Chinese financial market: stock valuation from a data analysis perspective and option valuation using truncated binomial trees

Hao Li

Submitted for the degree of Doctor of Philosophy
University of Sussex
September 2019

Declaration

I hereby declare that this thesis has not been and will not be, submitted in whole or in part to another University for the award of any other degree.

Hao Li

Abstract

This thesis comprises two parts. First, we try to answer the question in a data analysis perspective, which financial factor is more relevant to the market capitalisation movements. Second, due to price boundaries imposed by the market regulators, how could we price the financial options on such markets in a mathematically rigorous manner.

In the first part of this thesis, after collecting large amount of real-world financial data for companies listed on the Chinese financial market, we carry out a data analysis and set up linear regression models between market capitalisation and various financial data including PE ratio, total earning, etc. By calculating and comparing the coefficient of determination, those regression models are ranked. We find that assets and earnings are highly correlated with market capitalisations. To extract information and reduce noise, principal component analysis technique is also used. Combining all results, a relationship between market capitalisation and other financial data is revealed.

In the second part of this thesis, based on truncated binomial trees, several option valuation models are obtained. After introducing assumptions satisfying the specific price boundaries in the Chinese financial market, we derive an option valuation model from the Cox-Ross-Rubinstein model for European call options traded on the price-bounded financial market. A closed-form solution is obtained by assuming that security trading is continuous. Using the Chinese financial market data, empirical analysis result suggests that our modified model has more explanatory power than BS model in the Chinese financial market.

Acknowledgements

My thanks go to:

- my supervisors, Dr. Qi Tang and Dr. Ding Chen.
- Prof. Bertram Düring.
- Dr. Haslifah Hasim.
- Dr. Max Jensen.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis consists of 2 separate parts. They focus on stock valuation and option valuation respectively.

The first part of this thesis concentrates on a data study of the Chinese stock market. In this part, we will look at the relationship among the extended freely available historical financial data set. Shiller (2005) collected the price-earnings (PE) ratio values for the Standard & Poor's 500 (S&P 500) index every January from 1881 to 1989 and calculated 1-year return. Then a simple linear regression reveals the relationship between those two factors. Shiller's (2005) work suggested that lower PE could indicate higher return in the future based on his analysis of the S&P 500 data. This result was widely accepted by a group of fundamental investors (Graham & Dodd 2009). Empirical analysis using different datasets (Gottwald 2012, Basu 1977) and models based on varieties of PE (Azhar et al. 2009) all confirmed this idea. Following this, we will explore Shiller's (2005) work in the Chinese market and furthermore verify the relationship between return and earnings for individual Chinese companies. We also carry out a data analysis between stock market valuation return and various financial data including PE and earnings for individual Chinese companies. We hope to draw some conclusions between market capitalisation returns and the large set of financial data available for the Chinese market.

Besides earnings, Fama et al. (1969) suggested that any new information may affect valuation in different levels. Earnings and dividend policy announcements are the most analysed factors as new information. The advantage for the Chinese stock market is that it is born after the Internet era and a lot of financial data are available freely on the trading platforms for people who make the effort to collect. So we will use the big data technique and analyse these available financial data to find what quantity gives the best linear fitting for stock price. This is an extension of Shiller's (2005) study.

In the second part, several option valuation models based on binomial tree models are derived. A financial option is a contract which offers the buyer a right to buy or sell the underlying asset at a specific strike price (Hull 2015). In this thesis, we will focus on the call option, which gives the buyer a right to buy the underlying asset. To carry out the analysis, some assumptions must be introduced. For example, Black & Scholes (1973) assumed that the underlying asset logarithmic returns are normally distributed. Normal distribution is quite simple to analyse for option valuation. However empirical studies by Corrado & Su (1996*b*) showed that there is significant

disparity between normal distribution based BS model and real-world data. For options on the US financial market, higher-order-moment distributions could improve the valuation models (Rubinstein 1998). On some special financial markets, such as the Chinese financial market, the relevant authorities imposed price movement bounds on the underlying stock price movements (SSE 2018), we therefore need to introduce some modified models.

The price bounds in the Chinese financial market are imposed by the marker regulator to curb excessive gambling behaviour. Within each phase (one day for the Chinese financial market), there are a lower bound and an upper bound on the price movements for each stock listed on the market. No trade is allowed if the price moves beyond the boundaries. So it is clear that a plain distribution is truncated by the boundaries. In our research, since the normal distribution is widely used in traditional option valuation models, we truncate normal distributions for our re-weighting model and re-distributed model respectively. Since a binomial distribution can approximate a normal distribution, following Cox et al.'s (1979) approach, our models will be derived from a truncated binomial tree firstly. Then following Hull's (2015) idea, we show that the binomial distribution results approximate some modified normal distribution results if the binomial tree is fine enough. Using the Chinese financial market data, we will compare our model with the BS model.

The price boundaries among those financial markets are not always same. For example, on Borsa Istanbul (the Turkish securities exchange), there are circuit breakers at 0.9 and 1.1 of the previous day closing price for each stock (Istanbul 2015). If a stock triggers a circuit breaker, the stock trading pauses for 7 minutes. There is no circuit breaker in the Chinese financial market. Trade orders are still accepted, if the bid or offer prices are within the boundaries (more details will be explained in Chapter 4). In the Taiwanese financial market, there are different price boundaries at 0.05 and 0.1 below or above the previous day closing price for bonds and stocks respectively (TSE 2019). The Taiwanese market also has no circuit breaker. In this thesis, we will focus on price boundary system in the Chinese financial market.

These 2 parts share the data that we explain and pre-process in Part I. As our stock valuation model (the 16-variable model in Chapter 3) requires a lot of data, we hope to find a model that requires less data. So we view the data from another perspective and introduce option valuation models in Part II. Furthermore, in Part II, we set up a portfolio of stocks and options. We believe that options require less data to evaluate and portfolios reduce more uncertainty. So we can have a model that has more explanatory power and requires less data.

Part I

Stock market data analysis

Chapter 2

Data consideration for the Chinese stock market

When using fundamental data for investment reference, a large group of investors believe that price-earnings (PE) ratios have strong relationship with future returns of a stock (Gottwald 2012). It has been demonstrated that low PE stocks will have higher returns than high PE stocks over a long time period on collective basis (Basu 1977). This phenomenon has been widely verified (Gottwald 2012).

Shiller (2005) analysed the relationship between market capitalisation change rates and PE ratio using Standard & Poor's 500 (S&P 500) index. He collected the S&P 500 index price data every January as well as the total earnings of all component companies in the previous year and set up a linear regression between the market capitalisation change rates and PE ratio. Then regression coefficient (β) and the coefficient of determination (R^2) data are collected. By doing so, he suggested that the S&P 500 index would have higher rate of change in five year period, if the PE ratio is lower in January. Shiller concluded that the market capitalisation change rates and PE ratios are negatively correlated. However, the low R^2 obtained via regression indicated high degree of market volatility and uncertainty.

In this chapter, we will follow Shiller's (2005) idea to have a brief look at the relationship between return and price-earnings ratio in the Chinese stock market.

2.1 Linear regression analysis

In data analysis, one of the most commonly used methods to determine the relationship between two variables (or one against several variables) is regression analysis.

In this chapter, we will use the coefficient of determination R^2 to measure the goodness of fit. R^2 values are normally greater than or equal to 0 and less than or equal to 1. According to Freedman et al. (2007), an R^2 of 1 indicates that all data can be interpreted perfectly by the model, while an R^2 of 0 indicates that no data can be interpreted by the model.

Of course we are aware of the deficiencies of using R^2 as the only criterion. For example, R^2 test is not able to explain biases (Ozer 1985). Hu & Bentler (1999) suggested using two-index

strategy, combining two or more testes together to judge a model. Some researchers recommend "making a residual analysis" (Barrett 1974) with R^2 test. However this makes the analysis too complex in our context. Using one index, we can easily compare two models. We could also apply other methods, such as F-test or chi-squared test (Mendenhall & Sincich 2012). These tests are also useful to check the goodness of fit for a model. But similar to the R^2 test, each has its own limitations (Hooper et al. 2008). Since we are working on large number of data, we believe that the statistical summary of large numbers of R^2 can, from one angle, compensate the statistical shortcomings of R^2 when observing a few quantities.

Our next step is to define a critical value of R^2 to separate the well-fitted and poorly-fitted data sets. As observed by Taylor (1990), analyses based on laboratory data could have higher R^2 values than analyses based on naturally observed data. In finance data analysis, various values have been used as indicator of goodness of fit. For example, Fama & French's (1993) analysis showed that R^2 for the wide-accepted CAPM is less than 0.29.

To simplify the discussions, we adopt the simple threshold as $R^2 = 0.5$. This can be interpreted as the models could explain approximately 70% of standard deviation. We define the term *large* R^2 as $R^2 > 0.5$. And P(large R^2) $\Big($i.e. P$(R^2 > 0.5)\Big)$ as how many companies (in percentage) can be explained by the model.

2.2 The data

We download data between 19th of December 1990 and 30th of June 2015 from GTA CSMAR database. The data include:

- market capitalisation,
- net assets,
- retained earnings,
- fixed assets,
- intangible assets,
- total assets,
- current liability,
- accounts receivable,
- inventory,
- operating cash flow,
- current asset,
- operating income,
- operating cost,
- net profit,

- operating profit,
- total profit,
- after-tax profit and
- stock daily close prices,

for all 2378 companies listed on Shanghai Stock Exchange and Shenzhen Stock Exchange. Additionally, we download and use the CSI 300 index data in this thesis to reflect the overall performance of Chinese financial market. This index consists of stocks traded on both Shanghai Stock Exchange and Shenzhen Stock Exchange. In this database, some data cells are filled with "NaN", because the corresponding company was not traded on that day. So we clear our database by deleting NaN cells before the IPO (initial public offering) date for the corresponding companies and filling the rest by the most recently available historical data (see later).

Then we pre-process the data before we can carry out our analyses. Firstly, we sort the data in order by ascending date.

Secondly, we generate the daily data. Due to the limited available data, to increase the sample size, we decide to use the daily data in our analyses. However, except daily market capitalisation data, most data (for example, earnings data) we downloaded is by quarter. Therefore, filling all the daily values by the most recently available historical data, we generate the daily data. In other words, the values in that quarter are all the same.

