S_{\tau^{(1)}+\cdots+\tau^{(n)}}>B\}} = 1.
$$

So in this case, we only need to compute

$$
e^{-(r+\lambda)(\tau^{(1)}+\cdots+\tau^{(n)})} h^{DOP}(S_{\tau^{(1)}+\cdots+\tau^{(n)}}),
$$

which can be done in a way similar to that for the standard put, with h replaced by h^{DOP}. The result is

$$
E[e^{-(r+\lambda)(\tau^{(1)}+\cdots+\tau^{(n)})} h^{DOP}(S_{\tau^{(1)}+\cdots+\tau^{(n)}})] = \sum_{i=0}^{n} C_n^i h^{DOP}(S_0 e^{(2i-n)\alpha}) L_\alpha^i L_{-\alpha}^{n-i},
$$

where $L_\alpha = E\left[e^{-(r+\lambda)\tau_\alpha}\right]$ and $L_{-\alpha} = E\left[e^{-(r+\lambda)\tau_{-\alpha}}\right]$.

When $n = m$, $I_{\{S_0>B\}} I_{\{S_{\tau^{(1)}}>B\}} \cdots I_{\{S_{\tau^{(1)}+\cdots+\tau^{(n)}}>B\}} = I_{\{S_{\tau^{(1)}+\cdots+\tau^{(m)}}>B\}}$. So we need to compute

$$
\begin{aligned}
&E[e^{-(r+\lambda)(\tau^{(1)}+\cdots+\tau^{(m)})} I_{\{S_{\tau^{(1)}+\cdots+\tau^{(m)}}>B\}} h^{DOP}(S_{\tau^{(1)}+\cdots+\tau^{(m)}})] \\
&= \sum_{S_0 e^{k\alpha}>B} h^{DOP}(S_0 e^{k\alpha}) E[e^{-(r+\lambda)(\tau^{(1)}+\cdots+\tau^{(m)})} I_{\{S_{\tau^{(1)}+\cdots+\tau^{(m)}}=S_0 e^{k\alpha}\}}].
\end{aligned}
\tag{5.3.22}
$$

Both h^{DOP} and $E[e^{-(r+\lambda)(\tau^{(1)}+\cdots+\tau^{(m)})} I_{\{S_{\tau^{(1)}+\cdots+\tau^{(m)}}=S_0 e^{k\alpha}\}}] = C_m^i L_\alpha^i L_{-\alpha}^{m-i}, i = \frac{m+k}{2}$ are known, so is (5.3.22).

When $n > m$, we need to compute

$$E[e^{-(r+\lambda)(\tau^{(1)}+\cdots+\tau^{(n)})}I_{\{S_0>B\}}I_{\{S_{\tau^{(1)}}>B\}}\cdots I_{\{S_{\tau^{(1)}+\cdots+\tau^{(n)}}>B\}}h^{DOP}(S_{\tau^{(1)}+\cdots+\tau^{(n)}})].$$

By conditioning on $S_{\tau^{(1)}+\cdots+\tau^{(n)}}$, we can write it as:

$$\sum_{S_0e^{k\alpha}>B} E[e^{-(r+\lambda)(\tau^{(1)}+\cdots+\tau^{(n)})}I_{\{S_0>B\}}I_{\{S_{\tau^{(1)}}>B\}}\cdots I_{\{S_{\tau^{(1)}+\cdots+\tau^{(n)}}=S_0e^{k\alpha}\}}]h^{DOP}(S_0e^{k\alpha}).$$

The key input is $E[e^{-(r+\lambda)(\tau^{(1)}+\cdots+\tau^{(n)})}I_{\{S_0>B\}}I_{\{S_{\tau^{(1)}}>B\}}\cdots I_{\{S_{\tau^{(1)}+\cdots+\tau^{(n)}}=S_0e^{k\alpha}\}}]$, which admits a recursive relation, as shown below.

If the value of the sub-account at the n-th hitting time is $S_0e^{k\alpha}$, its value at the $(n-1)$-th hitting time can only be either $S_0e^{(k-1)\alpha}$ or $S_0e^{(k+1)\alpha}$. Taking advantage of this fact, we get

$$\begin{aligned}
&E[e^{-(r+\lambda)(\tau^{(1)}+\cdots+\tau^{(n)})}I_{\{S_0>B\}}I_{\{S_{\tau^{(1)}}>B\}}\cdots I_{\{S_{\tau^{(1)}+\cdots+\tau^{(n)}}=S_0e^{k\alpha}\}}]\\
&= E[e^{-(r+\lambda)(\tau^{(1)}+\cdots+\tau^{(n)})}I_{\{S_0>B\}}I_{\{S_{\tau^{(1)}}>B\}}\cdots I_{\{S_{\tau^{(1)}+\cdots+\tau^{(n-1)}}=S_0e^{(k-1)\alpha}\}}I_{\{S_{\tau^{(1)}+\cdots+\tau^{(n)}}=S_0e^{k\alpha}\}}]\\
&\quad + E[e^{-(r+\lambda)(\tau^{(1)}+\cdots+\tau^{(n)})}I_{\{S_0>B\}}I_{\{S_{\tau^{(1)}}>B\}}\cdots I_{\{S_{\tau^{(1)}+\cdots+\tau^{(n-1)}}=S_0e^{(k+1)\alpha}\}}I_{\{S_{\tau^{(1)}+\cdots+\tau^{(n)}}=S_0e^{k\alpha}\}}]\\
&= E[e^{-(r+\lambda)(\tau^{(1)}+\cdots+\tau^{(n)})}I_{\{S_0>B\}}I_{\{S_{\tau^{(1)}}>B\}}\cdots I_{\{S_{\tau^{(1)}+\cdots+\tau^{(n-1)}}=S_0e^{(k-1)\alpha}\}}I_{\{\tau^{(n)}=\tau_\alpha\}}]\\
&\quad + E[e^{-(r+\lambda)(\tau^{(1)}+\cdots+\tau^{(n)})}I_{\{S_0>B\}}I_{\{S_{\tau^{(1)}}>B\}}\cdots I_{\{S_{\tau^{(1)}+\cdots+\tau^{(n-1)}}=S_0e^{(k+1)\alpha}\}}I_{\{\tau^{(n)}=\tau_{-\alpha}\}}].
\end{aligned} \tag{5.3.23}$$

The first expectation term on the RHS of (5.3.23) can be computed in two steps. Conditioning on $\mathcal{F}_{\tau^{(1)}+\cdots+\tau^{(n-1)}}$, the conditional expectation is

$$\begin{aligned}
&e^{-(r+\lambda)(\tau^{(1)}+\cdots+\tau^{(n-1)})}I_{\{S_0>B\}}I_{\{S_{\tau^{(1)}}>B\}}\cdots I_{\{S_{\tau^{(1)}+\cdots+\tau^{(n-1)}}=S_0e^{(k-1)\alpha}\}}\\
&E[e^{-(r+\lambda)(\tau^{(n)})}I_{\{\tau^{(n)}=\tau_\alpha\}}|\mathcal{F}_{\tau^{(1)}+\cdots+\tau^{(n-1)}}]
\end{aligned}$$

$$= e^{-(r+\lambda)(\tau^{(1)}+\cdots+\tau^{(n-1)})} I_{\{S_0>B\}} I_{\{S_{\tau^{(1)}}>B\}} \cdots I_{\{S_{\tau^{(1)}+\cdots+\tau^{(n-1)}}=S_0 e^{(k-1)\alpha}\}}$$

$$E[e^{-(r+\lambda)(\tau^{(1)})} I_{\{\tau^{(1)}=\tau_\alpha\}}]$$

$$= e^{-(r+\lambda)(\tau^{(1)}+\cdots+\tau^{(n-1)})} I_{\{S_0>B\}} I_{\{S_{\tau^{(1)}}>B\}} \cdots I_{\{S_{\tau^{(1)}+\cdots+\tau^{(n-1)}}=S_0 e^{(k-1)\alpha}\}}$$

$$E[e^{-(r+\lambda)(\tau_\alpha)} I_{\{\tau^{(1)}=\tau_\alpha\}}]$$

$$= e^{-(r+\lambda)(\tau^{(1)}+\cdots+\tau^{(n-1)})} I_{\{S_0>B\}} I_{\{S_{\tau^{(1)}}>B\}} \cdots I_{\{S_{\tau^{(1)}+\cdots+\tau^{(n-1)}}=S_0 e^{(k-1)\alpha}\}} L_\alpha.$$

and the full expectation is

$$E[e^{-(r+\lambda)(\tau^{(1)}+\cdots+\tau^{(n-1)})} I_{\{S_0>B\}} I_{\{S_{\tau^{(1)}}>B\}} \cdots I_{\{S_{\tau^{(1)}+\cdots+\tau^{(n-1)}}=S_0 e^{(k-1)\alpha}\}}] L_\alpha.$$

Similarly, the second term on the RHS of (5.3.23) is

$$E[e^{-(r+\lambda)(\tau^{(1)}+\cdots+\tau^{(n-1)})} I_{\{S_0>B\}} I_{\{S_{\tau^{(1)}}>B\}} \cdots I_{\{S_{\tau^{(1)}+\cdots+\tau^{(n-1)}}=S_0 e^{(k+1)\alpha}\}}] L_{-\alpha}.$$

In summary, if we let

$$f(n,k) = E[e^{-(r+\lambda)(\tau^{(1)}+\cdots+\tau^{(n)})} I_{\{S_0>B\}} I_{\{S_{\tau^{(1)}}>B\}} \cdots I_{\{S_{\tau^{(1)}+\cdots+\tau^{(n)}}=S_0 e^{k\alpha}\}}],$$

then

$$f(n,k) = f(n-1,k+1)L_{-\alpha} + f(n-1,k-1)L_\alpha. \tag{5.3.24}$$

The recursion (5.3.24) starts at $n = m$ with initial condition

$$f(m,k) = E[e^{-(r+\lambda)(\tau^{(1)}+\cdots+\tau^{(m)})} I_{\{S_{\tau^{(1)}+\cdots+\tau^{(m)}}=S_0 e^{k\alpha}\}}] = C_m^i L_\alpha^i L_{-\alpha}^{m-i},$$

$$i = \frac{m+k}{2},$$

$$k = -m+2, -m+4, \cdots, m,$$

	$m=2$	$m=3$	$m=4$	$m=5$
moments-true:				
1st	-0.0107	0.0473	0.0911	0.1120
2nd	0.6101	1.2886	1.9160	2.1738
3rd	18.7373	21.1829	22.2021	23.8058
4th	1331.2273	1255.2929	1194.0548	1034.2894
moments-semi analytic:				
1st	-0.0124	0.0406	0.0826	0.1095
2nd	0.6328	1.3876	2.0269	2.2248
3rd	16.0309	19.8174	20.8156	21.2746
4th	1163.9510	1095.0649	974.4574	888.2653

Table 5.3.21: Moments comparison. The barrier level is $B = S_0 e^{-m\alpha}$. The row "moments-true" is obtained by simulation and the row "moments-semi analytic" is obtained by the semi-analytic algorithm with the assumption that individual costs are independent. The common parameters are: $T = 3, S_0 = K = 50, \mu = 0.1, \sigma = 0.2, r = 0.02, \alpha = 0.1, d = 0$.

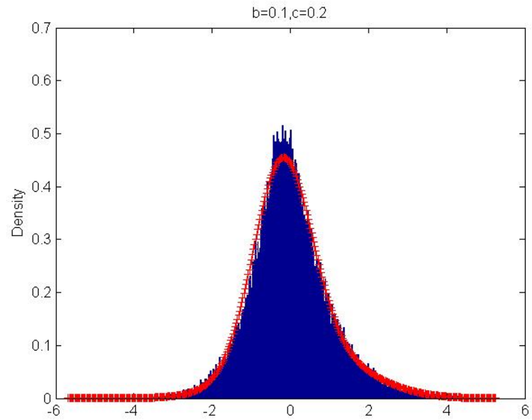
and boundary conditions

$$f(n, -m) = 0, \forall n \geq m, \tag{5.3.25}$$

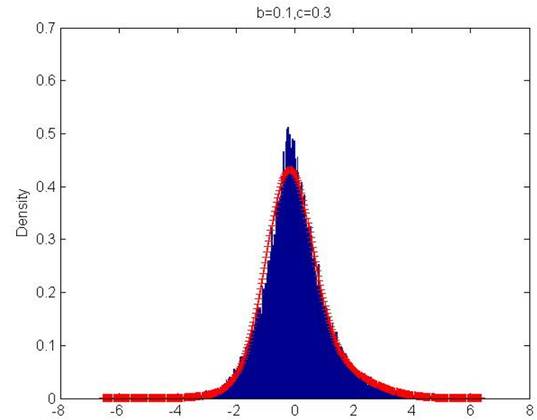
$$f(n, k) = 0, \forall k > n. \tag{5.3.26}$$

In Table 5.3.21, we compare the moments calculated by Monte Carlo simulation and the semi-analytic algorithm, assuming individual costs are independent. These two methods generate similar moments estimators.

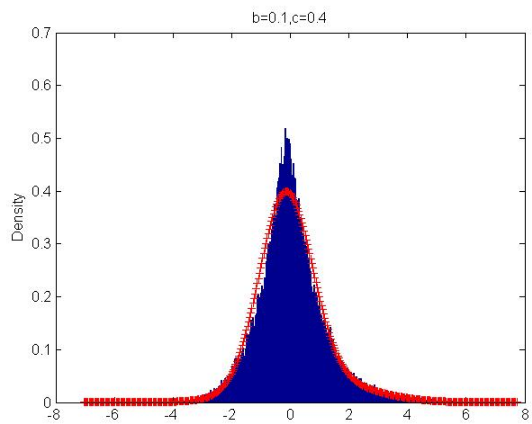
The large fourth moment of the cost distribution suggests a very heavy tail. Indeed, using Monte Carlo simulation with 10^6 iterations, we find, for $m = 3$ and the same common parameters as in Table 5.3.21, the 90% quantile of the cost distribution is 0.3737 while its 99% quantile is 4.6388.



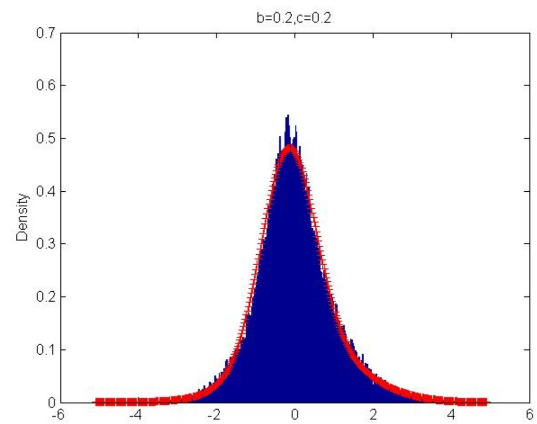
$b = 0.1, c = 0.2$



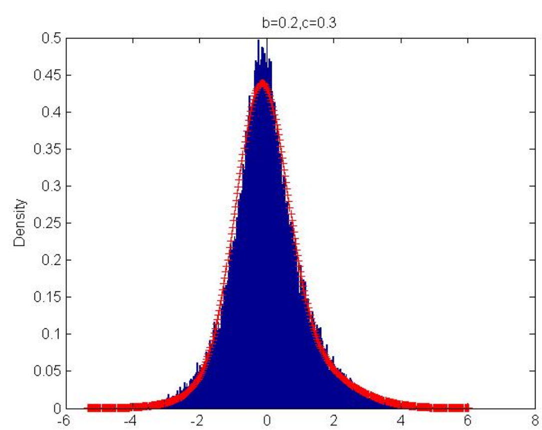
$b = 0.1, c = 0.3$



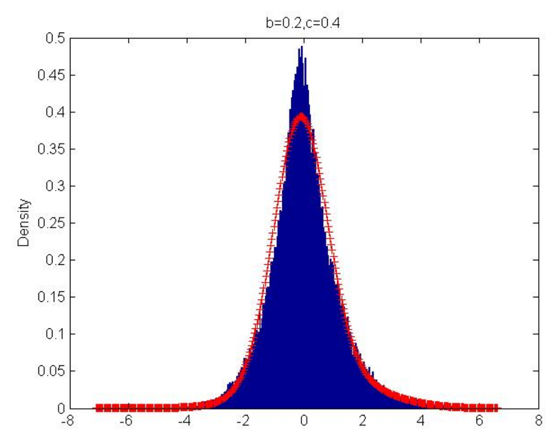
$b = 0.1, c = 0.4$



$b = 0.2, c = 0.2$



$b = 0.2, c = 0.3$



$b = 0.2, c = 0.4$

Figure 5.3.3: Density of the cost distribution of hedging spVA with different buffer and cap levels. The blue area represents the histogram obtained by simulation. The red plus sign line is the density of a mixture distribution of two normal, fitted to the values of the first four moments of the cost distribution given by the semi-analytic algorithm. The common parameters are $S_0 = 50, r = 0.03, \mu = 0.1, \sigma = 0.3, d = 0.02, T = 3, \alpha = 0.1$.