Finally, before approaching data for modelling, data normalisation procedure is required to remove the units and adjust values to a common scale. For data $D = \{d_1, \cdots, d_i, \cdots, d_n\}$, the normalised data $\tilde{D} = \{\tilde{d}_1, \cdots, \tilde{d}_i, \cdots, \tilde{d}_n\}$ are calculated as

$$\tilde{d}_i = \frac{d_i - m}{M - m}, \tag{2.1}$$

where m is the minimum and M is the maximum of D. For our ratio analyses requiring return and PE ratio data, we calculate the ratio results using the untreated data firstly, then normalise the results into the normalised return or the normalised PE. For the principal component analyses in Section 3.3, we will introduce several new variables (i.e. PC1, PC2, PC3 and PC4). These variables are first principal component results of different data sets (more details will be explained in Section 3.3). To obtain those results, we extract information from the untreated data using principal component analysis technique, before normalising the extracted result into one variable.

These data will be used to analyse the relationship between PE and return, between market capitalisation and other variables in this part. The raw data will also be used in the empirical analyses in the Part II.

2.3 The variables

In the Chinese stock market, there are large number of stock splits. For example, the number of shares outstanding of Global top e-commerce Co., Ltd. (Figure 2.1) was changed several times between 2015 and 2017. So its market capitalisation and stock price (Figure 2.2) went to different directions.

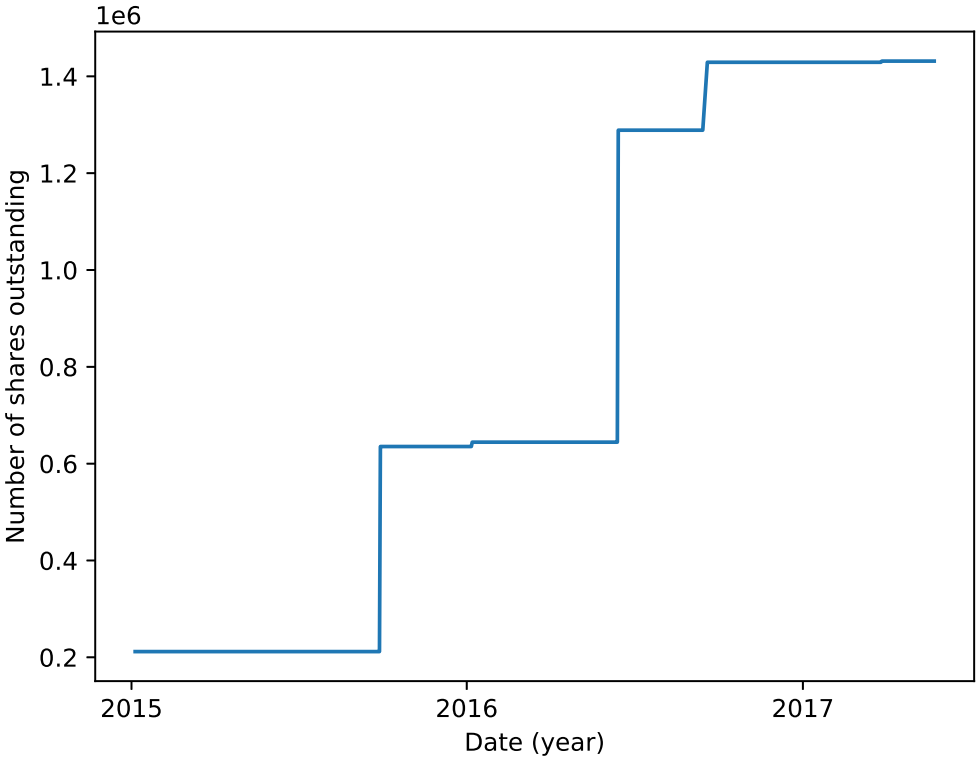


Figure 2.1: Number of shares outstanding: Global top e-commerce Co., Ltd.

Hence we believe that the stock price method used by Shiller (2005) cannot be applied to the Chinese stocks without careful adjustment. So we use the market capitalisation to represent the value of a company. Market capitalisation (CAP) is the total market value of all stocks of a company, which is defined as

$$CAP_i = \text{the price of stock } i \times \text{the number of stock } i. \tag{2.2}$$

The price in (2.2) is the latest trading price on a specific trading period. For daily data, the price is the last price on a trading day. Figure 2.3 suggested that market capitalisation and stock price are highly correlated. For over 73% companies, the correlation coefficient values between their market capitalisation and corresponding stock price data are larger than 0.707. According to Walpole et al. (2012, p. 433), it is clear that the correlation coefficient larger than 0.707 here implies a coefficient of determination larger than approximately 0.5. Recall that we adopted the threshold at $R^2 = 0.5$ (recall Section 2.1). We follow the same criterion here. Therefore we believe that market capitalisation can be used in our analyses to replace stock price in Shiller's (2005) research.

The PE ratio is calculated, using the market capitalisation and earnings data, as follows

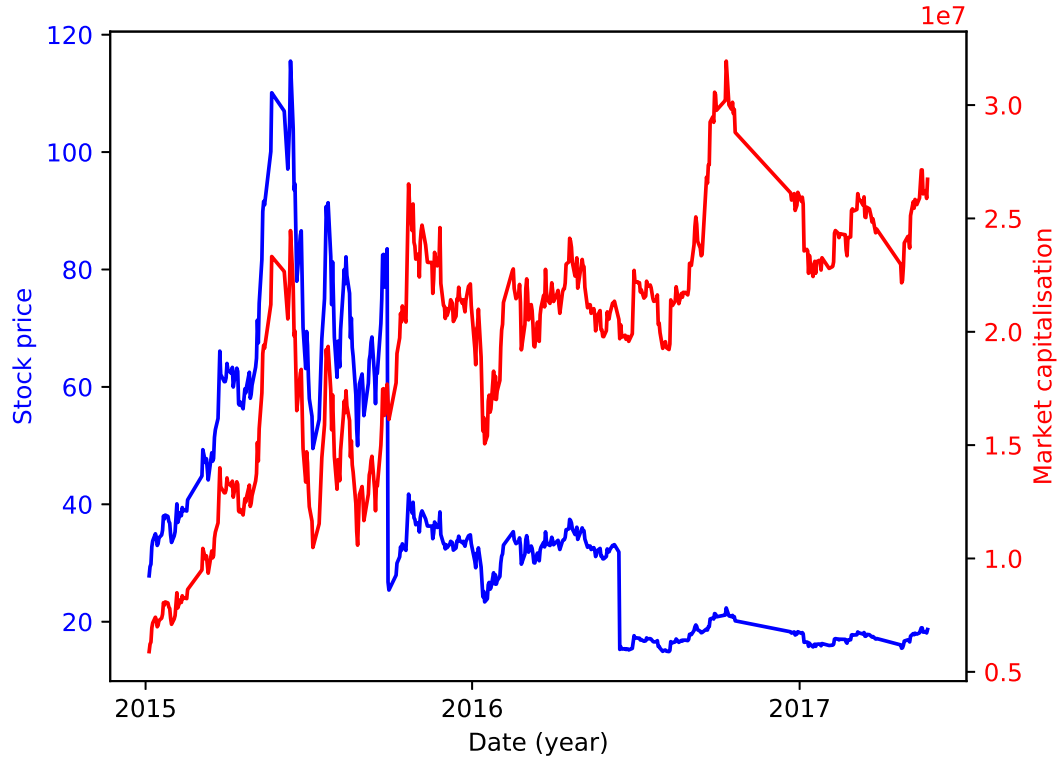


Figure 2.2: Market capitalisation and stock price: Global top e-commerce Co., Ltd.

$$\begin{aligned} PE &= \frac{stock\ price}{earnings\ per\ share} \\ &= \frac{stock\ price \times the\ number\ of\ shares}{earnings\ per\ share \times the\ number\ of\ shares} \\ &= \frac{market\ capitalisation}{total\ earnings}. \end{aligned} \tag{2.3}$$

The rate of change v of a variable λ over time period $[t, t+\delta t]$ is defined as

$$v_\lambda(t, \delta t) = \frac{\lambda_{t+\delta t} - \lambda_t}{\lambda_t}. \tag{2.4}$$

We will use this formula to calculate the *return*, which is the rate of change for market capitalisation. We carry out two analyses in this chapter. One analysis is to require holding the stock for at least 1 year. For this analysis, we select companies whose historical data existed for no less than 5 years prior to the investigation period. The other analysis is to require holding the stock for at most 10 years. For the latter analysis, we select companies whose historical data existed for no less than 15 years prior to the investigation period. Table 2.1 shows the number of companies we will use for different analyses.

Hence for the first analysis (1 year holding period) we will have 2270 companies, while for the

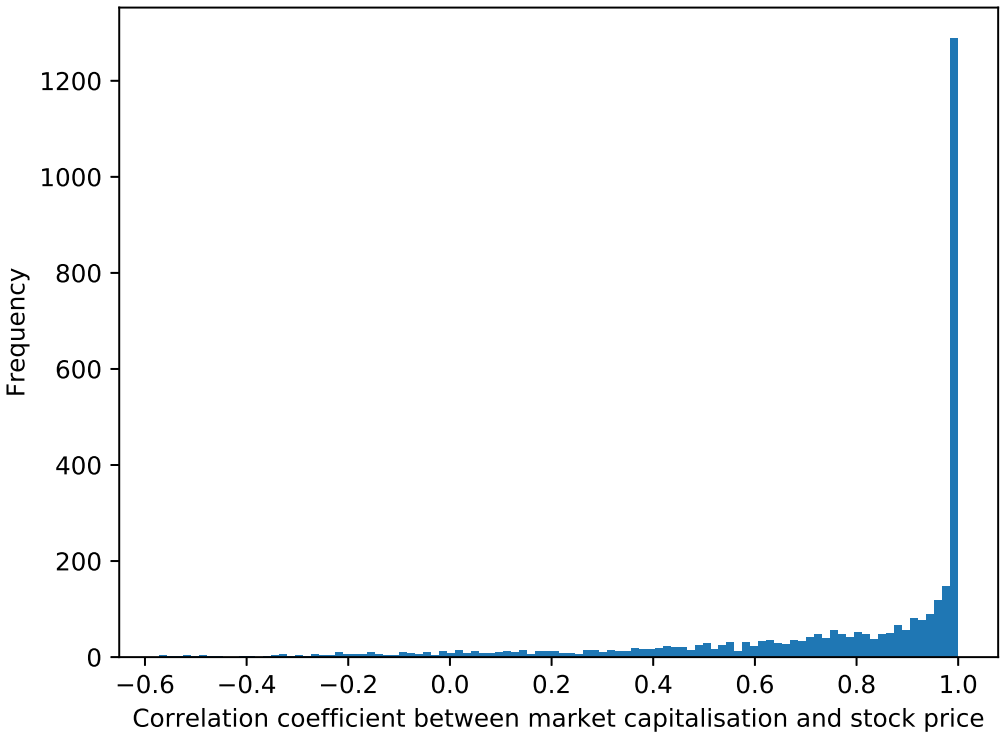


Figure 2.3: Histogram of correlation coefficient between market capitalisation and stock price.

Holding	Existence of historical data required	The num. of companies
1 year	5 years	2270
10 years	15 years	849

Table 2.1: The number of companies for different holding period.

second one (maximum 10 years holding) we will have 849 companies. We calculate the return on rolling basis. Since we only have the data before 01-07-2015 in our database, we only calculate 1-year return for data before 30-06-2014, and 10-year return before 30-06-2005.

2.4 1-year return and PE

In Shiller's (2005) study, he pointed out the following discrepancies in his analyses and conclusions:

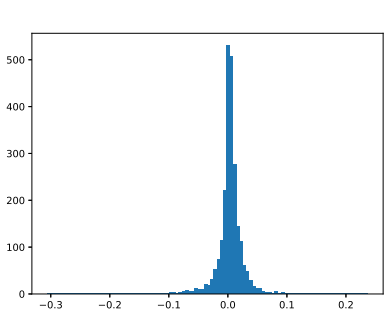
1. Only one data point on January for each year.
2. Only S&P 500 index data.
3. Low coefficient of determination (R^2) value.

To improve the study, we will try to investigate the problems from a different perspective. We will collect and analyse the rolling annualised return of daily data on every trading day from the

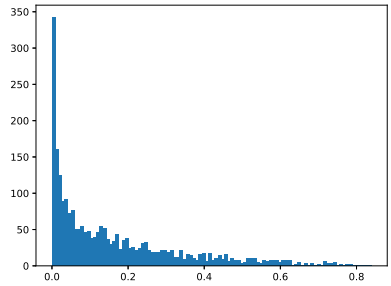
Chinese stock market. We will also collect and analyse data for all feasible companies listed on the Chinese stock market. By these, we hope to address these issues, through the use of statistical summaries.

Similar to Shiller's (2005) idea, the first analysis we will carry out is the regression between normalised market capitalisation 1-year return (R_1) and normalised price-earnings (PE) ratio.

Figure 2.4a shows the frequency of the regression coefficient for all companies, and Figure 2.4b shows the frequency of their coefficient of determination value. It is clear that most companies have very low R^2. This implies that this model has similar deficiencies as Shiller has concluded. The fact that most of the R^2 values from linear regression are below 0.5 implies that, from big data point of view, 1-year return and PE may not have close relationship. They might be a kind of random value pairs.



(a) Regression coefficient β histogram.



(b) Coefficient of determination R^2 histogram.

Figure 2.4: Regression relationship between 1-year return and PE.

Now we breakdown the above statistics into a summary table (Table 2.2). The number of companies with negative slope β value is greater than the ones with positive value. This shows that for most companies, the investors will have higher returns, if they buy the stocks when PE is lower.

	$R^2 > 0.5$	$R^2 < 0.5$	Total
$slope < 0$	9 companies	1676 companies	1685 companies
$slope > 0$	5 companies	580 companies	585 companies
Total	14 companies	2256 companies	2270 companies

Table 2.2: Breakdown of the statistics: the number of companies in different R^2 and slope value groups.

Clearly, negative β value implies that return and PE are negatively correlated for those companies. However, most companies with negative β have low R^2, meaning the match between the two variables has low quality. All these are similar to the observations made by Shiller for S&P 500 index. Over 74% (1685 out of 2270) companies have negative β and low R^2, implying that if we accept poor quality of fit, return and PE are negatively correlated as expected.

2.5 Using shorter or longer term returns

It is noticed that Figure 2.4b demonstrates that most of the companies have very low R^2 values for the one year return-PE regression over the entire life. We will try to modify our model by using shorter or longer term return to find a better fitting.

We use the same return-PE model as the previous section and breakdown the results into Figure 2.5. From this, we can see that there is no clear trend over holding period, the return-PE models don't interpret much data in the Chinese market.

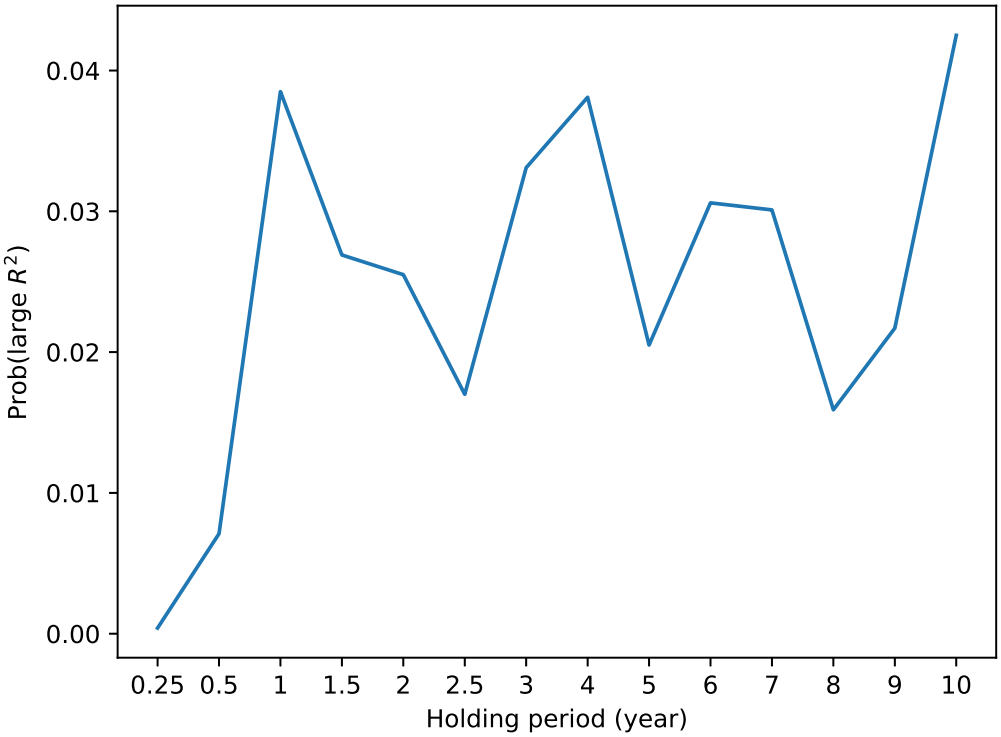


Figure 2.5: P($R^2 > 0.5$) for different holding periods, following the same idea of 1-year model.

Hence the return-PE model is not ideal and sole PE ratio is not able to describe the rise and fall of market capitalisation. Goedhart et al. (2005) reviewed a similar PE valuation model. Same to our results, the model has been demonstrated to have poor predictive power. To find better model, they proposed to introduce more factors such as operating items (operating income, operating leases, etc.). However they didn't investigate these factors further. Following their idea, we will look at broader factor sets involving more financial statement data. We derive that some factors might better describe the variations in the size of capitalisation.

Chapter 3

Data investigation of the Chinese stock market

As we discussed earlier, Shiller (2005) suggested that PE ratio could affect the future return of a stock market. This can be confirmed in US market on the index level, as Shiller (2005) has shown.

The defect of his investigation is that the regression involved demonstrated low R^2 value. We want to address this issue by using a large set of financial data for most of the companies from the Chinese stock market (whenever there is long enough history of being listed on the stock market), to seek if any financial data demonstrate better fitting properties. Besides PE, Nissim & Penman (2001) analysed more items, including common equity, operating assets and etc. They introduced several models and suggested that assets and operating liabilities have strong statistical relationship with common equity and market values. Fama et al. (1969) also reviewed stock splits, dividends and stock returns. They suggested that any new information could affect stock prices. It implies that broader data set should be introduced into research. We will focus on financial statement data as the starting point. Instead of ratio analyses, we carry out non-ratio analyses using the plain data in this chapter. So the normalised data variables we pre-processed in Section 2.2 will be used as the independent variables. We always use market capitalisation as the dependent variable for all models.

Following these ideas, we will investigate, in a systematic pattern, which quantities affect the rise and fall of market capitalisation on statistical basis. Our approach is to collect 16 widely used financial data components. We carry out regression analysis on 1 out of 16, 2 out of 16, 3 out of 16, etc. And finally, looking at how the principle components of these financial data affect the rise and fall of capitalisation.

3.1 Regression analysis

3.1.1 Single-variable models

In this section we carry out the first set of data analysis, we regress our dependent variable (normalised market capitalisation) against every single normalised financial data we selected.

Table 3.1 shows the summary for 1-variable models. We can find that the model with net assets

factor produces the best fit with our criteria. In the Chinese stock market, 33% companies' data can be interpreted by this model.

Feature	$\mu(\alpha)$	median(α)	$\mu(\beta)$	med(β)	$\mu(R^2)$	med(R^2)	P(largeR^2)
Net Assets	0.13	0.1	0.29	0.3	0.36	0.33	0.33
Total Assets	0.13	0.1	0.27	0.29	0.34	0.31	0.31
Retained Earnings	0.14	0.11	0.24	0.26	0.32	0.27	0.27
Fixed Assets	0.13	0.1	0.24	0.26	0.32	0.28	0.27
Current Asset	0.15	0.12	0.23	0.24	0.3	0.24	0.24
Current Liability	0.15	0.12	0.21	0.23	0.28	0.21	0.21
Inventory	0.15	0.13	0.2	0.2	0.27	0.2	0.21
Intangible Assets	0.16	0.14	0.17	0.17	0.26	0.2	0.19
Accounts Receivable	0.16	0.14	0.17	0.17	0.25	0.18	0.17
Net Profit	0.13	0.12	0.22	0.21	0.23	0.16	0.15
Operating Cash Flow	0.18	0.16	0.12	0.1	0.16	0.08	0.08
Operating Profit	0.17	0.15	0.15	0.13	0.16	0.1	0.07
Total Profit	0.16	0.15	0.15	0.14	0.16	0.1	0.07
Operating Income	0.2	0.18	0.1	0.08	0.15	0.09	0.05
Operating Cost	0.2	0.18	0.09	0.07	0.14	0.08	0.04
After-tax Profit	0.17	0.16	0.13	0.11	0.13	0.07	0.04

Table 3.1: Statistical breakdown for single variable models: mean of y-intercept, median of y-intercept, mean of slope, median of slope, mean of coefficient of determination, median of coefficient of determination and the P($R^2 > 0.5$) (order by descending probability).

Furthermore, 4 out of first 5 features are asset-related. It suggests that asset information strongly influence Chinese stock valuation. Below net profit, P($R^2 > 0.5$) drops to less than 0.08, which suggests that these features could explain very few data. Hence income statement items, including profit related features and operating related features, have low impacts on stock values. The mean results and median results show the same trend as P($R^2 > 0.5$) results, which confirms our conclusion.