Chapter 6

Extension to Regime Switching GBM

6.1 Regime Switching Models

Since its invention, the Black-Scholes option pricing model has been refined in various ways. The demand for capturing the structural changes in macroeconomic conditions, economic fundamentals, monetary policies and business environment motivates the introduction of Markov regime switching models to the econometrics, finance and actuarial science. These models use an embedded continuous time Markov chain to control the transitions between multiple states of certain economic, financial or actuarial factors, including aggregate return, volatility, interest rate, mortality and so on.

The origin of regime-switching models dates back to Quandt (1958) and Goldfeld and Quandt (1973), which discuss the parameter estimation of a linear regression system obeying two separate regimes. Whereafter, in an influential paper, Hamilton (1989) suggests Markov switching techniques as a method for modeling non-stationary time series. For option pricing under regime switching models, Naik (1993) provides an elegant treatment pricing European option under a regime switching model with two regimes; Buffington and Elliott (2002) extends the approximate valuation of American options due to Barone-Adesi and Whaley to a Black-Scholes economy with regime switching; Elliott et al. (2005) adopts a regime switching random Esscher transform to determine an equivalent martingale pricing measure and justify their pricing result by the minimal entropy martingale measure (MEMM); Elliott et al. (2007) studies the price of European and American option under a generalized Markov-modulated jump diffusion model using the generalized Esscher transform and coupled partial-differential-integral equations; Boyle and Draviamb (2007) considers the pricing of exotic options using a PDE method, Surkov et al. (2007) present a new, efficient algorithm, based on transform methods, which symmetrically treats the diffusive and integrals terms, is applicable to a wide class of path-dependent options (such as Bermudan, barrier, and shout options) and options on multiple assets, and naturally extends to regime-switching Levy models. In the actuarial literature, Hardy (2001) popularizes the use of regime switching model for pricing and hedging long term investment guarantee products and fits the model to the monthly data from the Standard and Poor's 500 and the Toronto Stock Exchange 300 indices using a discrete time regime-switching lognormal model; Siu (2005) considers the valuation of participating life insurance policies with surrender options in regime-switching models; Siu et al. (2008) extends the framework of Siu (2005) and investigates the valuation of participating life insurance policies without surrender options to a Markov, regime-switching, jump diffusion case; Lin et al. (2009) considers the pricing problem of various equity-linked annuities and variable annuities under a generalized geometric Brownian motion model with regime switching.

If the number of states in the regime switching model is large, we need either to solve a system of PDEs with the number of PDEs being the number of states of the embedded Markov chain or to perform multiple integrals numerically, both of which can be computationally inefficient. Therefore in practice, we prefer relatively simple model with just two states. It turns out that regime switching model with two regimes is often enough to describe the vicissitudes of the business world, for instance, the economic expansion and recession, the bull and the bear market and the public mentality of optimism and pessimism. See Chapter 11 of Taylor (2005) for empirical evidence.

In this chapter, we consider the discrete hedging problem of variable annuities under the assumption that the sub-account follows a GBM model with two regimes. To this end, we start with the derivation of the state-dependent, defective densities of the two-sided hitting times of the regime switching GBM.

6.2 Hitting Time Distribution for GBM with Two Regimes

In this section, we derive the hitting time distribution of a Regime Switching GBM with two regimes. In particular, we consider two independent arithmetic BM

$$X_t^1 = \mu_1 t + \sigma_1 W_t^1, X_t^2 = \mu_2 t + \sigma_2 W_t^2,$$

where W_t^1 and W_t^2 are a standard Brownian motions under the physical measure P.

Let J_t be an independent continuous Markov process with two states $\{1, 2\}$ and intensity matrix $Q = \begin{bmatrix} -\lambda & \lambda \\ v & -v \end{bmatrix}$, and $\{T_i | i \geq 0\}$ be the jump epochs of J_t. [1]

[1]We assume J_t is observable because, in the incomplete market model that we will introduce shortly, there are not any other securities or derivatives from which the parameters in the intensity matrix of J_t could be inferred.

We define the Markov additive process X_t as

$$
\begin{aligned}
X_t = X_0 + \sum_{n\geq 1} \sum_{1\leq i,j\leq 2, i\neq j} (X^i_{T_n} - X^i_{T_{n-1}}) I_{\{J_{T_{n-1}}=i, J_{T_n}=j, T_n\leq t\}} \\
+ \sum_{n\geq 1} \sum_{1\leq i\leq 2} (X^i_t - X^i_{T_{n-1}}) I_{\{J_{T_{n-1}}=i, T_{n-1}\leq t<T_n\}}.
\end{aligned} \tag{6.2.1}
$$

In other words, the evolution of X_t will switch from X^i_t to X^j_t when J_t transits from state i to state j, where $1 \leq i, j \leq 2, i \neq j$.

The GBM with regime switching is $S_t = S_0 e^{X_t}$. Since the stopping times of S_t can be translated to those of X_t, we focus on the latter hereafter.

Suppose $X_0 = 0$ and $J_0 = 1$, define the two-sided stopping times

$$T_{u,d} = \inf\{t > 0 | X_t = u \quad \text{or} \quad X_t = d\}, \tag{6.2.2}$$

$$T_u = \inf\{t > 0 | X_t = u, X_s > d (\forall 0 \leq s < t)\}, \tag{6.2.3}$$

$$T_d = \inf\{t > 0 | X_t = d, X_s < d (\forall 0 \leq s < t)\}. \tag{6.2.4}$$

As usual, $T_u < \infty$ or $T_{u,d} = T_u$ implies $T_d = \infty$ and $T_d < \infty$ or $T_{u,d} = T_d$ implies $T_u = \infty$. We derive the state-dependent Laplace transform of the T_u and T_d

$$L^u_{1,1}(\theta) = E_{0,1}[e^{-\theta T_u} I_{\{J_{T_u}=1\}}],$$

$$L^u_{1,2}(\theta) = E_{0,1}[e^{-\theta T_u} I_{\{J_{T_u}=2\}}],$$

$$L^d_{1,1}(\theta) = E_{0,1}[e^{-\theta T_d} I_{\{J_{T_d}=1\}}],$$

$$L^d_{1,2}(\theta) = E_{0,1}[e^{-\theta T_d} I_{\{J_{T_d}=2\}}],$$

$$L^u_{2,1}(\theta) = E_{0,2}[e^{-\theta T_u} I_{\{J_{T_u}=1\}}],$$

$$L^u_{2,2}(\theta) = E_{0,2}[e^{-\theta T_u} I_{\{J_{T_u}=2\}}],$$

$$L^d_{2,1}(\theta) = E_{0,2}[e^{-\theta T_d} I_{\{J_{T_d}=1\}}],$$

$$L^d_{2,2}(\theta) = E_{0,2}[e^{-\theta T_d} I_{\{J_{T_d}=2\}}],$$

where $\theta > 0$ and $E_{0,1}(\bullet)$ underscores $X_0 = 0, J_0 = 1$.

Let $\varphi_i(\alpha) = \mu_i\alpha + \frac{1}{2}\sigma_i^2\alpha^2, i = 1, 2$ be the Laplace exponent of X_t^i and

$$F(\alpha) = Q + \begin{bmatrix} \varphi_1(\alpha) & 0 \\ 0 & \varphi_2(\alpha) \end{bmatrix} = \begin{bmatrix} \varphi_1(\alpha) - \lambda & \lambda \\ v & \varphi_2(\alpha) - v \end{bmatrix},$$

then according to Asmussen and Kella (2000),

Theorem 6.2.1 (Asmussen and Kella)**.** *For any initial distribution of* (X_0, J_0) *and every complex* α *such that* $\varphi_i(\alpha)$ *exists,*

$$M^W(t, \alpha) = e^{\alpha X_t}\mathbf{1}_{J_t}e^{-F(\alpha)t}$$

is a (row) vector valued martingale. Consequently, if $h(\alpha)$ *and* $\theta(\alpha)$ *are a right eigenvector and eigenvalue of the matrix* $F(\alpha)$*, then with every initial distribution of* (X_0, J_0)

$$e^{\alpha X_t - \theta(\alpha)t}h_{J_t}(\alpha) \tag{6.2.5}$$

is a martingale, where $h_j(\alpha)$ *is the* j*-th element of* $h(\alpha)$*.*

The two (real) eigenvalues of $F(\alpha)$ are

$$\theta_{1,2}(\alpha) = \frac{1}{2}\left[(\varphi_1(\alpha) - \lambda + \varphi_2(\alpha) - v) \pm \sqrt{(\varphi_1(\alpha) - \lambda - \varphi_2(\alpha) + v)^2 + 4\lambda v}\right]. \tag{6.2.6}$$

For a given $\theta > 0$, there are four α's, denoted by $\alpha_k(\theta), k \in \{1, 2, 3, 4\}$, s.t. θ is an eigenvalue of $F(\alpha_k(\theta))$. In particular, they are the four roots of the following 4-th order polynomial

$$\frac{1}{4}\sigma_1^2\sigma_2^2\alpha^4 + \frac{1}{2}(\mu_2\sigma_1^2 + \mu_1\sigma_2^2)\alpha^3 + \left[\mu_1\mu_2 - \frac{1}{2}v\sigma_1^2 - \frac{1}{2}\lambda\sigma_2^2 - \frac{1}{2}\theta(\sigma_1^2 + \sigma_2^2)\right]\alpha^2 - [v\mu_1 + \lambda\mu_2 + \theta(\mu_1 + \mu_2)]\,\alpha + \theta(\lambda + v) + \theta^2 = 0.$$

For an eigenvalue θ of $F(\alpha)$, its eigenvector $h(\alpha)$ satisfies $\frac{h_1(\alpha)}{h_2(\alpha)} = \frac{-\lambda}{\varphi_1(\alpha) - \lambda - \theta}$. So we take, without loss of generality, $h_1(\alpha) = -\lambda, h_2(\alpha) = \varphi_1(\alpha) - \lambda - \theta$.

Now we make use of (6.2.5) to derive the state-dependent Laplace transforms.

Because for any $\theta > 0$,

$$e^{\alpha(\theta)X_t - \theta t} h_{J_t}(\alpha(\theta))$$

is a martingale, the optional sampling theorem implies

$$E_{0,1}\left[e^{\alpha(\theta)X_{T_{u,d}} - \theta T_{u,d}} h_{J_{T_{u,d}}}(\alpha(\theta))\right] = h_1(\alpha(\theta)). \tag{6.2.7}$$

On the other hand,

$$\begin{aligned}
&E_{0,1}\left[e^{\alpha(\theta)X_{T_{u,d}} - \theta T_{u,d}} h_{J_{T_{u,d}}}(\alpha(\theta))\right] \\
&\quad = E_{0,1}\left[e^{\alpha(\theta)u - \theta T_u} h_{J_{T_u}}(\alpha(\theta)) I_{\{T_{u,d}=T_u\}}\right] + E_{0,1}\left[e^{\alpha(\theta)d - \theta T_d} h_{J_{T_d}}(\alpha(\theta)) I_{\{T_{u,d}=T_d\}}\right] \\
&\quad = e^{\alpha(\theta)u} E_{0,1}\left[e^{-\theta T_u} h_{J_{T_u}}(\alpha(\theta))\right] + e^{\alpha(\theta)d} E_{0,1}\left[e^{-\theta T_d} h_{J_{T_d}}(\alpha(\theta))\right]
\end{aligned}$$

and

$$\begin{aligned}
E_{0,1}\left[e^{-\theta T_u} h_{J_{T_u}}(\alpha(\theta))\right] &= E_{0,1}\left[e^{-\theta T_u} h_{J_{T_u}}(\alpha(\theta)) I_{\{J_{T_u}=1\}}\right] + E_{0,1}\left[e^{-\theta T_u} h_{J_{T_u}}(\alpha(\theta)) I_{\{J_{T_u}=2\}}\right] \\
&= -\lambda E_{0,1}\left[e^{-\theta T_u} I_{\{J_{T_u}=1\}}\right] + [\varphi_1(\alpha(\theta)) - \lambda - \theta] E_{0,1}\left[e^{-\theta T_u} I_{\{J_{T_u}=2\}}\right] \\
&= -\lambda L^u_{1,1} + [\varphi_1(\alpha(\theta)) - \lambda - \theta] L^u_{1,2}, \\
E_{0,1}\left[e^{-\theta T_d} h_{J_{T_d}}(\alpha(\theta))\right] &= -\lambda E_{0,1}\left[e^{-\theta T_d} I_{\{J_{T_d}=1\}}\right] + [\varphi_1(\alpha(\theta)) - \lambda - \theta] E_{0,1}\left[e^{-\theta T_d} I_{\{J_{T_d}=2\}}\right] \\
&= -\lambda L^d_{1,1} + [\varphi_1(\alpha(\theta)) - \lambda - \theta] L^d_{1,2}.
\end{aligned}$$

Hence (6.2.7) becomes

$$-\lambda e^{\alpha(\theta)u} L^u_{1,1} + e^{\alpha(\theta)u}[\varphi_1(\alpha(\theta)) - \lambda - \theta] L^u_{1,2} - \lambda e^{\alpha(\theta)d} L^d_{1,1} + e^{\alpha(\theta)d}[\varphi_1(\alpha(\theta)) - \lambda - \theta] L^d_{1,2} = h_1(\alpha(\theta)). \tag{6.2.8}$$