3.1.2 Models with index data

Since the P($R^2 > 0.5$) values of all models in the last section are less than 0.5, we introduce index feature (CSI 300 data, recall Section 2.2) and try to improve the models. The index model itself (Table 3.2) could interpret 57% company data and adding another feature (net assets here, see Table 3.3) could interpret 82% company data. These results suggest that in our dataset the index is the best feature to help understanding stock values.

Feature	$\mu(\alpha)$	median(α)	$\mu(\beta)$	med(β)	$\mu(R^2)$	med(R^2)	P(largeR^2)
Index	0.06	0.05	0.68	0.73	0.54	0.57	0.57

Table 3.2: Statistical breakdown for regression relationship between market capitalisation and index.

3.1.3 Two-variable models

To select 2 features from 16 features, we use a 16-choose-2 combination C_2^{16}. Then we follow the same procedure as in last subsection. Table 3.4 shows the summary for 2-variable models.

The best model, model with net assets and retained earnings factor, can explain 44.3% companies data in the Chinese stock market. It is much better than the best 1-variable model $\Big(\text{P}(R^2 > 0.5) = 0.33\Big)$. Recalling that the best 1-variable model is the model with net assets, we find the best 2-variable model has one more factor, retained earnings, in conjunction with the net assets factor.

Feature	$\mu(\alpha)$	median(α)	$\mu(\beta)$	med(β)	$\mu(R^2)$	med(R^2)	P(largeR^2)
Net Assets	0.04	0.01	0.15	0.14	0.7	0.77	0.82
Total Assets	0.04	0.01	0.12	0.11	0.69	0.76	0.81
Retained Earnings	0.03	0.01	0.14	0.12	0.69	0.75	0.81
Fixed Assets	0.05	0.02	0.09	0.09	0.69	0.75	0.80
Current Asset	0.04	0.01	0.13	0.11	0.68	0.74	0.80
Inventory	0.05	0.02	0.09	0.07	0.66	0.72	0.77
Current Liability	0.05	0.02	0.08	0.07	0.67	0.73	0.76
Intangible Assets	0.05	0.03	0.07	0.05	0.66	0.72	0.76
Net Profit	0	-0.01	0.17	0.13	0.66	0.72	0.75
Accounts Receivable	0.05	0.03	0.07	0.05	0.65	0.71	0.73
Total Profit	0	-0.01	0.16	0.12	0.63	0.68	0.72
Operating Profit	0	-0.01	0.16	0.12	0.63	0.68	0.72
Operating Income	0.04	0.02	0.09	0.06	0.62	0.67	0.70
Operating Cost	0.05	0.02	0.08	0.05	0.62	0.66	0.70
After-tax Profit	0.01	0	0.14	0.1	0.61	0.66	0.69
Operating Cash Flow	0.05	0.03	0.06	0.03	0.61	0.65	0.68

Table 3.3: Statistical breakdown: models with index (order by descending probability).

After obtaining the 2-variable models, we add index feature into those models. Table 3.5 shows that adding index feature could interpret 44.9% more companies ($0.892 - 0.443$). This confirms our finding that index is the strongest factor.

3.1.4 Multi-variable models

After we find the best 2-variable model, we can carry out our procedure to find the best 3-variable model, by using 16-choose-3 combination C_3^{16}. Table 3.6 shows the results. To make the results more understandable, we present them in a dot plot (Figure 3.1). In Figure 3.1, each row represents a set of variables in a regression model. A blue square indicates that the feature, whose variable name is on the top of the figure, is present in the model. The numbers on the right of the figure show the P($R^2 > 0.5$) values. In this section, we will use dot plot figures to present results. We repeat the same approach for 4-variable model, 5-variable model and so on (the number of independent variables $v = 4, 5, \ldots, 16$). We summarise some best results of variable selecting into Figure 3.1 to 3.13. These results are ordered by descending P($R^2 > 0.5$).

For some models, the probability values are very close to each other. Figure 3.3 shows the first 10 best 5-variable models. The first 3 models even have exactly the same P($R^2 > 0.5$) value at 0.697. We believe that the main reason is our discrete indicators. There are totally 2378 companies in our dataset (recall Section 2.2). The value 0.697 shows that 1657 ($\approx 0.697 \times 2378$) companies can be explained by the model. It is clear that the number of companies is a non-negative integer. Hence P($R^2 > 0.5$) values are discrete. It can accept a countable number of values.

Obviously, median function is continuous. There is less chance to have exactly the same value (shown in Figure 3.3). However it is difficult to explain the underlying financial meaning of those median values. Therefore we stick with the original criteria P($R^2 > 0.5$).

From 6-variable models to 15-variable models, similarly, some models have the same value. Hence, for 5-variable or more models, we keep the best 3 models in our final selection summary figure (see Figure 3.14).

Feature 1	Feature 2	Probability
Retained Earnings	Net Assets	0.443
Fixed Assets	Net Assets	0.441
Total Assets	Net Assets	0.438
Current Asset	Net Assets	0.431
Intangible Assets	Net Assets	0.431
Current Liability	Net Assets	0.428
Total Assets	Retained Earnings	0.427
Inventory	Net Assets	0.423
Fixed Assets	Total Assets	0.417
Current Asset	Total Assets	0.416
Current Asset	Fixed Assets	0.414
Total Assets	Current Liability	0.412
Fixed Assets	Retained Earnings	0.411
Intangible Assets	Total Assets	0.411
Current Asset	Retained Earnings	0.405
Accounts Receivable	Net Assets	0.404
Intangible Assets	Retained Earnings	0.399
Inventory	Total Assets	0.396
Net Profit	Net Assets	0.396
Operating Cash Flow	Net Assets	0.394
Current Liability	Retained Earnings	0.393
Inventory	Retained Earnings	0.388
Operating Profit	Net Assets	0.388
Inventory	Fixed Assets	0.387
Total Profit	Net Assets	0.386
Accounts Receivable	Total Assets	0.383
Intangible Assets	Fixed Assets	0.383
Net Profit	Total Assets	0.381
After-tax Profit	Net Assets	0.380
Operating Cost	Net Assets	0.378
Fixed Assets	Current Liability	0.377
Operating Income	Net Assets	0.376
Operating Cash Flow	Total Assets	0.375
Net Profit	Fixed Assets	0.371
Current Asset	Current Liability	0.371
Current Asset	Intangible Assets	0.368
Total Profit	Total Assets	0.366
Operating Profit	Total Assets	0.366
Accounts Receivable	Fixed Assets	0.365
Operating Cost	Total Assets	0.363
Operating Income	Total Assets	0.362
Accounts Receivable	Retained Earnings	0.362
After-tax Profit	Total Assets	0.360
Inventory	Current Asset	0.360
Accounts Receivable	Current Asset	0.352
Operating Cash Flow	Retained Earnings	0.351
Intangible Assets	Current Liability	0.346
Net Profit	Retained Earnings	0.344
Inventory	Intangible Assets	0.343
Operating Profit	Fixed Assets	0.341
Total Profit	Fixed Assets	0.340
Operating Cash Flow	Fixed Assets	0.339
Net Profit	Current Asset	0.339
Inventory	Current Liability	0.336
Operating Profit	Retained Earnings	0.335
Total Profit	Retained Earnings	0.334
Operating Income	Retained Earnings	0.327
Operating Cash Flow	Current Asset	0.326
Operating Income	Fixed Assets	0.324
Operating Cost	Fixed Assets	0.324
Operating Cost	Retained Earnings	0.324
...	...	...

Feature 1	Feature 2	Probability
...	...	...
After-tax Profit	Retained Earnings	0.323
After-tax Profit	Fixed Assets	0.323
Net Profit	Current Liability	0.320
Accounts Receivable	Intangible Assets	0.320
Net Profit	Inventory	0.316
Accounts Receivable	Inventory	0.316
Accounts Receivable	Current Liability	0.314
Net Profit	Intangible Assets	0.313
Operating Profit	Current Asset	0.307
Total Profit	Current Asset	0.304
After-tax Profit	Current Asset	0.299
Net Profit	Accounts Receivable	0.296
Operating Cash Flow	Inventory	0.294
Operating Cost	Current Asset	0.291
Operating Income	Current Asset	0.291
Operating Cash Flow	Current Liability	0.288
Operating Profit	Current Liability	0.286
Total Profit	Current Liability	0.283
Operating Cash Flow	Intangible Assets	0.282
After-tax Profit	Current Liability	0.275
Operating Income	Current Liability	0.274
Operating Profit	Inventory	0.274
Operating Profit	Intangible Assets	0.271
Total Profit	Inventory	0.271
Operating Cost	Current Liability	0.270
Operating Cash Flow	Accounts Receivable	0.269
Total Profit	Intangible Assets	0.267
After-tax Profit	Inventory	0.263
After-tax Profit	Intangible Assets	0.255
Operating Income	Inventory	0.254
Operating Profit	Accounts Receivable	0.254
Net Profit	Operating Cash Flow	0.254
Total Profit	Accounts Receivable	0.253
Operating Cost	Inventory	0.251
Operating Income	Intangible Assets	0.251
Operating Cost	Intangible Assets	0.247
Operating Income	Accounts Receivable	0.239
After-tax Profit	Accounts Receivable	0.239
Operating Cost	Accounts Receivable	0.233
Operating Cost	Net Profit	0.208
Operating Income	Net Profit	0.204
Net Profit	After-tax Profit	0.195
Operating Profit	Net Profit	0.193
Total Profit	Net Profit	0.192
Operating Profit	Operating Cash Flow	0.154
Total Profit	Operating Cash Flow	0.153
Operating Cash Flow	After-tax Profit	0.143
Operating Income	Operating Cash Flow	0.140
Operating Cost	Operating Cash Flow	0.140
Operating Cost	Operating Profit	0.124
Operating Income	Total Profit	0.123
Operating Cost	Total Profit	0.122
Operating Income	Operating Cost	0.122
Operating Income	Operating Profit	0.122
Operating Profit	After-tax Profit	0.111
Total Profit	After-tax Profit	0.110
Operating Income	After-tax Profit	0.105
Operating Profit	Total Profit	0.104
Operating Cost	After-tax Profit	0.100

Table 3.4: P($R^2 > 0.5$): 2-variable models.