Recall that for a given $\theta > 0$, we can find four $\alpha_k(\theta), k \in \{1, 2, 3, 4\}$ satisfying (6.2.8), so we have four equations. And we also have four unknowns: $L^u_{1,1}, L^u_{1,1}, L^d_{1,1}, L^d_{1,2}$. So in principle, $L^u_{1,1}, L^u_{1,1}, L^d_{1,1}, L^d_{1,2}$ can be obtained. (It turns out that, in most common cases, $\alpha_k(\theta), k \in \{1,2,3,4\}$ are real and distinct). To be specific, let

$$A = \begin{bmatrix} -\lambda e^{\alpha_1(\theta)u} & e^{\alpha_1(\theta)u}[\mu_1\alpha_1(\theta) + \frac{1}{2}\sigma_1^2\alpha_1(\theta)^2 - \lambda - \theta] & -\lambda e^{\alpha_1(\theta)d} & e^{\alpha_1(\theta)d}[\mu_1\alpha_1(\theta) + \frac{1}{2}\sigma_1^2\alpha_1(\theta)^2 - \lambda - \theta] \\ -\lambda e^{\alpha_2(\theta)u} & e^{\alpha_2(\theta)u}[\mu_1\alpha_2(\theta) + \frac{1}{2}\sigma_1^2\alpha_2(\theta)^2 - \lambda - \theta] & -\lambda e^{\alpha_2(\theta)d} & e^{\alpha_2(\theta)d}[\mu_1\alpha_2(\theta) + \frac{1}{2}\sigma_1^2\alpha_2(\theta)^2 - \lambda - \theta] \\ -\lambda e^{\alpha_3(\theta)u} & e^{\alpha_3(\theta)u}[\mu_1\alpha_3(\theta) + \frac{1}{2}\sigma_1^2\alpha_3(\theta)^2 - \lambda - \theta] & -\lambda e^{\alpha_3(\theta)d} & e^{\alpha_3(\theta)d}[\mu_1\alpha_1(\theta) + \frac{1}{2}\sigma_1^2\alpha_3(\theta)^2 - \lambda - \theta] \\ -\lambda e^{\alpha_4(\theta)u} & e^{\alpha_4(\theta)u}[\mu_1\alpha_4(\theta) + \frac{1}{2}\sigma_1^2\alpha_4(\theta)^2 - \lambda - \theta] & -\lambda e^{\alpha_4(\theta)d} & e^{\alpha_4(\theta)d}[\mu_1\alpha_4(\theta) + \frac{1}{2}\sigma_1^2\alpha_4(\theta)^2 - \lambda - \theta] \end{bmatrix}$$

and

$$B_1 = \begin{bmatrix} -\lambda & -\lambda & -\lambda & -\lambda \end{bmatrix}'.$$

Then

$$\begin{bmatrix} L^u_{1,1} & L^u_{1,2} & L^d_{1,1} & L^d_{1,2} \end{bmatrix}' = A^{-1}B_1.$$

Similarly,

$$\begin{bmatrix} L^u_{2,1} & L^u_{2,2} & L^d_{2,1} & L^d_{2,2} \end{bmatrix}' = A^{-1}B_2,$$

with $B_2 = \begin{bmatrix} \mu_1\alpha_1(\theta) + \frac{1}{2}\sigma_1^2\alpha_1(\theta)^2 - \lambda - \theta \\ \mu_1\alpha_2(\theta) + \frac{1}{2}\sigma_1^2\alpha_2(\theta)^2 - \lambda - \theta \\ \mu_1\alpha_3(\theta) + \frac{1}{2}\sigma_1^2\alpha_3(\theta)^2 - \lambda - \theta \\ \mu_1\alpha_4(\theta) + \frac{1}{2}\sigma_1^2\alpha_4(\theta)^2 - \lambda - \theta \end{bmatrix}.$

Having obtained the state-dependent Laplace transform of T_u and T_d, we are now concerned with inverting them to get the state-dependent defective densities

$$f^u_{i,j}(t) = P_{0,i}(T_u \in dt, J_{T_u} = j), f^d_{i,j}(t) = P_{0,i}(T_d \in dt, J_{T_u} = j),$$

where $i, j \in \{1,2\}$.

As a matter of fact, the expressions for $L^u_{i,j}, i, j \in \{1,2\}$ are too complex to allow a direct analytic inversion. However, we are able find an analytic approximation for the state-dependent Laplace transform by sum of exponential functions.

To begin with, consider a smooth function $f(x)$ on $[0,1]$. According to Beylkin and

Monzon (2005), we define the continuous approximation problem as follow: Given the accuracy $\epsilon > 0$, find the (nearly) minimal number of complex weights ω_m and complex nodes $e^{t_m x}$ s.t.

$$\left| f(x) - \sum_{m=1}^{M} \omega_m e^{t_m x} \right| \leq \epsilon, \quad \forall x \in [0,1].$$

The first step in solving this continuous problem is to reformulate it as a discrete one. Namely, given $2N+1$ values of $f(x)$ on a uniform grid in $[0,1]$ and a target accuracy $\epsilon > 0$, find the (nearly) minimal number of complex weights ω_m and complex nodes γ_m s.t.

$$\left| f(\frac{k}{2N}) - \sum_{m=1}^{M} \omega_m \gamma_m^k \right| \leq \epsilon, \quad \forall 0 \leq k \leq 2N.$$

Once we obtain ω_m and γ_m, the solution to the continuous problem is $\sum_{m=1}^{M} \omega_m e^{t_m x}, \quad t_m = 2N \log(\gamma_m)$.

Section 4 of Beylkin and Monzon (2005) provides an algorithm for the computation of ω_m and γ_m

1. For a properly chosen N (we should slightly oversample $f(x)$ to guarantee that the function can be reconstructed from its samples), let $h(k) = f(\frac{k}{2N})$ and use this sample to construct the Hankel matrix H with $H_{k,l} = h(k+l)$;

2. Find the smallest eigenvalue of H that is bigger than the target accuracy $\epsilon > 0$ and denote the corresponding eigenvector by $u = [u_0, \cdots, u_N]$;

3. Find M roots of the so called c-eigenpolynomial $\sum_{k=0}^{N} u_k z^k$ in the significant region and denote the roots by $\gamma_1, \cdots, \gamma_M$. (Beylkin and Monzon (2005) does not provide clear guidance on how to choose the M roots, though. For our application, we just select those roots which are real or nearly real. The nearly real number refers to a complex number with very small imaginary part, which, we believe, is zero if not for the numerical error. It turns out that this criterion almost always select exactly M roots.)

4. Obtain the M weights corresponding to $\gamma_1, \cdots, \gamma_M$ by solving a least square problem

$$\text{minimize} \sum_{k=0}^{2N} \left(h_k - \sum_{m=1}^{M} \omega_m \gamma_m^k \right)^2 .$$

To approximation a function $f(x)$ on $[0, c]$, we can define a new function on $[0, 1]$ by $g(x) = f(cx)$ and find the sum of exponential approximation to g, say $\sum_{m=1}^{M} \omega_m e^{t_m x}$. Then the approximation to f is $\sum_{m=1}^{M} \omega_m e^{t_m \frac{x}{c}}$.

In Figure 6.2.1, we apply the algorithm to the 8 state-dependent Laplace transforms. The results suggest a very good approximation, as the two lines in each plot are almost indistinguishable.

Once we get the nodes t and weight ω, we can invert the sum of exponential functions to find an approximation to the state-dependent densities. For example, consider $f_{1,1}^u(t) = P_{0,1}(T_u \in dt, J_{T_u} = 1)$, because we know $L_{1,1}^u = E_{0,1}[e^{-\theta T_u} I_{\{J_{T_u}=1\}}] = \sum_{m=1}^{M} \omega_m e^{t_m x}$, the continuous density $f_{1,1}^u(t)$ can be estimated by a discrete probability distribution: $P_{0,1}(T_u = -t_m, J_{T_u} = 1) = \omega_m, 1 \leq m \leq M$. (In principle, T_u and T_d should only take positive value. But there are positive t's in Table 6.2.1 and Table 6.2.2. We remark that this does not cause any trouble, for those positive t's are always associated with extremely small ω's, which ,we believe, should in fact be 0. So we can drop the positive t's.)

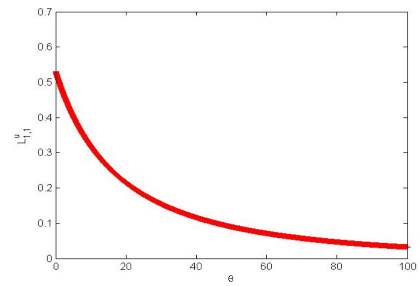
We will see later on that the discrete approximation to the continuous densities of the two-sided hitting times will reduce the burden of numerical computation of the expected cost significantly.

	t	ω
$L^u_{1,1}$	1.80852501856910	5.64943548810358e-10
	-1.06059605301666	0.0540292652465996
	-2.52907603949538	0.160085744041042
	-5.23319421181019	0.165102716812368
	-9.57317511046110	0.103377159102284
	-16.1619048107431	0.0397036897486903
	-26.2536426423025	0.00761835243509458
	-42.9604678809146	0.000421848827119129
$L^u_{1,2}$	-1.14784601679815	0.00253328594487044
	-3.04871571904201	0.00886794245087835
	-6.41798503951179	0.0108794512941907
	-11.8427914708896	0.00659607296831040
	-20.5712260042625	0.00183889125673382
	-35.6930763520289	0.000148511141090607
$L^d_{1,1}$	1.80852756817127	4.39924500978415e-10
	-1.06059905792554	0.0420786354962498
	-2.52908758423329	0.124677326691885
	-5.23321977964831	0.128586572262751
	-9.57320954563106	0.0805143140683346
	-16.1619407768352	0.0309231009106943
	-26.2536792193954	0.00593354332153920
	-42.9605076940982	0.000328556232407684
$L^d_{1,2}$	-1.14803445863434	0.00210134663721014
	-3.0518626504514	0.00738197397318248
	-6.42264322541062	0.00908930154130968
	-11.8477599424369	0.00551891627013721
	-20.5763897315360	0.00153915581156443
	-35.6987833276554	0.000124296504110904

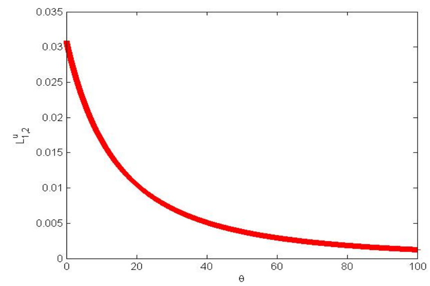
Table 6.2.1: The Nodes and weights. The common parameters are $\mu_1 = 0.1, \sigma_1 = 0.2, \mu_2 = 0.15, \sigma_2 = 0.3, \lambda = 1, v = 2, u = 0.05, d = -0.05$.

	t	ω
$L^u_{2,1}$	-1.15095811756262	0.00228863096892593
	-3.05700094828703	0.00806376972204257
	-6.43003774970880	0.00995760718492565
	-11.8562348816443	0.00605756637161561
	-20.5854249365050	0.00169031715922332
	-35.7088754004830	0.000136485367669418
$L^u_{2,2}$	-0.655026489777474	0.119847868824214
	-1.82964119366378	0.218616045751207
	-4.09323727933916	0.133155914839690
	-7.80531230318662	0.0383435399620687
	-13.9136117031091	0.00418901372426000
	-27.1393227256974	9.22969818685364e-05
$L^d_{2,1}$	-1.14492512868536	0.00183782068557754
	-3.04357722313321	0.00641435596222969
	-6.41057159920028	0.00784645262186679
	-11.8342822972467	0.00474812549018077
	-20.5621582137059	0.00132295942677462
	-35.6829543591211	0.000106857961816101
$L^d_{2,2}$	-0.654934164699926	0.101417891624054
	-1.82923376178448	0.185032527533187
	-4.09207626118483	0.112728645771948
	-7.80242327879989	0.0324762801803375
	-13.9054701524608	0.00355090584188897
	-27.1110276498739	7.80911940697021e-05

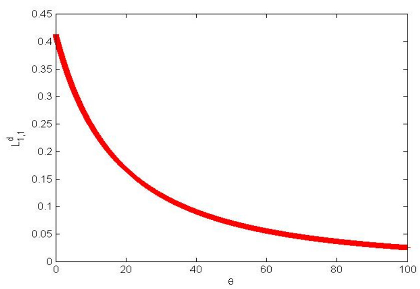
Table 6.2.2: The Nodes and weights (Continued). The common parameters are $\mu_1 = 0.1, \sigma_1 = 0.2, \mu_2 = 0.15, \sigma_2 = 0.3, \lambda = 1, v = 2, u = 0.05, d = -0.05$.



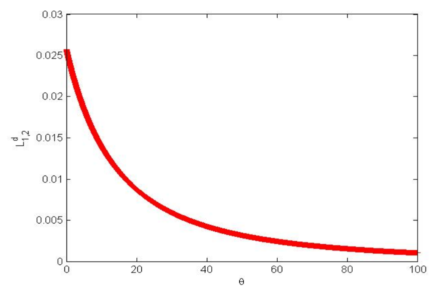
$L^u_{1,1}$



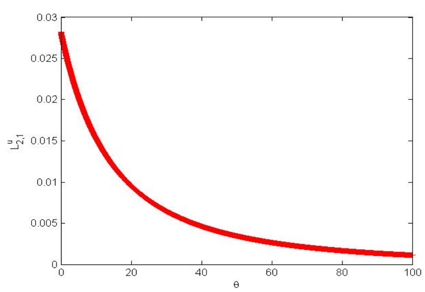
$L^u_{1,2}$



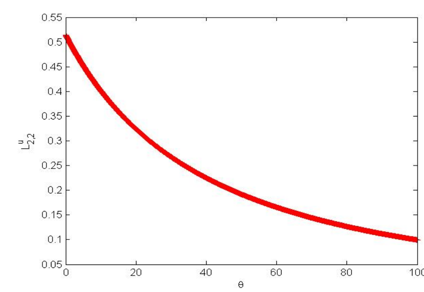
$L^d_{1,1}$



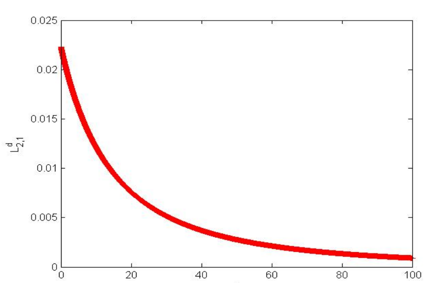
$L^d_{1,2}$



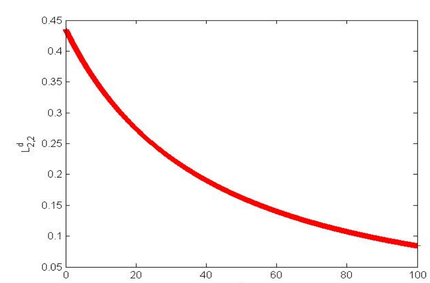
$L^u_{2,1}$



$L^u_{2,2}$



$L^d_{2,1}$



$L^d_{2,2}$

Figure 6.2.1: Sum of exponential approximation to the state-dependent Laplace transform. The black solid lines are the state-dependent Laplace transforms and the red plus sign lines are the sum of exponential approximations. The common parameters are $\mu_1 = 0.1, \sigma_1 = 0.2, \mu_2 = 0.15, \sigma_2 = 0.3, \lambda = 1, v = 2, u = 0.05, d = -0.05$.

6.3 The Expected Cost

In this section, we derive a new recursive formula for computing the expected cost of discretely hedging a put option with the underlying asset following a GBM model with two regimes.

Consider the first cost

$$e^{-r(\tau_1 \wedge \epsilon_1^\lambda)} \left[P^R_{\tau_1 \wedge \epsilon_1^\lambda} - \Delta^R_0 e^{d(\tau_1 \wedge \epsilon_1^\lambda)} S_{\tau_1 \wedge \epsilon_1^\lambda} - M^R_0 e^{r(\tau_1 \wedge \epsilon_1^\lambda)} \right], \tag{6.3.1}$$

where $\tau_1 \stackrel{d}{=} T_{u,d}$, ϵ_1^λ is an independent exponential r.v. with mean $\frac{1}{\lambda}$, P^R_t, Δ^R_t and M^R_t are the time t price, Delta and value of the money market account of a put option written on an underlying asset whose dynamics follows a GBM with two regimes.

The expectation of (6.3.1) is denoted by $h^R(S_0, J_0)$. We will show how to compute $h^R(S_0, J_0)$ using numerical integration in section 6.4.

Now assume $h^R(S_0, J_0)$ is known, we move on to the second cost

$$I_{\{\tau_1 < \epsilon_1^\lambda\}} e^{-r(\tau_1 + \tau_2 \wedge \epsilon_2^\lambda)} \left[P^R_{\tau_1 + \tau_2 \wedge \epsilon_2^\lambda} - \Delta^R_{\tau_1} e^{d(\tau_2 \wedge \epsilon_2^\lambda)} S_{\tau_1 + \tau_2 \wedge \epsilon_2^\lambda} - M^R_{\tau_1} e^{r(\tau_2 \wedge \epsilon_2^\lambda)} \right], \tag{6.3.2}$$

where $\tau_2 \stackrel{d}{=} T_{u,d}$ is conditionally independent of τ_1 and $\epsilon_2^\lambda = \epsilon_1^\lambda - \tau_1$.

Conditioning on τ_1 and $\tau_1 < \epsilon_1^\lambda$, the conditional expectation is

$$\begin{aligned} & e^{-(r+\lambda)\tau_1} E_{\tau_1} \left(e^{-r(\tau_2 \wedge \epsilon_2^\lambda)} \left[P^R_{\tau_1 + \tau_2 \wedge \epsilon_2^\lambda} - \Delta^R_{\tau_1} e^{d(\tau_2 \wedge \epsilon_2^\lambda)} S_{\tau_1 + \tau_2 \wedge \epsilon_2^\lambda} - M^R_{\tau_1} e^{r(\tau_2 \wedge \epsilon_2^\lambda)} \right] | \epsilon_1^\lambda > \tau_1 \right) \\ & = e^{-(r+\lambda)\tau_1} h^R(S_{\tau_1}, J_{\tau_1}). \end{aligned}$$

In general, the conditional expectation for the $(n+1)$-th cost is

$$e^{-(r+\lambda)(\tau_1 + \cdots + \tau_n)} h^R(S_{\tau_1 + \cdots + \tau_n}, J_{\tau_1 + \cdots + \tau_n}). \tag{6.3.3}$$

For the computation of the unconditional expectation, we take $n = 2$ as an example

$$
\begin{aligned}
&E_{0,1}\left[e^{-(r+\lambda)(\tau_1+\tau_2)}h^R(S_{\tau_1+\tau_2}, J_{\tau_1+\tau_2})\right] \\
&= E_{0,1}\left[e^{-(r+\lambda)(\tau_1+\tau_2)}h^R(S_{\tau_1+\tau_2}, J_{\tau_1+\tau_2})\right] \\
&= \sum_{s=S_0e^{2\alpha},S_0,S_0e^{-2\alpha}} \sum_{j=1,2} E_{0,1}\left[e^{-(r+\lambda)(\tau_1+\tau_2)}h^R(S_{\tau_1+\tau_2}, J_{\tau_1+\tau_2})I_{\{S_{\tau_1+\tau_2}=s, J_{\tau_1+\tau_2}=j\}}\right]. \quad (6.3.4)
\end{aligned}
$$

Each term in the sum of (6.3.4) can be further simplified in similar ways. For instance,

$$
\begin{aligned}
&E_{0,1}\left[e^{-(r+\lambda)(\tau_1+\tau_2)}h^R(S_{\tau_1+\tau_2}, J_{\tau_1+\tau_2})I_{\{S_{\tau_1+\tau_2}=S_0e^{2\alpha}, J_{\tau_1+\tau_2}=1\}}\right] \\
&= h^R(S_0e^{2\alpha}, 1)E_{0,1}\left[e^{-(r+\lambda)(\tau_1+\tau_2)}I_{\{S_{\tau_1+\tau_2}=S_0e^{2\alpha}, J_{\tau_1+\tau_2}=1\}}\right],
\end{aligned}
$$

and

$$
\begin{aligned}
&E_{0,1}\left[e^{-(r+\lambda)(\tau_1+\tau_2)}I_{\{S_{\tau_1+\tau_2}=S_0e^{-2\alpha}, J_{\tau_1+\tau_2}=1\}}\right] \\
&= E_{0,1}\left[e^{-(r+\lambda)(\tau_\alpha+\tilde{\tau}_\alpha)}I_{\{\tau_1=\tau_\alpha, \tau_2=\tilde{\tau}_\alpha\}}I_{\{J_{\tau_\alpha+\tilde{\tau}_\alpha}=1\}}\right] \\
&= E_{0,1}\left[e^{-(r+\lambda)(\tau_\alpha+\tilde{\tau}_\alpha)}I_{\{\tau_1=\tau_\alpha, \tau_2=\tilde{\tau}_\alpha\}}I_{\{J_{\tau_\alpha}=1\}}I_{\{J_{\tau_\alpha+\tilde{\tau}_\alpha}=1\}}\right] \\
&\quad +E_{0,1}\left[e^{-(r+\lambda)(\tau_\alpha+\tilde{\tau}_\alpha)}I_{\{\tau_1=\tau_\alpha, \tau_2=\tilde{\tau}_\alpha\}}I_{\{J_{\tau_\alpha}=2\}}I_{\{J_{\tau_\alpha+\tilde{\tau}_\alpha}=1\}}\right]
\end{aligned}
\quad (6.3.5)
$$

Using the identity $E(XI_A) = E(X|A)P(A)$ for $A = \{J_{\tau_\alpha} = 1\}$ and $A = \{J_{\tau_\alpha} = 2\}$, we have

$$
\begin{aligned}
&E_{0,1}\left[e^{-(r+\lambda)(\tau_1+\tau_2)}I_{\{S_{\tau_1+\tau_2}=S_0e^{-2\alpha}, J_{\tau_1+\tau_2}=1\}}\right] \\
&= E_{0,1}\left[e^{-(r+\lambda)(\tau_\alpha+\tilde{\tau}_\alpha)}I_{\{\tau_1=\tau_\alpha, \tau_2=\tilde{\tau}_\alpha\}}I_{\{J_{\tau_\alpha+\tilde{\tau}_\alpha}=1\}}|J_{\tau_\alpha}=1\right]P(J_{\tau_\alpha}=1) \\
&\quad + E_{0,1}\left[e^{-(r+\lambda)(\tau_\alpha+\tilde{\tau}_\alpha)}I_{\{\tau_1=\tau_\alpha, \tau_2=\tilde{\tau}_\alpha\}}I_{\{J_{\tau_\alpha+\tilde{\tau}_\alpha}=1\}}|J_{\tau_\alpha}=2\right]P(J_{\tau_\alpha}=2) \quad (6.3.6)
\end{aligned}
$$

For $E_{0,1}\left[e^{-(r+\lambda)(\tau_\alpha+\tilde{\tau}_\alpha)}I_{\{\tau_1=\tau_\alpha,\tau_2=\tilde{\tau}_\alpha\}}I_{\{J_{\tau_\alpha+\tilde{\tau}_\alpha}=1\}}|J_{\tau_\alpha}=1\right]$, we further condition on τ_1 and the conditional expectation is

$$
\begin{aligned}
& e^{-(r+\lambda)\tau_\alpha}I_{\{\tau_1=\tau_\alpha\}}|J_{\tau_\alpha}=1 \quad E_{\tau_\alpha}\left[e^{-(r+\lambda)\tilde{\tau}_\alpha}I_{\{\tau_2=\tilde{\tau}_\alpha\}}I_{\{J_{\tau_\alpha+\tilde{\tau}_\alpha}=1\}}|J_{\tau_\alpha}=1\right] \\
&= e^{-(r+\lambda)\tau_\alpha}I_{\{\tau_1=\tau_\alpha\}}|J_{\tau_\alpha}=1 \quad E_{\tau_\alpha}\left[e^{-(r+\lambda)\tilde{\tau}_\alpha}I_{\{J_{\tau_\alpha+\tilde{\tau}_\alpha}=1\}}|J_{\tau_\alpha}=1\right] \\
&= e^{-(r+\lambda)\tau_\alpha}I_{\{\tau_1=\tau_\alpha\}}|J_{\tau_\alpha}=1 \quad L^\alpha_{1,1}(r+\lambda),
\end{aligned}
$$

where $L^\alpha_{1,1}(r+\lambda)=E_{0,1}\left[e^{-(r+\lambda)T_\alpha}I_{\{J_{T_\alpha}=1\}}\right]$.

So the unconditional expectation is

$$
L^\alpha_{1,1}(r+\lambda)E_{0,1}\left[e^{-(r+\lambda)\tau_\alpha}I_{\{\tau_1=\tau_\alpha\}}|J_{\tau_\alpha}=1\right]
$$

and the first term on the RHS of (6.3.6) is

$$
\begin{aligned}
& L^\alpha_{1,1}(r+\lambda)E_{0,1}\left[e^{-(r+\lambda)\tau_\alpha}I_{\{\tau_1=\tau_\alpha\}}|J_{\tau_\alpha}=1\right]P(J_{\tau_\alpha}=1) \\
&= L^\alpha_{1,1}(r+\lambda)E_{0,1}\left[e^{-(r+\lambda)\tau_\alpha}I_{\{\tau_1=\tau_\alpha\}}I_{\{J_{\tau_\alpha}=1\}}\right] \\
&= L^\alpha_{1,1}(r+\lambda)E_{0,1}\left[e^{-(r+\lambda)\tau_\alpha}I_{\{J_{\tau_\alpha}=1\}}\right] \\
&= L^\alpha_{1,1}(r+\lambda)L^\alpha_{1,1}(r+\lambda)
\end{aligned}
$$

Analogously, the second term is $L^\alpha_{1,2}(r+\lambda)L^\alpha_{2,1}(r+\lambda)$, so

$$
\begin{aligned}
& E_{0,1}\left[e^{-(r+\lambda)(\tau_1+\tau_2)}h^R(S_{\tau_1+\tau_2},J_{\tau_1+\tau_2})I_{\{S_{\tau_1+\tau_2}=S_0e^{2\alpha},J_{\tau_1+\tau_2}=1\}}\right] \\
&= h^R(S_0e^{2\alpha},1)[L^\alpha_{1,1}(r+\lambda)L^\alpha_{1,1}(r+\lambda)+L^\alpha_{1,2}(r+\lambda)L^\alpha_{2,1}(r+\lambda)],
\end{aligned}
$$

where for compactness, we write $L^{\pm\alpha}_{i,j}(r+\lambda)=L^{\pm\alpha}_{i,j}, \quad i,j\in\{1,2\}$.

Applying the same operations to other term in the sum of (6.3.4), we get

$$
\begin{aligned}
&E_{0,1}\left[e^{-(r+\lambda)(\tau_1+\tau_2)}h^R(S_{\tau_1+\tau_2},J_{\tau_1+\tau_2})\right]\\
&\quad= h^R(S_0e^{2\alpha},1)[L^{\alpha}_{1,1}L^{\alpha}_{1,1}+L^{\alpha}_{1,2}L^{\alpha}_{2,1}]+h^R(S_0e^{2\alpha},2)[L^{\alpha}_{1,1}L^{\alpha}_{1,2}+L^{\alpha}_{1,2}L^{\alpha}_{2,2}]\\
&\qquad+h^R(S_0,1)[L^{\alpha}_{1,1}L^{-\alpha}_{1,1}+L^{\alpha}_{1,2}L^{-\alpha}_{2,1}+L^{-\alpha}_{1,1}L^{\alpha}_{1,1}+L^{-\alpha}_{1,2}L^{\alpha}_{2,1}]\\
&\qquad+h^R(S_0,2)[L^{\alpha}_{1,1}L^{-\alpha}_{1,2}+L^{\alpha}_{1,2}L^{-\alpha}_{2,2}+L^{-\alpha}_{1,1}L^{\alpha}_{1,2}+L^{-\alpha}_{1,2}L^{\alpha}_{2,2}]\\
&\qquad+h^R(S_0e^{-2\alpha},1)[L^{-\alpha}_{1,1}L^{-\alpha}_{1,2}+L^{-\alpha}_{1,2}L^{-\alpha}_{2,2}]+h^R(S_0e^{-2\alpha},2)[L^{-\alpha}_{1,1}L^{-\alpha}_{1,2}+L^{-\alpha}_{1,2}L^{-\alpha}_{2,2}].
\end{aligned}
$$

For the unconditional expectation of the $(n+1)$-th cost $E_{0,i}\left[e^{-(r+\lambda)(\tau_1+\cdots+\tau_n)}h^R(S_{\tau_1+\cdots+\tau_n},J_{\tau_1+\cdots+\tau_n})\right]$, we need to define

$$
\mathbf{L}_u=\begin{bmatrix} L^u_{1,1} & L^u_{1,2}\\ L^u_{2,1} & L^u_{2,2}\end{bmatrix},\quad \mathbf{L}_d=\begin{bmatrix} L^d_{1,1} & L^d_{1,2}\\ L^d_{2,1} & L^d_{2,2}\end{bmatrix}
$$

and then

$$
\begin{aligned}
&E_{0,i}\left[e^{-(r+\lambda)(\tau_1+\cdots+\tau_n)}h^R(S_{\tau_1+\cdots+\tau_n},J_{\tau_1+\cdots+\tau_n})\right]\\
&=\sum_{k=0}^{n}\sum_{j=1,2}h^R(S_0e^{(2k-n)\alpha},j)\left[\sum_{l_1,\cdots,l_n=0,1;l_1+\cdots+l_n=k}(\mathbf{L}_u^{l_1}\mathbf{L}_d^{1-l_1})\cdots(\mathbf{L}_u^{l_n}\mathbf{L}_d^{1-l_n})\right]_{i,j}.
\end{aligned}
\tag{6.3.7}
$$

The efficient computation of $f(n,k)=\left[\sum_{l_1,\cdots,l_n=0,1;l_1+\cdots+l_n=k}(\mathbf{L}_u^{l_1}\mathbf{L}_d^{1-l_1})\cdots(\mathbf{L}_u^{l_n}\mathbf{L}_d^{1-l_n})\right]$ makes use of the following recursion.

$$
\begin{aligned}
&\sum_{\substack{l_1,\cdots,l_n,l_{n+1}=0,1;\\ l_1+\cdots+l_n+l_{n+1}=k}}(\mathbf{L}_u^{l_1}\mathbf{L}_d^{1-l_1})\cdots(\mathbf{L}_u^{l_n}\mathbf{L}_d^{1-l_n})(\mathbf{L}_u^{l_{n+1}}\mathbf{L}_d^{1-l_{n+1}})\\
&=\sum_{\substack{l_1,\cdots,l_n,l_{n+1}=0,1;\\ l_1+\cdots+l_n=k,l_{n+1}=0}}(\mathbf{L}_u^{l_1}\mathbf{L}_d^{1-l_1})\cdots(\mathbf{L}_u^{l_n}\mathbf{L}_d^{1-l_n})\mathbf{L}_d+\sum_{\substack{l_1,\cdots,l_n,l_{n+1}=0,1;\\ l_1+\cdots+l_n=k-1,l_{n+1}=1}}(\mathbf{L}_u^{l_1}\mathbf{L}_d^{1-l_1})\cdots(\mathbf{L}_u^{l_n}\mathbf{L}_d^{1-l_n})\mathbf{L}_u,
\end{aligned}
$$

and hence

$$
f(n+1,k)=f(n,k)\mathbf{L}_d+f(n,k-1)\mathbf{L}_u,\quad 1\le k<n+1.
$$

	$\lambda = 0.5$ $v = 1$	$\lambda = 0.5$ $v = 2$	$\lambda = 1$ $v = 0.5$	$\lambda = 1$ $v = 2$	$\lambda = 2$ $v = 0.5$	$\lambda = 2$ $v = 1$
moments-true:						
1st	0.0246	0.0209	0.0311	0.0222	0.0258	0.0245
2nd	1.5331	1.1092	1.8356	1.2066	1.2879	1.4221
3rd	11.5395	8.9031	16.0664	9.8240	12.8538	15.9128
4th	291.5601	206.3097	440.4688	259.6881	364.0162	460.4080
moments-independence:						
1st	0.0246	0.0209	0.0311	0.0222	0.0258	0.0245
2nd	1.5177	1.0950	1.8148	1.2036	1.2854	1.4174
3rd	10.7765	8.2642	16.1584	9.3789	13.2212	15.9899
4th	280.4126	197.3913	431.2186	251.1223	364.2944	455.1345
moments-semi analytic:						
1st	0.0280	0.0267	0.0346	0.0261	0.0291	0.0275
2nd	1.5057	1.0841	1.7014	1.1691	1.2131	1.4033
3rd	11.2062	8.3594	15.7990	9.6107	12.2099	15.1269
4th	275.3842	165.0003	418.8227	213.6247	334.6223	425.0708

Table 6.3.1: Moments comparison. The row "moments-true" and the row "moments-independence" are obtained by Monte Carlo simulation with 100000 iterations. The row "moments-semi analytic" is obtained using the semi-analytic algorithm. The common parameters are: $T = 3, S_0 = K = 50, r = 0.02, \alpha = 0.05, \mu_1 = 0.1, \mu_2 = 0.15, \sigma_1 = 0.3, \sigma_2 = 0.4, d = 0.03, J_0 = 1$.