Feature 1	Feature 2	Probability
Retained Earnings	Net Assets	0.892
Fixed Assets	Net Assets	0.892
Fixed Assets	Retained Earnings	0.891
Total Assets	Net Assets	0.891
Total Assets	Retained Earnings	0.889
Current Liability	Retained Earnings	0.883
Current Asset	Total Assets	0.883
Fixed Assets	Total Assets	0.883
Current Asset	Fixed Assets	0.882
Current Asset	Retained Earnings	0.882
Current Liability	Net Assets	0.882
Inventory	Net Assets	0.879
Intangible Assets	Net Assets	0.879
Inventory	Retained Earnings	0.879
Current Asset	Net Assets	0.876
Intangible Assets	Retained Earnings	0.875
Accounts Receivable	Net Assets	0.875
Intangible Assets	Total Assets	0.874
Inventory	Fixed Assets	0.872
Current Asset	Current Liability	0.871
Total Assets	Current Liability	0.871
Net Profit	Net Assets	0.870
Operating Profit	Net Assets	0.868
Accounts Receivable	Retained Earnings	0.868
Net Profit	Fixed Assets	0.868
Net Profit	Total Assets	0.866
Inventory	Total Assets	0.866
Intangible Assets	Fixed Assets	0.866
Current Asset	Intangible Assets	0.865
Total Profit	Net Assets	0.865
Net Profit	Retained Earnings	0.865
Operating Profit	Fixed Assets	0.863
Operating Profit	Total Assets	0.863
Fixed Assets	Current Liability	0.863
Inventory	Current Asset	0.862
Total Profit	Fixed Assets	0.860
Accounts Receivable	Current Asset	0.860
Accounts Receivable	Total Assets	0.860
Total Profit	Total Assets	0.859
Net Profit	Current Asset	0.859
Operating Profit	Retained Earnings	0.858
Accounts Receivable	Fixed Assets	0.858
Total Profit	Retained Earnings	0.857
After-tax Profit	Net Assets	0.857
Inventory	Intangible Assets	0.854
After-tax Profit	Total Assets	0.853
Operating Cash Flow	Net Assets	0.853
Operating Profit	Current Asset	0.851
Net Profit	Intangible Assets	0.851
Total Profit	Current Asset	0.850
Net Profit	Inventory	0.850
Operating Cash Flow	Total Assets	0.850
Intangible Assets	Current Liability	0.849
After-tax Profit	Fixed Assets	0.849
After-tax Profit	Retained Earnings	0.849
Operating Cash Flow	Retained Earnings	0.848
Net Profit	Accounts Receivable	0.845
Inventory	Current Liability	0.845
Net Profit	Current Liability	0.844
Operating Income	Net Assets	0.843
Operating Cost	Net Assets	0.843
...	...	...

Feature 1	Feature 2	Probability
...	...	...
Operating Cost	Retained Earnings	0.842
Operating Income	Retained Earnings	0.842
Accounts Receivable	Intangible Assets	0.842
Accounts Receivable	Current Liability	0.841
Operating Cash Flow	Fixed Assets	0.840
Operating Profit	Inventory	0.838
Operating Income	Fixed Assets	0.838
Operating Profit	Intangible Assets	0.838
After-tax Profit	Current Asset	0.838
Accounts Receivable	Inventory	0.837
Total Profit	Intangible Assets	0.836
Operating Income	Total Assets	0.836
Total Profit	Inventory	0.836
Operating Cost	Total Assets	0.835
Operating Cost	Fixed Assets	0.833
Operating Profit	Current Liability	0.832
Operating Cash Flow	Current Asset	0.832
Total Profit	Current Liability	0.832
Total Profit	Accounts Receivable	0.829
Operating Cost	Current Asset	0.827
Operating Profit	Accounts Receivable	0.827
Operating Income	Current Asset	0.825
After-tax Profit	Inventory	0.824
After-tax Profit	Intangible Assets	0.823
Operating Income	Intangible Assets	0.818
After-tax Profit	Current Liability	0.816
Operating Cash Flow	Inventory	0.815
Operating Cash Flow	Current Liability	0.815
Operating Cash Flow	Intangible Assets	0.814
Operating Cost	Intangible Assets	0.814
After-tax Profit	Accounts Receivable	0.809
Net Profit	Operating Cash Flow	0.809
Operating Income	Current Liability	0.807
Operating Cost	Current Liability	0.806
Operating Income	Accounts Receivable	0.805
Operating Cost	Net Profit	0.805
Operating Income	Net Profit	0.803
Operating Income	Inventory	0.802
Operating Cost	Inventory	0.802
Operating Cash Flow	Accounts Receivable	0.801
Operating Cost	Accounts Receivable	0.800
Operating Income	Operating Cost	0.797
Operating Income	Operating Profit	0.797
Operating Cost	Total Profit	0.795
Operating Cost	Operating Profit	0.794
Operating Income	Total Profit	0.792
Operating Profit	Net Profit	0.788
Net Profit	After-tax Profit	0.787
Operating Profit	Operating Cash Flow	0.782
Total Profit	Operating Cash Flow	0.780
Total Profit	Net Profit	0.778
Operating Cost	After-tax Profit	0.777
Operating Income	After-tax Profit	0.777
Total Profit	After-tax Profit	0.770
Operating Income	Operating Cash Flow	0.768
Operating Profit	After-tax Profit	0.767
Operating Cost	Operating Cash Flow	0.765
Operating Profit	Total Profit	0.765
Operating Cash Flow	After-tax Profit	0.759

Table 3.5: P($R^2 > 0.5$): 2-variable models with index.

Feature 1	Feature 2	Feature 3	P($R^2 > 0.5$)
Net Assets	Retained Earnings	Fixed Assets	0.546
Net Assets	Retained Earnings	Intangible Assets	0.537
Net Assets	Fixed Assets	Intangible Assets	0.537
Net Assets	Retained Earnings	Current Liability	0.534
Net Assets	Current Asset	Total Assets	0.534
Net Assets	Fixed Assets	Total Assets	0.534
Net Assets	Retained Earnings	Total Assets	0.533
Net Assets	Fixed Assets	Current Asset	0.532
Net Assets	Fixed Assets	Current Liability	0.531
Net Assets	Retained Earnings	Current Asset	0.530

Table 3.6: The first 10 3-variable models in descending order of P($R^2 > 0.5$) (cf. Figure 3.1).

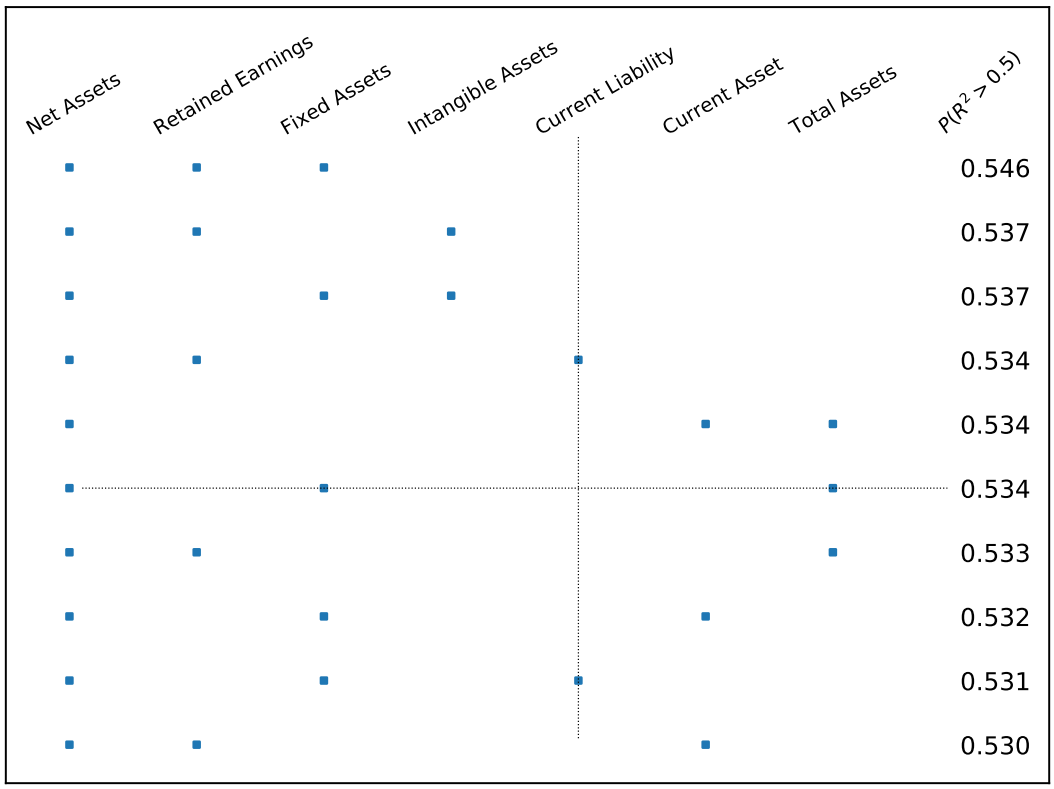


Figure 3.1: Dot plot and P($R^2 > 0.5$) values for the first 10 3-variable models (cf. Table 3.6).

Finally, we summarise the result into Figure 3.14. This figure shows the result of selecting models. The indicator integers at the left of the figure show the number of independent variables v in each model. Since some models have same v values, we omit the indicator integers below the first model which has the same v value. We suppose to use this selection result in the following sections for further model analyses.

To reduce outliers, we use confidence interval level at 0.85 (85%) for net assets beta. Then we re-calculate the beta mean values and large R^2 probability values for selected models, using confidence interval level at 85% for the net assets beta. Figure 3.15a shows the summary of P($R^2 > 0.5$). In 16-variable model, 92.3% companies data can be explained properly. Figure 3.15b shows the summary of beta mean for net assets against the number of independent variables (v) in a model. The result is clearly unsatisfactory. Different from intuitive expectation, the results are not convergent. There is big gap around $v \in [5, 10]$. Especially, the beta mean values become negative at $v = 6$.

After adding index feature (see Figure 3.16), our 8-or-more-variable models could explain almost all companies (Figure 3.16a), which is much better than the models without index feature. Although there are still some jumps for beta mean (Figure 3.16a), all values are positive.