For the higher moments, we assume that the individual costs are independent and this again turns out to be reasonable. See Table 6.3.1 for the comparison using the results from Monte Carlo simulation. We then fit the cost distribution with a mixture model of two normals by matching the first 4 raw moments and use the fitted distribution to approximate the quantiles. The fitted densities are plotted in Figure 6.3.1 and the quantiles are listed in Table 6.3.2.

6.4 Expectation of the First Cost

In this section, we compute the function Expectation of the First Cost-$h^R(S_0, J_0)$, using numerical integration.

	$\lambda = 0.5$ $v = 1$	$\lambda = 0.5$ $v = 2$	$\lambda = 1$ $v = 0.5$	$\lambda = 1$ $v = 2$	$\lambda = 2$ $v = 0.5$	$\lambda = 2$ $v = 1$
quantiles-true:						
90%	1.4271	1.1152	1.3103	1.1418	0.9825	1.0513
95%	1.8606	1.5466	1.6185	1.5240	1.2448	1.3534
97.5%	2.2401	1.9354	1.8913	1.8756	1.4969	1.6340
99%	2.6982	2.4048	2.2525	2.3400	1.8188	1.9881
quantiles-semi analytic:						
90%	1.4245	1.3256	1.3334	1.2599	1.0461	1.0119
95%	1.8345	1.7036	1.7245	1.6221	1.3511	1.3116
97.5%	2.2043	2.0401	2.0862	1.9473	1.6225	1.5891
99%	2.6839	2.4606	2.5961	2.3626	1.9602	1.9844

Table 6.3.2: Quantiles comparison. The true quantiles are obtained by Monte Carlo simulation with 100000 iterations. The common parameters are: $T = 3, S_0 = K = 50, r = 0.02, \alpha = 0.05, \mu_1 = 0.1, \mu_2 = 0.15, \sigma_1 = 0.3, \sigma_2 = 0.4, d = 0.03, J_0 = 1$.

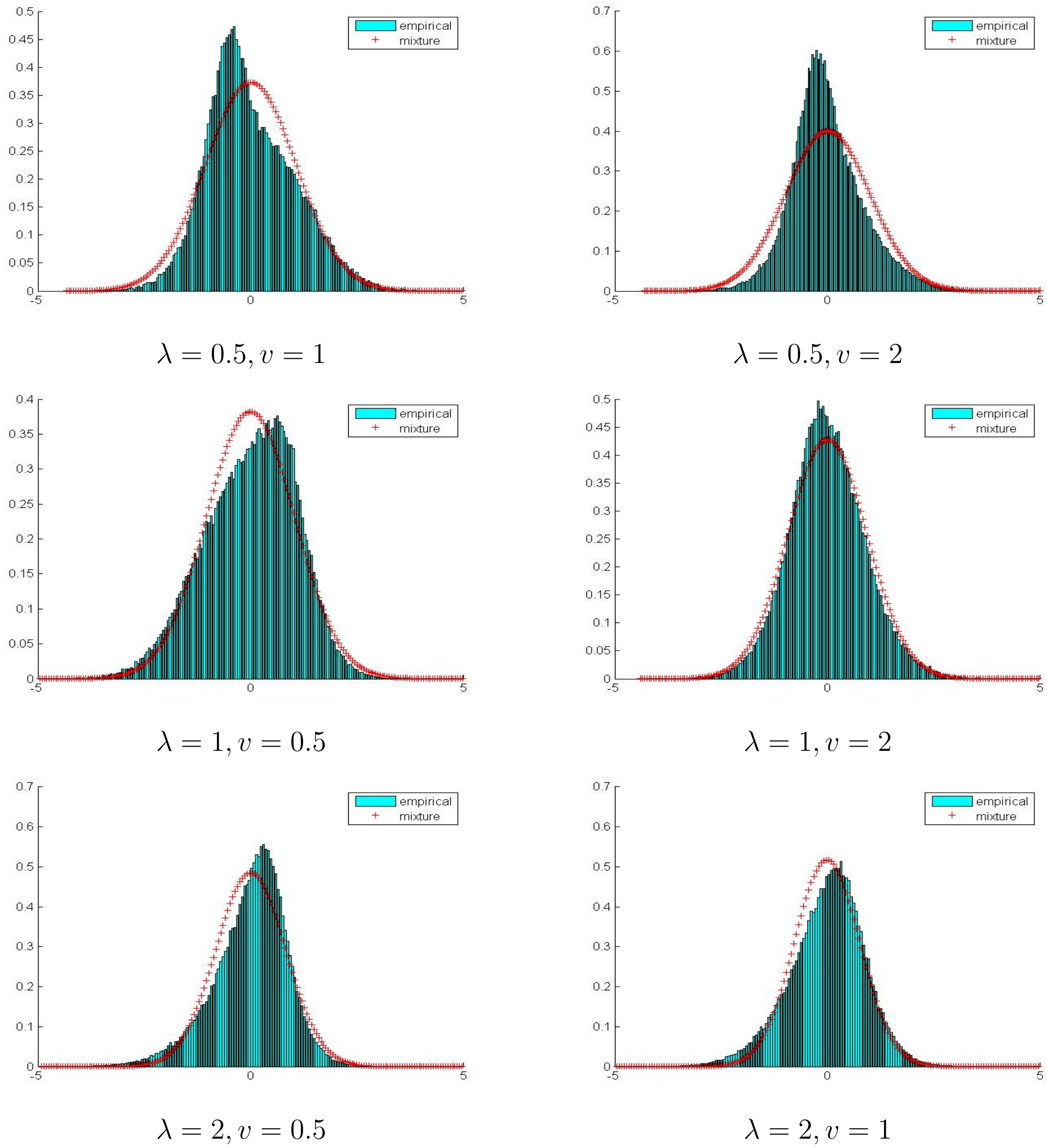
Recall that the first cost is given by

$$\begin{aligned} & e^{-r(\tau_1 \wedge \epsilon_1^\lambda)} \left[P^R_{\tau_1 \wedge \epsilon_1^\lambda} - \Delta^R_0 e^{d(\tau_1 \wedge \epsilon_1^\lambda)} S_{\tau_1 \wedge \epsilon_1^\lambda} - M^R_0 e^{r(\tau_1 \wedge \epsilon_1^\lambda)} \right] \\ & = I_{\{\tau_1 < \epsilon_1^\lambda\}} e^{-r\tau_1} \left[P^R_{\tau_1} - \Delta^R_0 e^{d\tau_1} S_{\tau_1} - M^R_0 e^{r\tau_1} \right] + I_{\{\tau_1 \geq \epsilon_1^\lambda\}} \left[P^R_{\epsilon_1^\lambda} - \Delta^R_0 e^{d\epsilon_1^\lambda} S_{\epsilon_1^\lambda} - M^R_0 e^{r\epsilon_1^\lambda} \right] \end{aligned} \tag{6.4.8}$$

The second term on the RHS of (6.4.8) is of minor importance, for at least two reasons:

1. When ϵ_1^λ is large and the bandwidth α is small, the sub-account will almost always hit the band before ϵ_1^λ, especially with high volatility or drift. So the second term vanishes in this case.

2. When ϵ_1^λ is small and the sub-account does not hit the band before ϵ_1^λ, the second term-the difference between the value of the option and the hedging portfolio-is nonzero but should be very small. Because neither time (ϵ_1^λ is small) nor the underlying asset price (the sub-account does not hit the band) has changed too much.



$\lambda = 0.5, v = 1$ $\lambda = 0.5, v = 2$

$\lambda = 1, v = 0.5$ $\lambda = 1, v = 2$

$\lambda = 2, v = 0.5$ $\lambda = 2, v = 1$

Figure 6.3.1: Density comparison. The histograms are obtained by Monte Carlo simulation with 100000 iterations. The common parameters are: $T = 3, S_0 = K = 50, r = 0.02, \alpha = 0.05, \mu_1 = 0.1, \mu_2 = 0.15, \sigma_1 = 0.3, \sigma_2 = 0.4, d = 0.03, J_0 = 1$.

Henceforth we ignore the second term and focus on the first of (6.4.8). Table 6.3.1 examines the accuracy of this simplification. The row "moments-semi analytic" uses the semi-analytic algorithm to compute the moments of the total re-balancing cost, which are close to the value obtained by Monte Carlo Simulation.

Denote $e^{-r\tau_1}\left[P^R_{\tau_1} - \Delta^R_0 e^{d\tau_1} S_{\tau_1} - M^R_0 e^{r\tau_1}\right]$ by $q(\epsilon^\lambda_1, \tau_1)$, the expectation of the first term on the RHS of (6.4.8) is

$$E\left[I_{\{\tau_1<\epsilon^\lambda_1\}} q(\epsilon^\lambda_1, \tau_1)\right] = E\left[E_{\epsilon^\lambda_1}\left(I_{\{\tau_1<\epsilon^\lambda_1\}} q(\epsilon^\lambda_1, \tau_1)\right)\right],$$

where the inner expectation can be further decomposed

$$\begin{aligned}
&E_{\epsilon^\lambda_1}\left[I_{\{\tau_1<\epsilon^\lambda_1\}} q(\epsilon^\lambda_1, \tau_1)\right] \\
&\quad = E_{\epsilon^\lambda_1}\left[I_{\{\tau_1<\epsilon^\lambda_1\}} q(\epsilon^\lambda_1, \tau_1) I_{\{\tau_1=\tau_\alpha\}}\right] + E_{\epsilon^\lambda_1}\left[I_{\{\tau_1<\epsilon^\lambda_1\}} q(\epsilon^\lambda_1, \tau_1) I_{\{\tau_1=\tau_{-\alpha}\}}\right] \\
&\quad = E_{\epsilon^\lambda_1}\left[I_{\{\tau_\alpha<\epsilon^\lambda_1\}} q(\epsilon^\lambda_1, \tau_\alpha) I_{\{\tau_1=\tau_\alpha\}}\right] + E_{\epsilon^\lambda_1}\left[I_{\{\tau_{-\alpha}<\epsilon^\lambda_1\}} q(\epsilon^\lambda_1, \tau_{-\alpha}) I_{\{\tau_1=\tau_{-\alpha}\}}\right] \\
&\quad = E_{\epsilon^\lambda_1}\left[I_{\{\tau_\alpha<\epsilon^\lambda_1\}} q(\epsilon^\lambda_1, \tau_\alpha)\right] + E_{\epsilon^\lambda_1}\left[I_{\{\tau_{-\alpha}<\epsilon^\lambda_1\}} q(\epsilon^\lambda_1, \tau_{-\alpha})\right] \\
&\quad = E_{\epsilon^\lambda_1}\left[I_{\{\tau_\alpha<\epsilon^\lambda_1\}} q(\epsilon^\lambda_1, \tau_\alpha) I_{\{J_{\tau_\alpha}=1\}}\right] + E_{\epsilon^\lambda_1}\left[I_{\{\tau_\alpha<\epsilon^\lambda_1\}} q(\epsilon^\lambda_1, \tau_\alpha) I_{\{J_{\tau_\alpha}=2\}}\right] \\
&\qquad + E_{\epsilon^\lambda_1}\left[I_{\{\tau_{-\alpha}<\epsilon^\lambda_1\}} q(\epsilon^\lambda_1, \tau_{-\alpha}) I_{\{J_{\tau_{-\alpha}}=1\}}\right] + E_{\epsilon^\lambda_1}\left[I_{\{\tau_{-\alpha}<\epsilon^\lambda_1\}} q(\epsilon^\lambda_1, \tau_{-\alpha}) I_{\{J_{\tau_{-\alpha}}=2\}}\right]. \qquad (6.4.9)
\end{aligned}$$

In order to compute the expectation terms in (6.4.9), we need expressions for the (state-dependent) option price P^R and Delta Δ^R under the two state regime-switching model. However, it is well known that this model implicitly implies the incompleteness of the underlying financial markets. In other words, there are infinitely many equivalent martingale measures. See Naik (1993) for more details.

Various approaches have been proposed to select an equivalent martingale measure for option pricing in an incomplete market. In essence, these methods choose the (unique) martingale measures that optimize certain criteria, which include minimizing the quadratic utility of the losses caused by incomplete hedging (Follmer and Schweizer (1991), Follmer and Sondermann (1986) and Schweizer (1996)) and an utility optimization problem based on marginal rate of substitution (Davis (1997)).

Gerber and Shiu (1994) pioneered the use of the Esscher transform, a time-honored tool in actuarial science introduced by Esscher (1932), for derivative pricing in incomplete market. It is shown that the martingale measure induced by Esscher transform maximizes the expected power utility. In our analysis, we adopt a particular form of the Esscher transform introduced in Elliott et al. (2005) to determine the equivalent martingale measure for pricing under the regime switching model. The choice of this version of the Esscher transform has been justified by Siu (2008) and Siu (2011) using a saddle point of a stochastic differential game and the minimization of relative entropy, respectively.

Suppose the continuous Markov chain underlying the regime switching model is J_t, the risk free interest rate is r and the dynamic of logarithm return process is

$$X_t = X_0 + \int_0^t \left(\mu_{J_s} - \frac{1}{2}\sigma_{J_s}^2 \right) ds + \int_0^t \sigma_{J_s} dW_s,$$

where $J_t \in \{1, 2\}$, W_t is a standard Brownian motion under the physical probability measure and W_t is independent of ξ_t.

Following Lin et al. (2009), define

$$\theta_t^\star = \frac{\mu_{J_t} - r}{\sigma_{J_t}}$$

and a new probability measure $Q_{\theta^\star}$

$$\frac{dQ_{\theta^\star}}{dP}|_{\mathcal{G}_t} = \Lambda_t^\star = \frac{e^{-\int_0^t \theta_u^\star dW_u}}{E_P\left[e^{-\int_0^t \theta_u^\star dW_u} | \mathcal{F}_t^J\right]},$$

where $\mathcal{F}_t^J$ and $\mathcal{F}_t^W$ are the filtration generated by $\{\xi_t\}$ and $\{W_t\}$ respectively, $\mathcal{G}_t = \mathcal{F}_t^J \vee \mathcal{F}_t^W$ is the minimal σ-algebra containing both $\mathcal{F}_t^J$ and $\mathcal{F}_t^W$.