These observations lead us to guess that there are some noises in the data. In some models, these noises enhance each other and affect the whole analysis. Hence, we need to reduce the noises. Therefore, we will use the principal component analysis technique to extract the real information

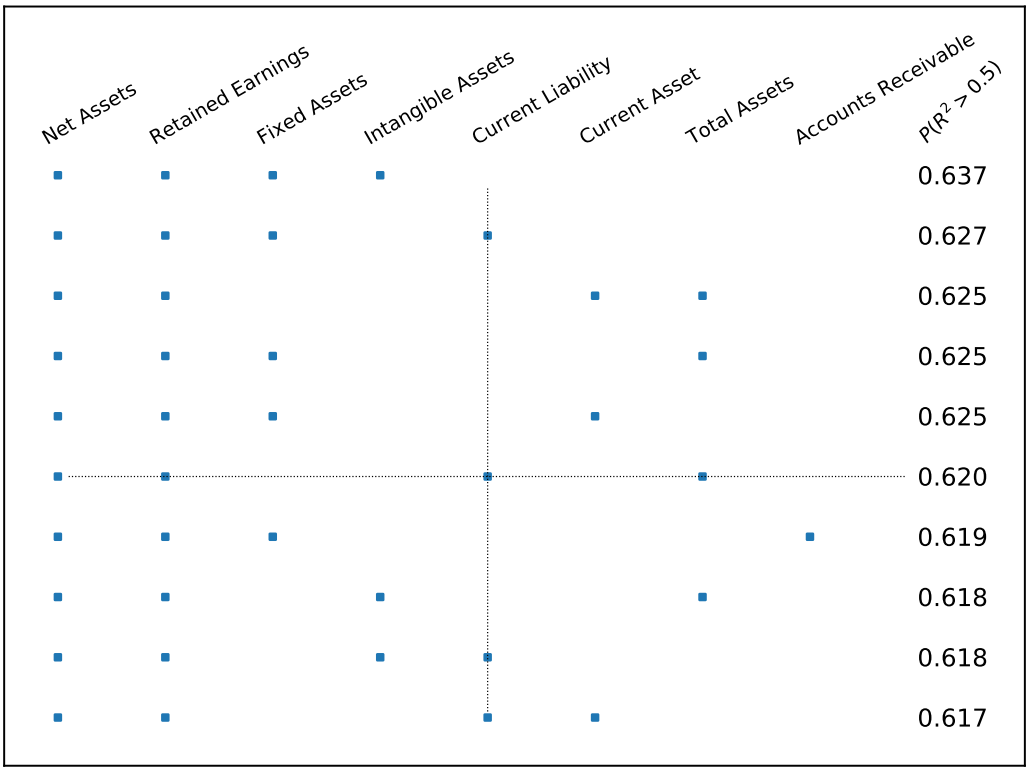


Figure 3.2: Variable selecting results and $\mathrm{P}(R^2 > 0.5)$ values for 4-variable models (best 10 results).

from the variables and reduce noises.

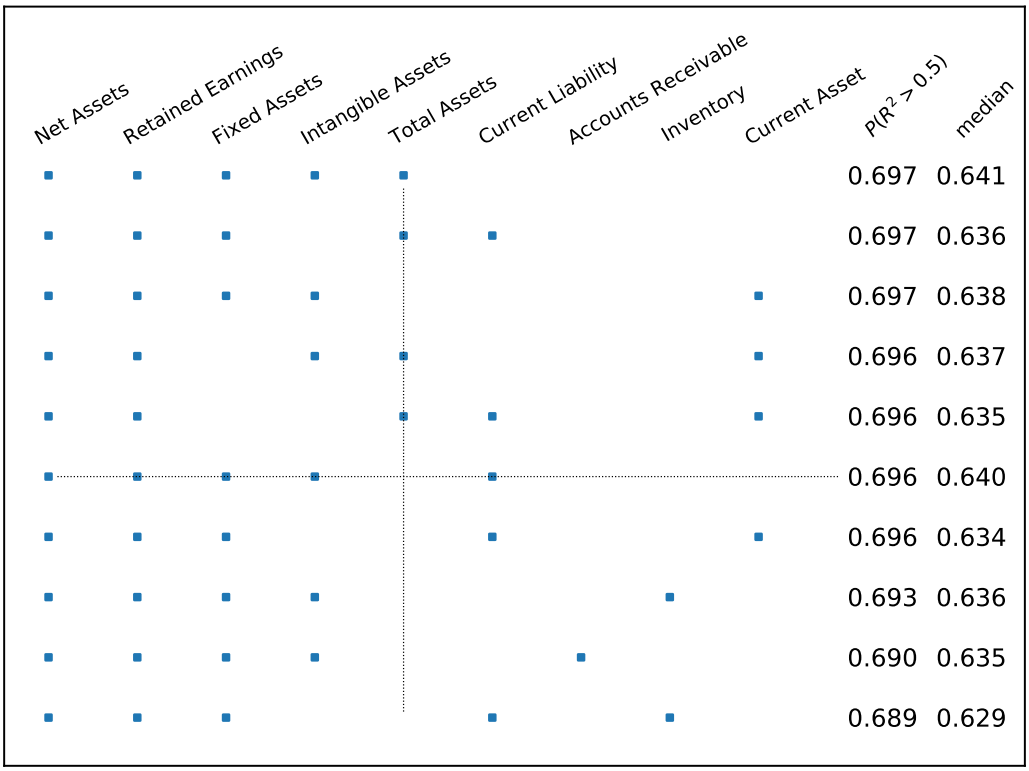


Figure 3.3: The first 10 5-variable models in descending order of P($R^2 > 0.5$) and their median values.

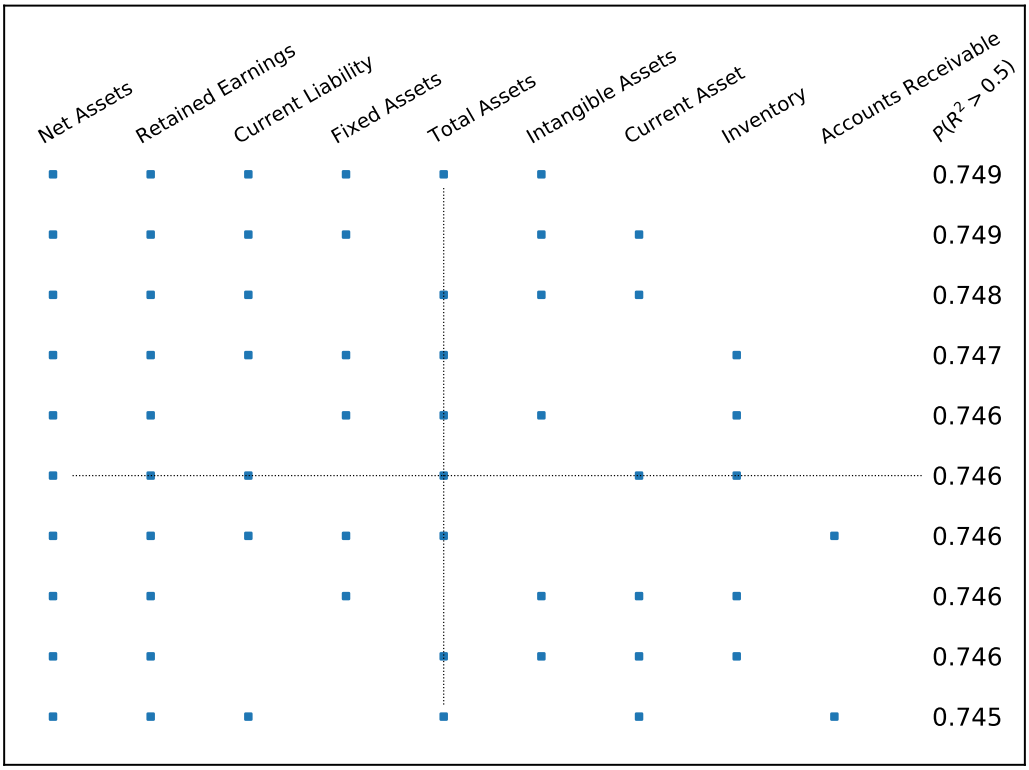


Figure 3.4: Variable selecting results and $\mathrm{P}(R^2 > 0.5)$ values for 6-variable models (best 10 results).

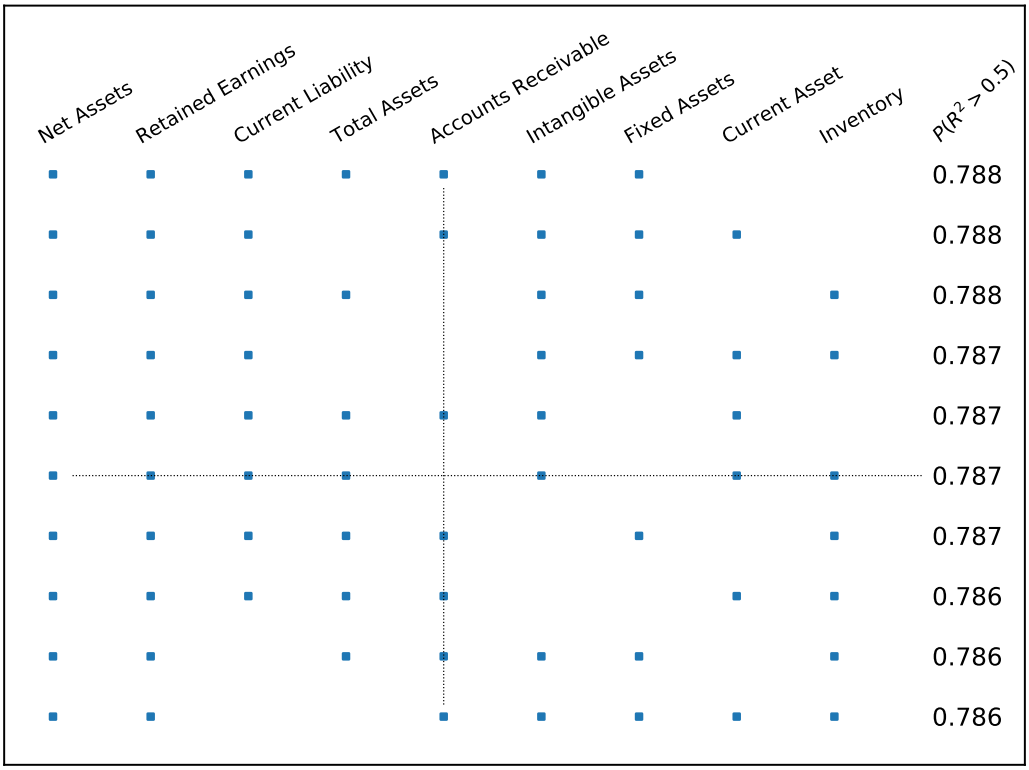


Figure 3.5: Variable selecting results and $\mathrm{P}(R^2 > 0.5)$ values for 7-variable models (best 10 results).