Since $\Lambda_t^\star$ is $\mathcal{G}_t$ adapted and

$$\frac{d\Lambda_t^\star}{\Lambda_t^\star} = -\theta_t^\star dW_t,$$

From $\frac{d\Lambda_t^\star}{\Lambda_t^\star} = -\theta_t^\star dW_t$, we know that $\{\Lambda_t^\star\}$ is a $(\{\mathcal{G}_t\}, P)$-(local)-martingale. If $\{\theta_t^\star\}$ satisfies the Novikov's condition, $\{\Lambda_t^\star\}$ is a $(\{\mathcal{G}_t\}, P)$-martingale. Then by Girsanov's theorem, $\bar{W}_t = W_t + \theta_t^\star dt, t \geq 0$ is a $(\{\mathcal{G}_t\}, Q_{\theta^\star})$-standard Brownian motion.

The sub-account account value $S_t = e^{X_t}$ has the following dynamics under $Q_{\theta^\star}$

$$\frac{dS_t}{S_t} = rdt + \sigma_{J_t} d\bar{W}_t,$$

and because we do not have any other securities or derivatives in our incomplete market model, we assume the intensity matrix Q of ξ_t is observable/estimable.

The state-dependent price of the put option at any time $t < T$ is given by the $Q_{\theta^\star}$ expectation of its discounted payoff

$$P^R(S_t, J_t) = E_{Q_{\theta^\star},t}\left[e^{-r(T-t)}(K - S_T)_+\right]. \tag{6.4.10}$$

This expectation is computed by first conditioning the path of J from t to T to get

$$E_{Q_{\theta^\star},t}\left[e^{-r(T-t)}(K - S_T)_+ | J_s, t < s \leq T\right] = Ke^{-r(T-t)}N(-d_2) - S_t e^{-d(T-t)}N(-d_1),$$

where

$$\begin{aligned} d_1 &= \frac{\log(\frac{S_t}{K}) + (r-d)(T-t) + \frac{1}{2}\int_t^T \sigma_{\xi_s}^2 ds}{\sqrt{\int_t^T \sigma_{\xi_s}^2 ds}} \\ &= \frac{\log(\frac{S_t}{K}) + (r-d)(T-t) + \frac{1}{2}[\sigma_1^2 Z_1 + \sigma_2^2(T-t-Z_1)]}{\sqrt{\sigma_1^2 Z_1 + \sigma_2^2(T-t-Z_1)}}, \\ d_2 &= d_1 - \sqrt{\sigma_1^2 Z_1 + \sigma_2^2(T-t-Z_1)} \end{aligned}$$

with Z_1 the occupation time of J in state 1 over the interval $[t, T]$.

Then the unconditional expectation is taken w.r.t. Z_1, whose density can be found in

Naik (1993)

$$
\begin{aligned}
& f_1(x, T-t) \\
& \quad = e^{-\lambda_1 x - \lambda_2 (T-t-x)}[\delta_0(T-t-x) \\
& \qquad + (\frac{\lambda_1 \lambda_2 x}{T-t-x})^{\frac{1}{2}} B_1(2[\lambda_1\lambda_2 x(T-t-x)]^{\frac{1}{2}}) + \lambda_1 B_0(2[\lambda_1\lambda_2 x(T-t-x)]^{\frac{1}{2}})] \quad (6.4.11)
\end{aligned}
$$

if $J_t = 1$ and

$$
\begin{aligned}
& f_2(x, T-t) \\
& \quad = e^{-\lambda_1 x - \lambda_2 (T-t-x)}[\delta_0(x) \\
& \qquad + (\frac{\lambda_1 \lambda_2 (T-t-x)}{x})^{\frac{1}{2}} B_1(2[\lambda_1\lambda_2 x(T-t-x)]^{\frac{1}{2}}) + \lambda_2 B_0(2[\lambda_1\lambda_2 x(T-t-x)]^{\frac{1}{2}})]
\end{aligned}
$$
(6.4.12)

if $J_t = 2$, where the intensity matrix of J_t is $Q = \begin{bmatrix} -\lambda_1 & \lambda_1 \\ \lambda_2 & -\lambda_2 \end{bmatrix}$ under $Q_{\theta^\star}$, δ_0 is the Dirac's delta funtion and $B_p(x)$ is the modified Bessel function

$$B_p(x) = \sum_{k=0}^{\infty} \frac{(x/2)^{2k+p}}{k!(k+p)!}.$$

So

$$P^R(S_t, J_t) = E_{Z_1}\left[Ke^{-r(T-t)}N(-d_2) - S_t e^{-d(T-t)}N(-d_1)\right],$$

and

$$
\begin{aligned}
\Delta^R(S_t, J_t) &= \frac{\partial P^R(S_t, J_t)}{\partial S_t} \\
&= E_{Z_1}\left[\frac{\partial}{\partial S_t}\left(Ke^{-r(T-t)}N(-d_2) - S_t e^{-d(T-t)}N(-d_1)\right)\right] \\
&= E_{Z_1}\left[-e^{-d(T-t)}N(-d_1)\right].
\end{aligned}
$$

At this point, we remark that the increased computational complexity due to the absence of closed form expressions for the price and Delta of the put option is offset in part by the use of a discrete distribution as an approximation to the continuous, state-dependent density of the hitting time. Indeed, the integral w.r.t. the hitting time is now substantially reduced to a simple sum with just a few terms.

6.5 The Management Fee

In this section, we address some technical issues related to the application of the "Percentile Premium Principle" to calculate the fair value of the management fee under the regime switching GBM model.

As usual, the cost of continuous hedging is given by the price of the put option $P^R(S_0, J_0)$, which we have obtained in Section 6.4. The expected revenue, on the other side, is taken under the $Q_{\theta_\star}$ measure,

$$E^{Q_{\theta^\star}}[\int_0^T e^{-rt}\delta S_t dt].$$

Recall that $\frac{dS_t}{S_t} = rdt + \sigma_{J_t} d\bar{W}_t$ under $Q_{\theta^\star}$, we have $E^{Q_{\theta^\star}}[S_t] = S_0 e^{rt}$ and therefore

$$E^{Q_{\theta^\star}}[\int_0^T e^{-rt}\delta S_t dt] = \delta S_0 T.$$

The regular fee, which covers the cost of continuous hedging cost, is then the solution to $P^R(S_0, J_0) = \delta S_0 T$.

To determine the loading, we use the equation

$$95\% \text{ quantile of rebalancing cost} = \text{expected revenue},$$

where the LHS is computed by the semi-analytic algorithm and RHS is taken under the physical measure P

$$E^P[\int_0^T e^{-rt}\delta S_t dt] = \int_0^T e^{-rt}\delta E^P(S_t)dt = \int_0^T e^{-rt}\delta S_0 E^P(e^{X_t})dt.$$

Now we show how to compute $E^P(e^{X_t})$. Suppose $X_0 = 0, J_0 = 1$, the drift and volatility of X are μ_i and σ_i in state $i \in \{1,2\}$ (we write the drift as μ_i for convenience, whereas it should be $\mu_i - \delta - \frac{1}{2}\sigma_i^2$ after the management fee is deducted).
According to Theorem 6.2.1,

$$e^{\alpha X_t - \theta(\alpha)t} h_{J_t}(\alpha),$$

is a martingale, where the intensity matrix of J_t is $Q = \begin{bmatrix} -\lambda & \lambda \\ v & -v \end{bmatrix}$, $\theta(\alpha)$ is defined in (6.2.6), $h_1(\alpha) = -\lambda, h_2(\alpha) = \varphi_1(\alpha) - \lambda - \theta(\alpha)$ and $\varphi_i(\alpha) = \mu_i\alpha + \frac{1}{2}\sigma_i^2\alpha^2, i = 1,2$. Let $\alpha = 1$, we have

$$E^P[e^{X_t - \theta(1)t} h_{J_t}(1)] = h_{J_0}(1) = -\lambda,$$

or equivalantly

$$E^P[e^{X_t} h_{J_t}(1)] = -\lambda e^{\theta(1)t}. \tag{6.5.13}$$

The LHS of (6.5.13) can be decomposed to

$$\begin{aligned} &E^P[e^{X_t} h_{J_t}(1) I_{\{J_t=1\}}] + E^P[e^{X_t} h_{J_t}(1) I_{\{J_t=2\}}] \\ &= -\lambda E^P[e^{X_t} I_{\{J_t=1\}}] + (\varphi_1(1) - \lambda - \theta(1)) E^P[e^{X_t} I_{\{J_t=2\}}]. \end{aligned}$$

So

$$-\lambda E^P[e^{X_t} I_{\{J_t=1\}}] + (\varphi_1(1) - \lambda - \theta(1)) E^P[e^{X_t} I_{\{J_t=2\}}] = -\lambda e^{\theta(1)t}. \tag{6.5.14}$$

Since we have two values for $\theta(1)$ (given by (6.2.6)), (6.5.14) offers two equations, from which $E^P[e^{X_t} I_{\{J_t=1\}}]$ and $E^P[e^{X_t} I_{\{J_t=2\}}]$ can be solved. Sum them together, we get

	$\lambda = 0.5$ $v = 1$	$\lambda = 0.5$ $v = 2$	$\lambda = 1$ $v = 0.5$	$\lambda = 1$ $v = 2$	$\lambda = 2$ $v = 0.5$	$\lambda = 2$ $v = 1$
Regular Fee	0.1026	0.1001	0.1100	0.1038	0.1146	0.1120
Loading	0.0110	0.0091	0.0093	0.0089	0.0072	0.0078

Table 6.5.3: The Management fee for discrete hedging. The common parameters are: $T = 3, S_0 = K = 50, r = 0.02, \alpha = 0.05, \mu_1 = 0.1, \mu_2 = 0.15, \sigma_1 = 0.3, \sigma_2 = 0.4, d = 0.03, J_0 = 1$.

$E^P[e^{X_t}]$.

In Table 6.5.3, we calculate the regular fee and loading for different transition rates of the two states.

Under a GBM model with no regime switching, the regular fee/loading is 0.0945/ for state 1: $\mu = 0.1, \sigma = 0.3$ and 0.1220/ for state 2: $\mu = 0.15, \sigma = 0.4$. Comparing these values with those in Table 6.5.3, we observe an interesting disparity between continuous and discrete hedging, in terms of the cost. With the introduction of regime switching, the regular fee-the cost of continuous hedging-lies strictly within its two extremes (the value for a single state), while the loading-the cost of discrete hedging-becomes uniformly larger, in response to the increased level of stochasticity.

There is no obvious trend among the discrete hedging costs for different transition rates. As we have commented in Section 5.1, the mean return rate μ and the volatility σ have opposite impacts on the loading, whereas μ and σ in one state should in principle be uniformly larger or smaller than they are in the other state, for higher returns usually come with higher risks.

6.6 Application to Structured Product based VA

It is possible to generalize the algorithm for vanilla put option to any VA product with a path-independent payoff. In this section, we consider, under the regime switching GBM model, the hedging cost analysis of the Structured Product based VA with buffer protection, introduced in Section 5.3.1.

The generalization is straightforward since the only part we need to modify in the original algorithm is the function $h^R(S_0, J_0)$. Denote by $h^{spR}(S_0, J_0)$ the expectation of first rebalancing cost, given the initial sub-account value S_0 and initial state of the underlying Markov chain J_0

$$e^{-r(\tau_1\wedge\epsilon_1^\lambda)}\left[P^{spR}_{\tau_1\wedge\epsilon_1^\lambda} - \Delta_0^{spR} e^{d(\tau_1\wedge\epsilon_1^\lambda)} S_{\tau_1\wedge\epsilon_1^\lambda} - M_0^{spR} e^{r(\tau_1\wedge\epsilon_1^\lambda)}\right],$$

where $P_t^{spR}, \Delta_t^{spR}, M_t^{spR}$ are the price, Delta and money account value for an option with payoff (5.3.15), respectively; τ_1 is the hitting time of the band and ϵ_1^λ is the exponential maturity of the contract.

In Table 6.6.4 and 6.6.5, we calculate the management fee of the spVA under regime switching GBM model for several contract and model parameters, which lead to the following observations

1. Like the case of the GBM model, the regular fee obtained under the regime switching GBM model is also negative, exposing the fact that the policyholder essentially assumes a short put position;

2. The regular fee is independent of μ; Moreover, if the underlying Markov chain spends more time in the high volatility state (as the intensities change from $\lambda = 1, v = 2$ to $\lambda = 2, v = 1$), the regular fee decreases. This is consistent with our earlier finding for the spVA under the GBM model, that the higher the volatility, the lower the regular fee;

3. In contrast with our earlier finding that the increase in μ or σ will result in smaller loading, the loading in the case of regime switching GBM model varies oppositely. When the underlying Markov chain spends more time in the state in which both μ and σ are higher, the loading increases;

4. The impact of b and c on the regular fee and the loading are the same as in the

	$b=0.1$ $c=0.2$	$b=0.1$ $c=0.3$	$b=0.1$ $c=0.4$	$b=0.2$ $c=0.2$	$b=0.2$ $c=0.3$	$b=0.2$ $c=0.4$
regular fee	-0.0630	-0.0763	-0.0897	-0.0549	-0.0691	-0.0833
loading	0.0087	0.0091	0.0096	0.0077	0.0081	0.0090

Table 6.6.4: Management fee for spVA under the regime switching GBM model. b is the buffer level and c the cap level. The common parameters are: $T = 3, S_0 = 50, \mu_1 = 0.1, \mu_2 = 0.15, \sigma_1 = 0.3, \sigma_2 = 0.4, r = 0.02, \alpha = 0.05, \lambda = 1, v = 2$.

	$b=0.1$ $c=0.2$	$b=0.1$ $c=0.3$	$b=0.1$ $c=0.4$	$b=0.2$ $c=0.2$	$b=0.2$ $c=0.3$	$b=0.2$ $c=0.4$
regular fee	-0.0681	-0.0815	-0.0950	-0.0594	-0.0737	-0.0880
loading	0.0092	0.0097	0.0104	0.0082	0.0090	0.0094

Table 6.6.5: Management fee for spVA under the regime switching GBM model. b is the buffer level and c the cap level. The common parameters are: $T = 3, S_0 = 50, \mu_1 = 0.1, \mu_2 = 0.15, \sigma_1 = 0.3, \sigma_2 = 0.4, r = 0.02, \alpha = 0.05, \lambda = 2, v = 1$.

case of GBM model.

Chapter 7

Future Work and Conclusion

In the future, we plan to generalize the semi-analytic algorithm for the hedging cost analysis of more VA products, especially those with high-water-mark and withdrawal features. In this chapter, we first present some preliminary results that we have obtained for the high-water-mark VA in Section 7.1. The discrete hedging problem for the GMWB, however, is extremely challenging. Even in the continuous setting, the price of the GMWB can only be calculated by Monte Carlo simulation or sophisticated analytical approximation. Finally we conclude the thesis in Section 7.2.

7.1 Preliminary Results on the High-Water-Mark VA

A high-water-mark VA is very similar to the floating strike European lookback option with payoff $M_T - S_T$, where S_T is the value of the sub-account at maturity T and $M_T = \max_{0 \le t \le T} S_t$ is the running maximum of S. For this kind of VA, we consider the following 4 discrete hedging strategies

1. S_t based: we rebalance the hedging portfolio whenever the value of the underlying asset S_t hits a two-sided band with width α: $[S_t e^{-\alpha}, S_t e^{\alpha}]$;

	S_t based	M_t based	M_t and S_t based	Δ_t based
mean	-0.4182	-7.2049	1.2970	-0.2740
std	2.6744	18.8359	6.5807	1.5499
skewness	-2.2890	0.0701	1.2038	-0.1121
kurtosis	20.2518	2.9146	5.6948	4.8534
90% quartile	1.9511	17.9791	10.1311	1.5488
95% quartile	2.8603	24.0801	13.7253	2.1594
97.5% quartile	3.7876	29.0879	17.2060	2.7790
99% quartile	5.1369	35.1116	21.7860	3.6232

Table 7.1.1: Summary statistics: the cost distribution of the 4 hedging strategies for a high-water-mark VA. The common parameters are $T = 3, S0 = 50, r = 0.02, \mu = 0.1, \sigma = 0.3$. To make a fair comparison, the bandwidth parameter α are tuned so that the hedging frequency of the 4 hedging strategies are roughly the same. In particular, S_t based: $\alpha = 0.05$ with average number of rebalance 100; M_t based: $\alpha = 0.003$ with average number of rebalance 106; M_t and S_t based: $\alpha = 0.016$ with average number of rebalance 101; Δ_t based: $\alpha = 0.11$ with average number of rebalance 99.

2. M_t based: we rebalance the hedging portfolio whenever M_t-the running maximum of S_t-reaches a new level $M_t e^{\alpha}$;

3. M_t and S_t based: we rebalance the hedging portfolio whenever M_t rises to $M_t e^{\alpha}$ or S_t falls to $S_t e^{-\alpha}$;

4. Δ_t based: we rebalance the hedging portfolio whenever the absolute change of the option Delta exceed α(in other words, the band is $[\Delta_t - \alpha, \Delta_t + \alpha]$).

Table 7.1.1 summarize the distributional statistics of the these four strategies. Clearly, the two-sided underlier-based hedging is no longer the favorite, as the Δ_t based exhibits smaller variance, smaller kurtosis and lighter right tail for a given hedging frequency. So we turn to the analysis of the Δ_t based discrete hedging. According to Proposition 6.7.2 of Musiela and Rutkowski (2011), the time t price of a high-water-mark VA is

$$LP_t = -sN(-\hat{d}) + Me^{-r\tau}N(-\hat{d}+\sigma\sqrt{\tau}) + s\frac{\sigma^2}{2r}N(\hat{d}) - e^{-r\tau}s\frac{\sigma^2}{2r}(\frac{M}{s})^{2r\sigma^{-2}}N(\hat{d} - 2r\sigma^{-1}\sqrt{\tau}), \tag{7.1.1}$$

where $s = S_t, M = M_t, \tau = T - t$, r is the risk free interest rate and $\hat{d} = \frac{\log(s/M)+(r+\frac{1}{2}\sigma^2)\tau}{\sigma\sqrt{\tau}}$.
The analysis of the Δ_t based discrete hedging requires the (two-sided) hitting times densities of Δ_t. Interestingly, Δ_t turns out to be a (though highly nonlinear) function of only two variables-the time to maturity τ and the ratio $\frac{M_t}{S_t}$ (of course, Delta depends on other model parameters, such as r and σ. But here we only look at time-varying ones).
So as a first attempt, we have identified the two-sided hitting time densities of $\frac{M_t}{S_t}$.
Assuming a GBM model with $S_t = S_0 e^{X_t}$ and $X_t = (\mu - d - \frac{1}{2}\sigma^2)t + \sigma B_t = \alpha t + \sigma B_t, X_0 = 0$, then $\frac{M_t}{S_t} = e^{\bar{X}_t - X_t}$, where B_t is a standard Brownian motion under the physical measure P and $\bar{X}_t = \max_{0 \le s \le t} X_s$. So the two-sided hitting problem of $\frac{M_t}{S_t}$ can be translated to that of the reflected processes $\bar{X}_t - X_t$.
We define, in a more general sense, the reflected process of X_t by $Y_t = (s \vee \bar{X}_t) - X_t, Y_0 = s - x = z \ge 0$. For $k > z > 0$, we are interested in the following stopping times of Y_t
$T_{0,k} := \inf\{t \ge 0 : Y_t \notin (0, k)\}$;
T_k: the first time when Y_t hits k without hitting 0 earlier;
T_0: the first time when Y_t hits 0 without hitting k earlier.
Note that $T_k < \infty (T_0 < \infty)$ implies $T_0 = \infty (T_k = \infty)$.

From $X_0 = x < s$ and the definition of Y_t, we know the first time Y_t hits 0 is the first time X_t rises to s and before that, $Y_t = s - X_t$. So

$$\begin{aligned} T_k &= \inf\{t : Y_t = k, Y_u > 0 (0 \le u \le t)\} \\ &= \inf\{t : X_t = s - k, X_u < s (0 \le u \le t)\}, \end{aligned}$$

i.e. T_k is the two-sided stopping time τ_{s-k} of X_t, with $X_0 = x$, hitting $s - k$ before s.
Similarly, T_0 is the two-sided stopping time τ_s of X_t, with $X_0 = x$, hitting s before $s - k$.
According to Proposition 1 of Avram et al. (2004), the Laplace transform of T_k and T_0

are

$$
\begin{aligned}
E_x(e^{-qT_k}) &= E_x[e^{-qT_k} I_{\{\tau_s < \tau_{s-k}\}}] + E_x[e^{-qT_k} I_{\{\tau_s > \tau_{s-k}\}}] \\
&= E_x[e^{-qT_k} I_{\{\tau_s > \tau_{s-k}\}}] = E_x[e^{-q\tau_{s-k}} I_{\{\tau_s > \tau_{s-k}\}}] \\
&= Z^{(q)}(x-s+k) - W^{(q)}(x-s+k)\frac{Z^{(q)}(k)}{W^{(q)}(k)} \qquad (7.1.2)
\end{aligned}
$$

and

$$
\begin{aligned}
E_x(e^{-qT_0}) &= E_x[e^{-qT_0} I_{\{\tau_s < \tau_{s-k}\}}] + E_x[e^{-qT_0} I_{\{\tau_s > \tau_{s-k}\}}] \\
&= E_x[e^{-q\tau_s} I_{\{\tau_s < \tau_{s-k}\}}] \\
&= \frac{W^{(q)}(s-x+k)}{W^{(q)}(k)} \qquad (7.1.3)
\end{aligned}
$$

respectively.

$W^{(q)}(x)$ and $Z^{(q)}(x)$ in (7.1.2) and (7.1.3) are called the scale functions and admit explicit forms for drifted Brownian motion X_t

$$
\begin{aligned}
W^{(q)}(x) &= \frac{2}{\sigma^2 \epsilon} e^{\gamma x} \sinh(\epsilon x), \\
Z^{(p)}(x) &= e^{\gamma x} \cosh(\epsilon x) - \frac{\gamma}{\epsilon} e^{\gamma x} \sinh(\epsilon x), \\
\gamma &= -\frac{\alpha}{\sigma^2}, \\
\epsilon &= \epsilon(q) = \sqrt{\frac{\alpha^2}{\sigma^4} + \frac{2q}{\sigma^2}} = \sqrt{\gamma^2 + \frac{2q}{\sigma^2}}.
\end{aligned}
$$

With these specifications, (7.1.2) and (7.1.3) become

$$
\begin{aligned}
E_x(e^{-qT_k}) &= e^{\gamma m} \cosh(\epsilon m) - e^{\gamma m} \frac{\sinh(\epsilon m)\cosh(\epsilon k)}{\sinh(\epsilon k)} \\
&= e^{\gamma m} \frac{\sinh(\epsilon z)}{\sinh(\epsilon k)}, \\
m &= k - z > 0 \qquad (7.1.4)
\end{aligned}
$$

and

$$E_x(e^{-qT_0}) = e^{\gamma(m-k)}\frac{\sinh(\epsilon m)}{\sinh(\epsilon k)}. \tag{7.1.5}$$

The inverse transform of (7.1.4) and (7.1.5) can be written in closed forms, thanks to Roberts G.E. and Kaufman H. (1966). In particular, the inverse transform of (7.1.4), i.e. the density of T_k is

$$f_{T_k}(t) = -e^{\gamma m}\frac{\sigma^2}{2}e^{-\frac{\sigma^2\gamma^2}{2}t}\frac{1}{\sigma\sqrt{\frac{\pi t}{2}}}\sum_{n=-\infty}^{\infty}\left[e^{-\frac{2k^2}{\sigma^2 t}(\frac{z}{2k}+\frac{1}{2}+n)^2}\frac{2k}{\sigma^2 t}(\frac{z}{2k}+\frac{1}{2}+n)\right], \tag{7.1.6}$$

and the inverse transform of (7.1.5), i.e. the density of T_0 is

$$f_{T_0}(t) = -e^{\gamma(m-k)}\frac{\sigma^2}{2}e^{-\frac{\sigma^2\gamma^2}{2}t}\frac{1}{\sigma\sqrt{\frac{\pi t}{2}}}\sum_{n=-\infty}^{\infty}\left[e^{-\frac{2k^2}{\sigma^2 t}(\frac{m}{2k}+\frac{1}{2}+n)^2}\frac{2k}{\sigma^2 t}(\frac{m}{2k}+\frac{1}{2}+n)\right] \tag{7.1.7}$$

Note that once Y_t hits 0, the band for re-balancing will become one-sided, since Y_t cannot assume negative values. So we also need the density of the one-sided hitting time, which is defined as

$$\bar{\sigma}_k = \inf\{t : Y_t = k, Y_0 = 0\}.$$

Theorem 1 in Avram et al. (2004) provides the Laplace transform of $\bar{\sigma}_k$

$$\begin{aligned} E_0[e^{-q\bar{\sigma}_k}] &= Z^{(u)}(k) - W^{(u)}(k)\frac{uW^{(u)}(k)}{W^{(u)\prime}(k)} \\ &= e^{\gamma k}\cosh(\epsilon k) - \frac{\gamma}{\epsilon}e^{\gamma k}\sinh(\epsilon k) - \frac{\frac{2q}{\sigma^2}e^{\gamma k}\sinh^2(\epsilon k)}{\gamma\epsilon\sinh(\epsilon k)+\epsilon^2\cosh(\epsilon k)}. \end{aligned} \tag{7.1.8}$$

Unfortunately, we are not able to invert (7.1.8) analytically. Instead, we use the method developed in Beylkin and Monzon (2005) to approximate (7.1.8) with exponential sums $\sum_{m=1}^{M}\omega_m e^{t_m u}$. As we have seen in Chapter 6, this is equivalent to approximate a continuous r.v. with a discrete one. Figure 7.1.1 shows the satisfactory accuracy of this approximation.

Next, we want to find the density of $Y_t \in dy$ with $Y_0 = s - x = z > 0$, given that Y_t

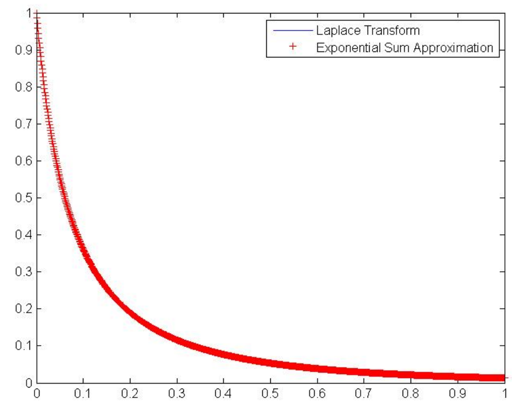


Figure 7.1.1: Exponential sum approximation to the Laplace transform of $\bar{\sigma}_k$. $k = 1$. The parameters in the approximation are $M = 10, \omega = [0.0000, 0.0317, 0.1411, 0.2206, 0.2327, 0.1929, 0.1205, 0.0496, 0.0103, 0.0006], t = [2.6696, -1.5012, -3.2368, -6.2839, -11.1481, -18.4692, -29.3484, -45.7304, -71.2230, -114.2089]$

did not hit 0 or k in the time interval $[0, t)$. This can be first translated to the problem of finding the density of $X_t \in d(s-y)$ with $X_0 = x$, given that X_t did not hit $s-k$ or s during $[0,t)$, and then to finding the density of $X_t \in d(k-y)$ with $X_0 = k - z$, given that X_t did not hit 0 or k during $[0, t)$. The Laplace transform of the last density, called the potential measure, is provided in Theorem 8.7 of Kyprianou A. E. (2006)

$$
\begin{aligned}
u^{(q)}(k-z, k-y) &= \int_0^\infty e^{-qt} P_z(Y_t \in dy, T_{0,k} > t)dt \\
&= \int_0^\infty e^{-qt} P_{k-z}(X_t \in d(k-y), \tau_0 \wedge \tau_k > t)dt \\
&= \frac{W^{(q)}(m)W^{(q)}(y)}{W^{(q)}(k)} - W^{(q)}(h) \\
&= \frac{2}{\sigma^2 \epsilon} e^{\gamma h} \left(\frac{\sinh(\epsilon m)\sinh(\epsilon y)}{\sinh(\epsilon k)} - \sinh(\epsilon h) I_{\{h>0\}} \right), \\
m &= k - z > 0, \\
h &= y - z. \qquad (7.1.9)
\end{aligned}
$$

The inversion is also achieved by exponential sum approximation. See Figure 7.1.2 for the result. As we mentioned before, when Y_t hits 0, the band turns to one-sided, so the

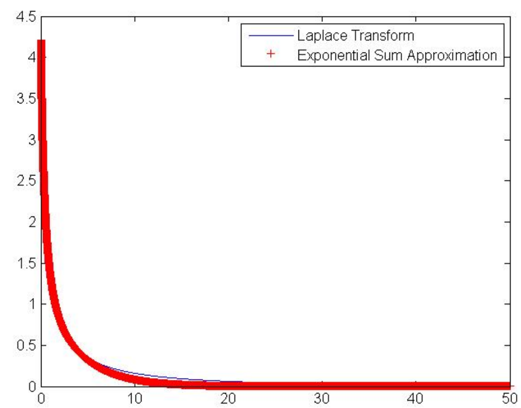


Figure 7.1.2: Exponential sum approximation to $u^{(q)}(k-z, k-y)$. $k = 1, y = 0.6, z = 0.5$. The parameters in the approximation are $M = 6, \omega = [1.2166, 1.0187, 1.5790, 0.3438, 0.0520, 0.0021], t = [-0.2676, -3.4871, -1.3696, -6.8517, -11.9951, -20.2963]$.

potential measure also needs to be adjusted. Specifically, we are interested in $P_{Y_0=0}(Y_t \in dy, \bar{\sigma}_k > t)$. From Theorem 8.11 of Kyprianou A. E. (2006), we know

$$\begin{aligned} \bar{U}^{(q)}(0, dy) &= \int_0^\infty e^{-qt} P_{Y_0=0}(Y_t \in dy, \bar{\sigma}_k > t) dt \\ &= \left[W^{(q)}(k) \frac{W^{(q)\prime}(y)}{W^{(k)\prime}(k)} - W^{(q)}(y) \right] dy \\ &= \frac{2}{\sigma^2 \epsilon} e^{\gamma y} \left[\sinh(\epsilon k) \frac{\frac{\gamma}{\epsilon} \sinh(\epsilon y) + \cosh(\epsilon y)}{\frac{\gamma}{\epsilon} \sinh(\epsilon k) + \cosh(\epsilon k)} - \sinh(\epsilon y) \right]. \quad (7.1.10) \end{aligned}$$

Again, the sum of exponential approximation is used, as shown in Figure 7.1.3.

So far, we have obtained the hitting time densities and the potential measures for the two-sided stopping times of the ratio $\frac{M_t}{S_t}$. Since Δ_t of the high-water-mark VA is a function of $\frac{M_t}{S_t}$ and the time to maturity, the quantities we have found here form a basis for the computation of their counterparts in terms of Δ_t.