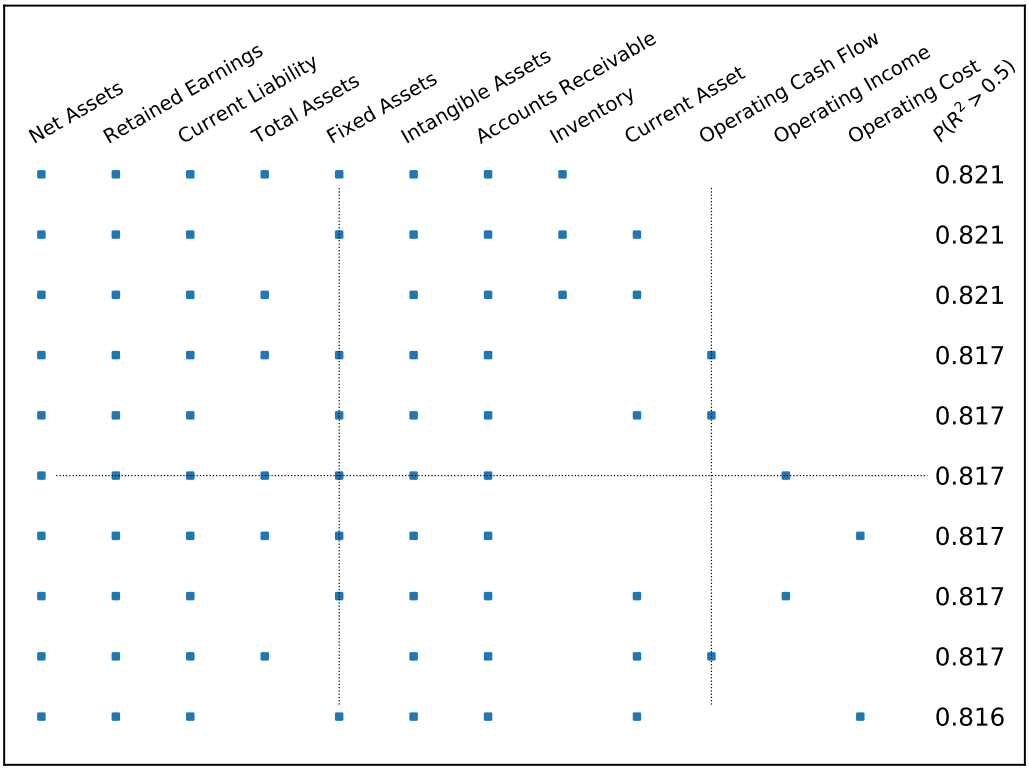


Figure 3.6: Variable selecting results and $\mathrm{P}(R^2 > 0.5)$ values for 8-variable models (best 10 results).

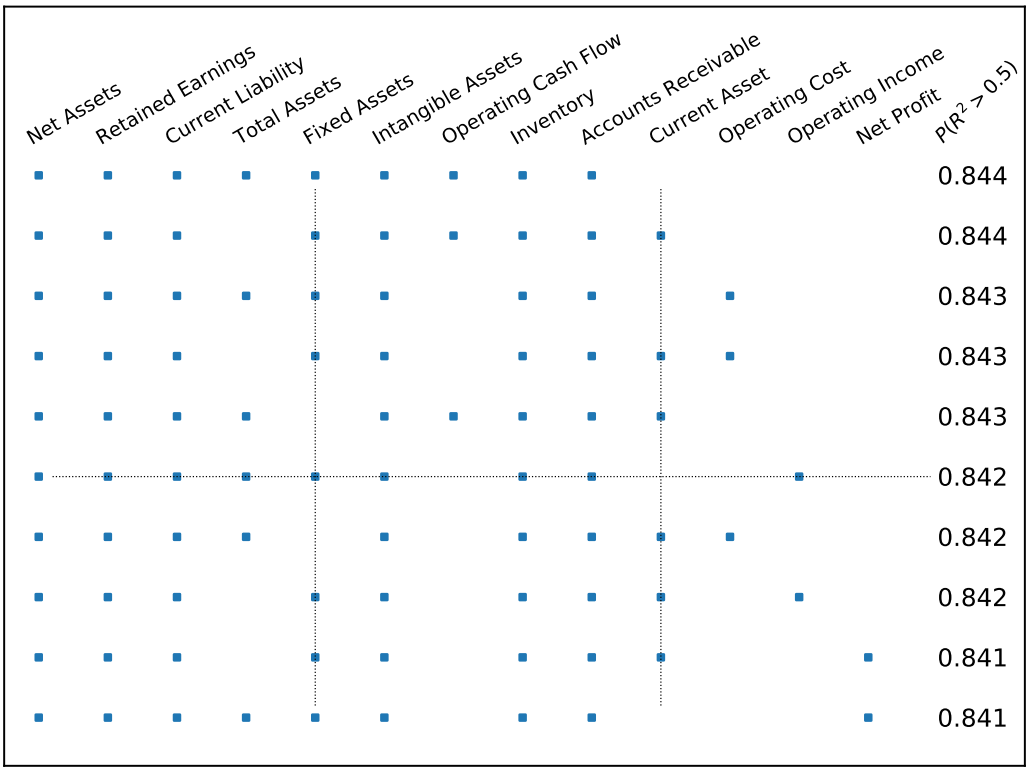


Figure 3.7: Variable selecting results and $\mathrm{P}(R^2 > 0.5)$ values for 9-variable models (best 10 results).

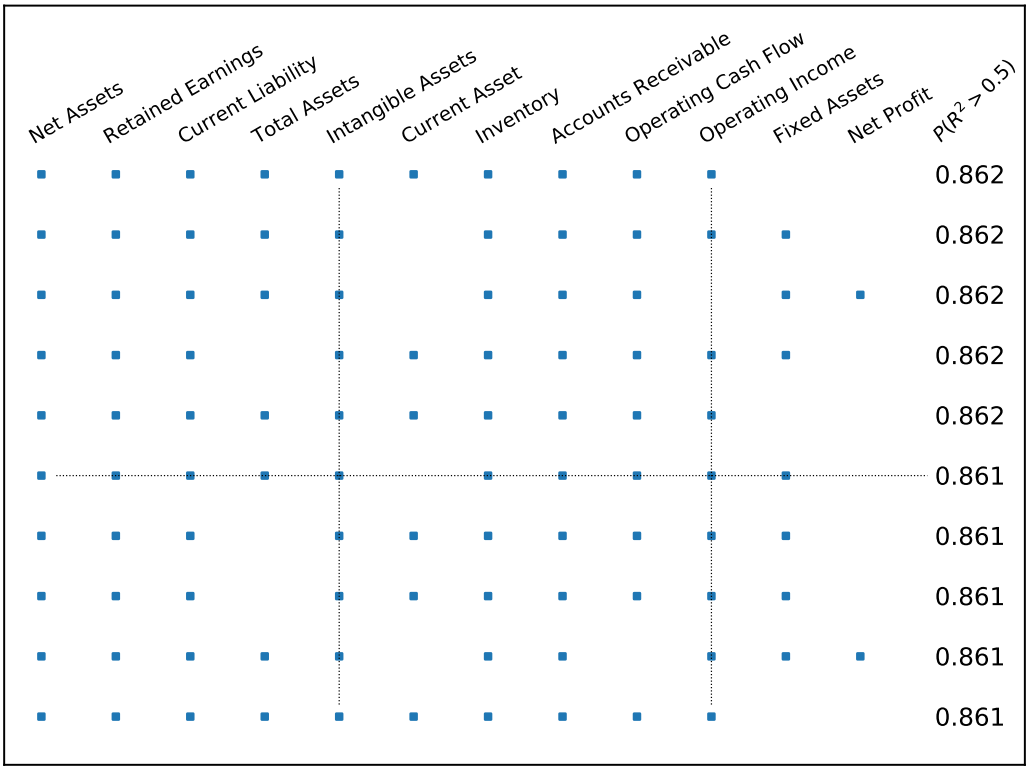


Figure 3.8: Variable selecting results and P($R^2 > 0.5$) values: 10-variable models (best 10 results).

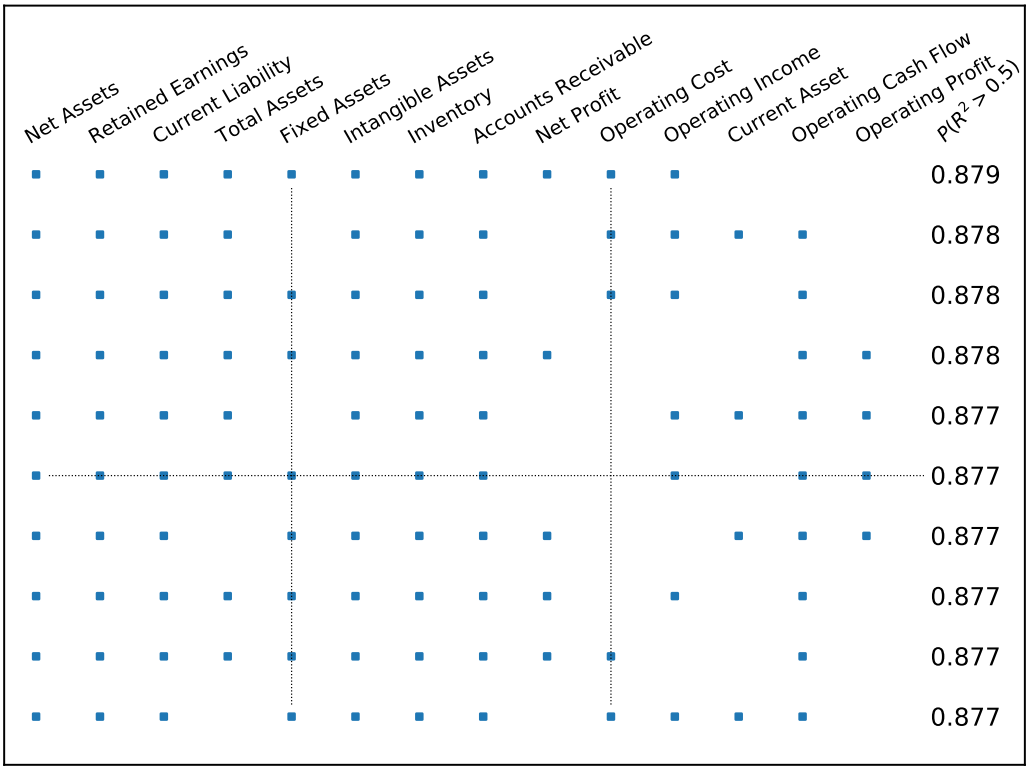


Figure 3.9: Variable selecting results and P($R^2 > 0.5$) values: 11-variable models (best 10 results).