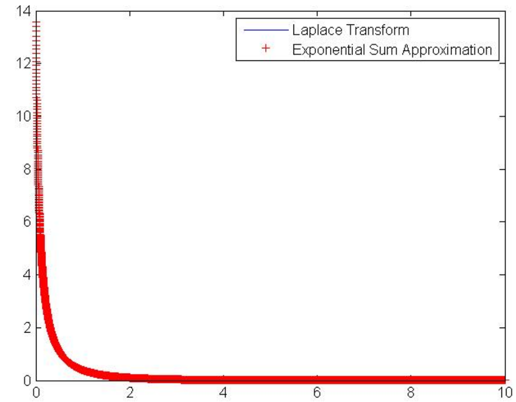


Figure 7.1.3: Exponential Sum Approximation to $\bar{U}^{(q)}(0,dy)$. $k=1, y=0.5$. The parameters in the approximation are: $M=10, \omega=[0.5085, 1.7775, 2.6773, 2.9029, 2.5599, 1.8144, 0.9581, 0.3255, 0.0545, 0.0025], t=[-0.8801, -2.3806, -5.0701, -9.2649, -15.3960, -24.1686, -36.7642, -55.1767, -83.0301, -128.7477]$.

7.2 Conclusion

In this thesis, we first investigate various discrete hedging strategies for put option and compare their relative efficiencies based on the severity of extreme losses. We identify the two-sided underlier-based hedging as the most suitable in that it produces the lightest right tail for a given hedging frequency. Then we assume the GBM model to develop a semi-analytic framework for the cost analysis of move-based discrete hedging and with the resulting cost distribution, we propose the "Percentile Premium Principle", which breaks the premium an insurer should charge into two parts-the regular fee for covering the cost of continuous hedging and the loading for the additional cost arising from discrete re-balances. We demonstrate the rationale for the new premium principle by applying, with necessary extensions, the semi-analytic algorithm to the pricing of some popular VA designs, including GMMB, annual ratchet VA and structured product based VA with buffer/contingent protection. It turns out that the loading, once deemed negligible by most VA providers, is too significant to be ignored. Finally, we generalize the algorithm to the case of the regime switching GBM model with two regimes, which allows a better modeling of the underlying economy over long period.

The key idea behind our semi-analytic algorithm is maturity randomization, which is first introduced for financial application in Carr (1998). The option value suffers no time decay once we replace the fixed maturity by an independent exponential r.v.. This characteristic paves the way for a recursive formulation of each re-balancing cost. We also showed the total expected cost associated with the random maturity is the Laplace transform of that with the fixed maturity, so the latter can be retrieved through numerical inversion.

Bibliography

Angelini F. and Herzel S. (2009). Measuring the Error of Dynamic Hedging: a Laplace Transform Approach. The Journal of Computational Finance, 13(2), 47-72.

AnnuityDigest. http://www.annuitydigest.com/b/what-driving-rush-variable-annuity-exit

Asmussen S. and Kella O. (2000). A Multi-dimensional Martingale for Markov Additive Processes and Its Applications. Advances in Applied Probability, 32(2), 376-393.

AXA Equitable. Structured Capital Strategies. http://www.axa-equitable.com/Elements/channel/products/structured-capital-strategies.html

Avram F., Kyprianou A. E. and Pistorius M. R. (2004). Exit Problems for Spectrally Negative Levy Processes and Applications to (Canadized) Russian Options. The Annals of Applied Probability, 14(1), 215-238.

Bauer D., Kling A. and Russ J. (2008). A Universal Pricing Framework for Guaranteed Minimum Benefits in Variable Annuities. ASTIN Bulletin, 38(2), 621-651.

Belanger A., Forsyth P.A. and Labahn G. (2009). Valuing the Guaranteed Minimum Death Benefit Clause with Partial Withdrawals. Applied Mathematical Finance, 16, 451-496.

Bertsimas D., Kogan L. and Lo A. (2000). When is Time Continuous? Journal of Financial Economics, 55(2), 173-204.

Beylkin G. and Monzon L. (2005). On Approximation of Functions by Exponential Sums. Applied and Computational Harmonic Analysis, 19, 17-48.

Black F. and Scholes M. (1973). The Pricing of Options and Corporate Liabilities. Journal of Political Economy, 81, 637-654.

Blamont D. and Sagoo P. (2009). Pricing and Hedging of Variable Annuities. Life & Pensions, 2, 39-44.

Boyle P. and Emanuel D. (1980). Discretely Adjusted Option Hedges. Journal of Financial Economics, 8, 259-282.

Boyle P., Kolkiewicz A.W. and Tan K.S. (2001). Valuation of the Reset Options Embedded in Some Equity-linked Insurance Products. North American Actuarial Journal, 5(3), 1-18.

Boyle P. and Draviamb T. (2007). Pricing Exotic Options under Regime Switching. Insurance: Mathematics and Economics, 40(2), 267-288.

Brancik L. (2011). Numerical Inverse Laplace Transforms for Electrical Engineering Simulation, MATLAB for Engineers - Applications in Control, Electrical Engineering, IT and Robotics. InTech. Available from: http://www.intechopen.com/books/matlab-for-engineersapplications-in-control-electrical-engineering-it-and-robotics/numerical-inverse-laplace-transforms-forelectrical-engineering-simulation

Buffington J. and Elliott R.J. (2002). American Options with Regime Switch- ing. International Journal of Theoretical and Applied Finance, 5, 497-514.

Carr P. (1998). Randomization and the American put. Review of Financial Studies, 11, 597-626.

Cheah P., Fraser D. and Reid N. (1993). Some Alternatives to Edgeworth. Canadian Journal of Statistics, 21, 131-138.

Chi Y. and Lin X.S. (2012). Are Flexible Premium Variable Annuities Under-priced? ASTIN Bulletin, forthcoming.

Coleman T.F., Kim Y., Li Y. and Patron, M. (2007). Robustly Hedging Variable Annuities with Guarantees under Jump and Volatility Risks. Journal of Risk and Insurance, 74, 347-376.

Dai M., Kuen Kwok Y. and Zong J. (2008). Guaranteed Minimum Withdraw Benefit in Variable Annuities. Mathematical Finance, 18, 595-611.

Davis, M. (1997). Option Pricing in Incomplete Markets. In Mathematics of Derivative Securities, ed. M. A. H. Dempster and S. R. Pliska, 216-226. Cambridge University Press.

Deng G., Husson T. and McCann, C.J. (2012). Structured Product Based Variable Annuities. 2012 Academy of Financial Services Annual Meeting Paper. Available at SSRN: http://ssrn.com/abstract=2049513 or http://dx.doi.org/10.2139/ssrn.2049513

Dupire B. (2005). Optimal Process Approximation: Application to Delta Hedging and Technical Analysis. Quantitative finance: Developments, Applications and Problems, Cambridge UK. http://www.newton.ac.uk/programmes/DQF/seminars/070714501.pdf

Durrett R. (2010). Probability: Theory and Examples, 4th edition. Cambridge University Press.

Elliott R.J., Chan L. and Siu T.K. (2005). Option Pricing and Esscher Trans form Under Regime Switching. Annals of Finance, 1(4), 423-432.

Elliott R., Siu T. K., Chan L. and Lau J. W. (2007). Pricing Options under a Generalized Markov-modulated Jump-diffusion Model. Stochastic Analysis and Applications, 25(4), 821-843.

Esscher F. (1932). On the Probability Function in the Collective Theory of Risk. Scandinavian Actuarial Journal, 15, 175-195.

Follmer H. and Schweizer M. (1991). Hedging of Contingent Claims under Incomplete Information. In Applied Stochastic Analysis, ed. M. H. A. Davis, and R. J. Elliott, 389-414. Gordon and Breach.

Follmer H. and Sondermann D. (1986). Hedging of Contingent Claims under Incomplete Information. In Contributions to Mathematical Economics, ed. W. Hildenbrand and A. Mas-Colell, 205-223. North Holland.

Forsyth P., Vetzal K. and Windcliff H. (2003). Hedging Segregated Fund Guarantees, Chapter 10, The Pension Challenge: Risk Transfers and Retirement Income Security, Ed. O. S. Mitchell, and K. Smetters, 214-237. Oxford University Press.

Gerber H. and Shiu E. (1994). Option Pricing by Esscher Transforms (with discussions). Transactions of the Society of Actuaries, 46, 99-191.

Glasserman P. (2003). Monte Carlo Methods in Financial Engineering, Springer, New York.

Goldfeld M. and Quandt E. (1973). A Markov Model for Switching Regressions. Journal of Econometrics, 1, 3-16.

Hamilton D. (1989). A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. Econometrica, 57, 357-384.

Hayashi T. and Mykland P. (2005). Evaluating Hedging Errors: An Asymptotic Approach. Mathematical Finance, 15(2), 309-343.

Hardy M.R. (2001). A Regime-Switching Model of Long-Term Stock Returns. North American Actuarial Journal, 5(2), 41-53.

Hardy M.R. (2003). Investment Guarantees: Modeling and Risk Management for Equity-linked Life Insurance. Wiley (New York).

Hardy M.R. (2004). Ratchet Equity Indexed Annuities. In 14th Annual International AFIR Colloquium.

He H., Keirstead W.P. and Rebholz J. (1998). Double Lookbacks. Mathematical Finance, 8, 201-228.

Henrotte P. (1993). Transaction Costs and Duplication Strategies. Working Paper. Stanford University.

Hollenbeck K. (1998). INVLAP.M: A Matlab Function for Numerical Inversion of Laplace Transforms by the de Hoog Algorithm. http://www.isva.dtu.dk/staff/karl/invlap.htm

Hull J. (2011). Options, Futures, and Other Derivatives, 8th edition. Prentice Hall.

Insured Retirement Institute (2011). The 2011 IRI Fact Book, http://www.IRIonline.org

Jaimungal S. and Young V. (2005). Pricing Equity-linked Pure Endowments with Risky Assets that Follow Levy Processes. Insurance: Mathematics and Economics, 36,(3), 329-346.

Jaimungal S., Donnelly R. and Rubisov D. (2012). Valuing GWBs with Stochastic Interest Rates and Volatility. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1984885

Karatzas I. and Shreve S.E. (1991). Brownian Motion and Stochastic Calculus, 2nd edition. Springer.

Kim J. and Levisohn B. (2010). Structured Notes: Not as Safe as They Seem. The Wall Street Journal. http://online.wsj.com/article/SB10001424052748704804504575606843404836032.html

Kyprianou A. E. (2006). Introductory Lectures on Fluctuations of Levy Processes with Applications. Springer.

Lin X.S. (1998). Double Barrier Hitting Time Distributions with Applications to Exotic Options. Insurance: Mathematics and Economics, 23, 45-58.

Lin X.S. and Tan K.S. (2003). Valuation of Equity-indexed Annuities under Stochastic Interest Rates. North American Actuarial Journal, 7(3), 72-91.

Lin, X. S. (2006). Introductory Stochastic Analysis for Finance and Insurance. John Wiley & Sons.

Lin X.S., Tan K.S. and Yang H. (2009). Pricing Annuity Guarantees under a Regime-switching Model. North American Actuarial Journal, 13(3), 316-338.

Marquardt T., Platen E. and Jaschke S. (2008). Valuing Guaranteed Minimum Death Benefit Options in Variable Annuities under a Benchmark Approach. Available at: http://www.business.uts.edu.au/qfrc/research/research_papers/rp221.pdf

Marshall C., Hardy M. and Saunders D. (2010). Valuation of a Guaranteed Minimum Income Benefit. North American Actuarial Journal, 14(1), 38-58.

Martellini L. and Priaulet P. (2002). Competing Methods for Option Hedging in the Presence of Transaction Costs. The Journal of Derivatives, Spring 2002, 9(3), 26-38.

McDonald R. (2009). Derivatives Markets,3rd edition. Prentice Hall.

McNeil A., Frey R. and Embrecht P. (2005). Quantitative Risk Management: Concepts, Techniques, and Tools. Princeton University Press.

Melnikov A. and Romanyuk Y. (2008). Efficient Hedging and Pricing of Equity-linked Life Insurance Contracts on Several Risky Assets. International Journal of Theoretical and Applied Finance, 11(3), 295-320.

Milevsky M. and Posner S. (2001). The Titanic Option: Valuation of the Guaranteed Minimum Death Benefit in Variable Annuities and Mutual Funds. The Journal of Risk and Insurance, 68(1), 93-128.

Milevsky M.A. and Salisbury T.S. (2006). Financial Valuation of Guaranteed Minimum Withdrawal Benefits. Insurance: Mathematics and Economics, 38(1), 21-38.

Moller T. (2001). Hedging Equity-linked Life Insurance Contracts. North American Actuarial Journal, 5(2), 79-95.

Musiela M. and Rutkowski M. (2011). Martingale Methods in Financial Modeling, 2nd edition. Springer.

Naik V. (1993). Option Valuation and Hedging Strategies with Jumps in the Volatility of Asset Returns. Journal of Finance, 48(5), 1969-84.

Ng A.C. and Li J.S. (2011). Valuing Variable Annuity Guarantees with the Multivariate Esscher Transform. Insurance Mathematics and Economics, 49, 393-400.

Oliver Wyman Limited (2007). VA VA Voom: Variable Annuities Are in Pole Position to Meet the Requirements of the European Asset Protection Market. http://www.mmc.com/knowledgecenter/OliverWymanVariableAnnuities.pdf

Peng J., Leung K. and Kowk Y. (2010). Pricing Guaranteed Minimum Withdrawal Benefits under Stochastic Interest Rates. Quantitative Finance, 12(6), 933-941.

Piscopo G. and Haberman S. (2011). The Valuation of Guaranteed Lifelong Withdrawal Benefit Options in Variable Annuity Contracts and the Impact of Mortality Risk. North American Actuarial Journal, 15(1), 59-76.

Quandt E. (1958). The Estimation of Parameters of Linear Regression System Obeying Two Separate Regimes. Journal of the American Statistical Association, 55, 873-880.

Roberts G.E. and Kaufman H. (1966). Table of Laplace Transforms. Philadelphia Saunders.

Samuelson P.A. (1973). Mathematics of Speculative Price. SIAM Review, 15(1), 1-42.

Schweizer M. (1996). Approximation Pricing and the Variance-Optimal Martingale Measure. Annals of Probability, 24, 206-236.

Shah P. and Bertsimas D. (2008). An Analysis of the Guaranteed Withdrawal Benefits for Life Option. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1312727

Siu T. K. (2005). Fair Valuation of Participating Policies with Surrender Options and Regime Switching. Insurance: Mathematics and Economics, 37(3), 533-552.

Siu T.K. (2008). A Game Theoretic Approach to Option Valuation Under Markovian Regime-Switching Models. Insurance: Mathematics and Economics 42(3), 1146-1158.

Siu T. K., Lau J. W. and Yang H. (2008). Pricing Participating Products under a Generalized Jump-Diffusion. Journal of Applied Mathematics and Stochastic Analysis, 2008, Article ID 474623, 30 pages.

Siu T.K. (2011). Regime Switching Risk: To Price or Not To Price?. International Journal of Stochastic Analysis, 2011, Article ID 843246, 14 pages.

Surkov V., Kenneth J. and Jaimungal S. (2007). Fourier Space Time Stepping for Option Pricing with Levy Models. Journal of Computational Finance, 12(2), 1-29.

Taylor S.J. (2005). Asset Price Dynamics, Volatility, and Prediction. Princeton University Press.

Toft K. (1996). On the Mean-Variance Trade-off in Option Replication with Transactions Costs. The Journal of Financial and Quantitative Analysis, 31(2), 233-263.

Wang Y. (2009). Quantile Hedging for Guaranteed Minimum Death Benefits. Insurance: Mathematics and Economics, 45(3), 449-458.

Windcliff H., Forsyth P. and Vetzal K. (2001). Valuation of Segregated Funds: Shout Options with Maturity Extensions. Insurance: Mathematics and Economics, 29(1), 1-21.