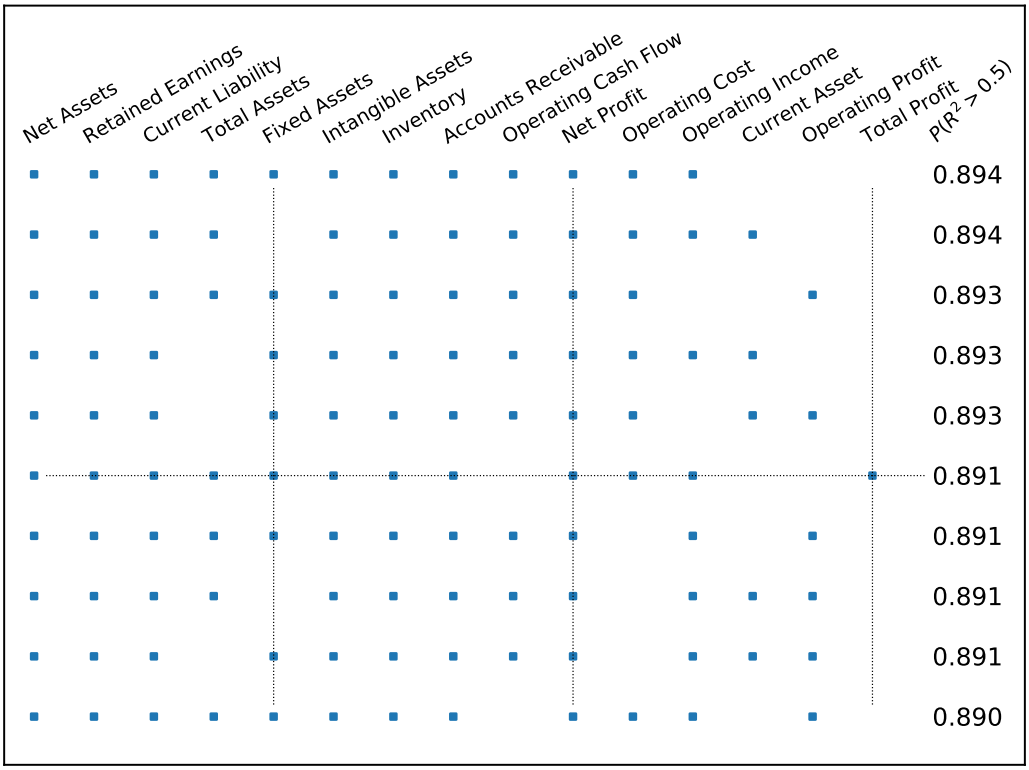


Figure 3.10: Variable selecting results and $\mathrm{P}(R^2 > 0.5)$ values: 12-variable models (best 10 results).

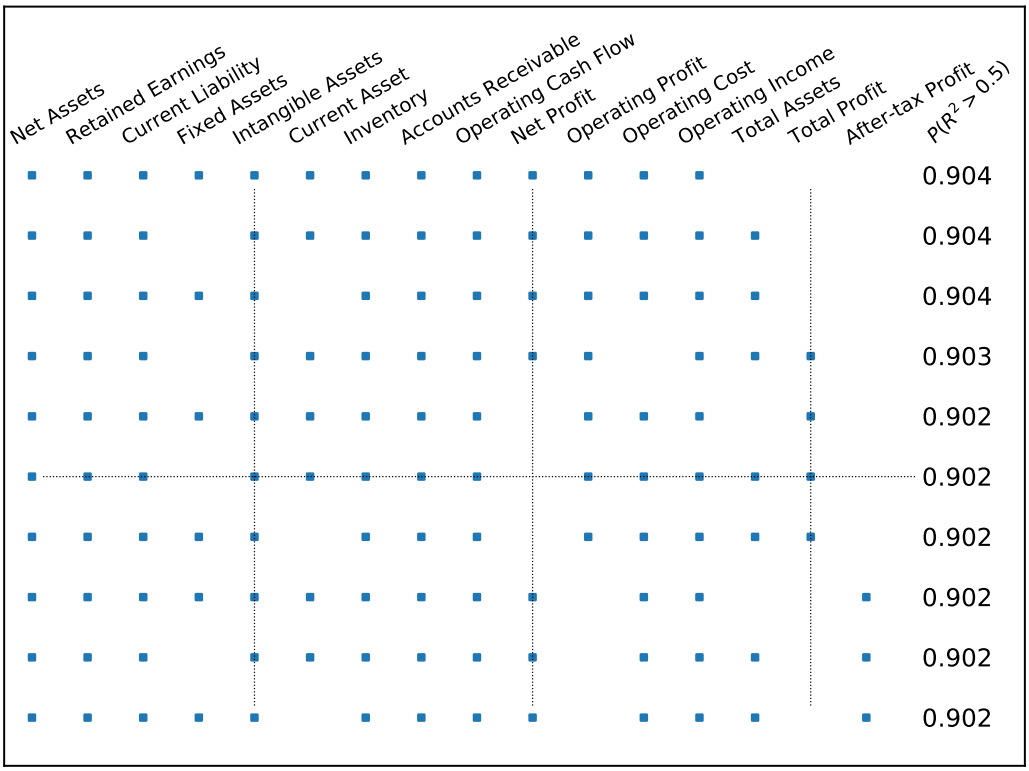


Figure 3.11: Variable selecting results and $\mathrm{P}(R^2 > 0.5)$ values: 13-variable models (best 10 results).

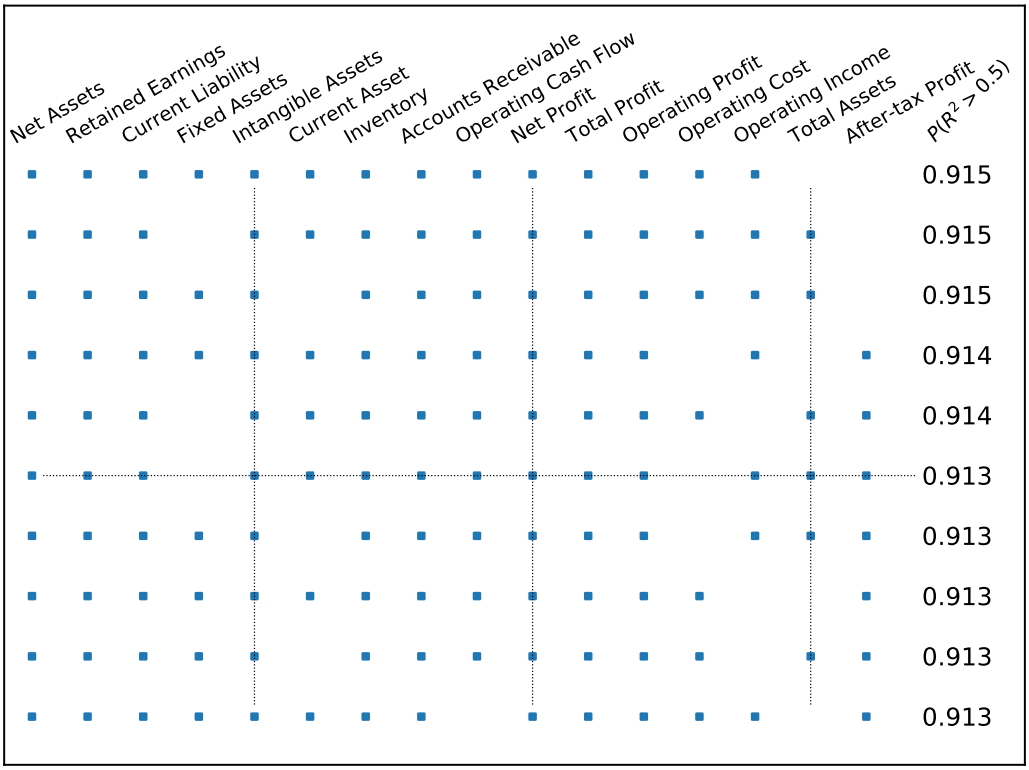


Figure 3.12: Variable selecting results and $\mathrm{P}(R^2 > 0.5)$ values: 14-variable models (best 10 results).

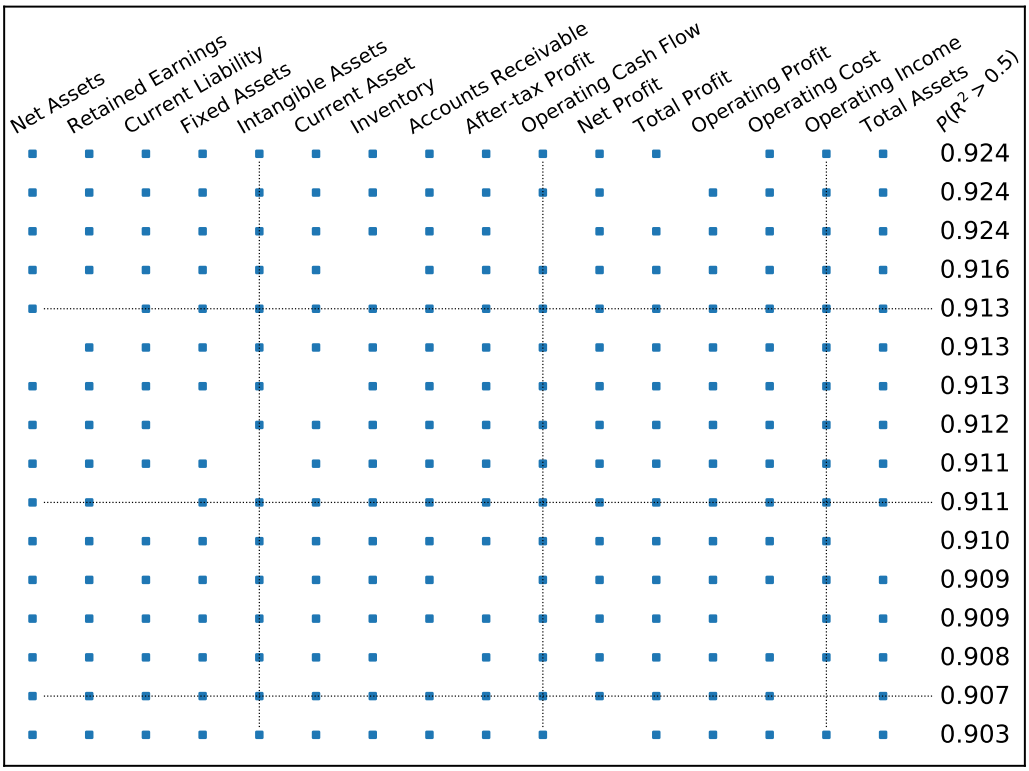


Figure 3.13: Variable selecting results and $\mathrm{P}(R^2 > 0.5)$ values: 15-variable models.

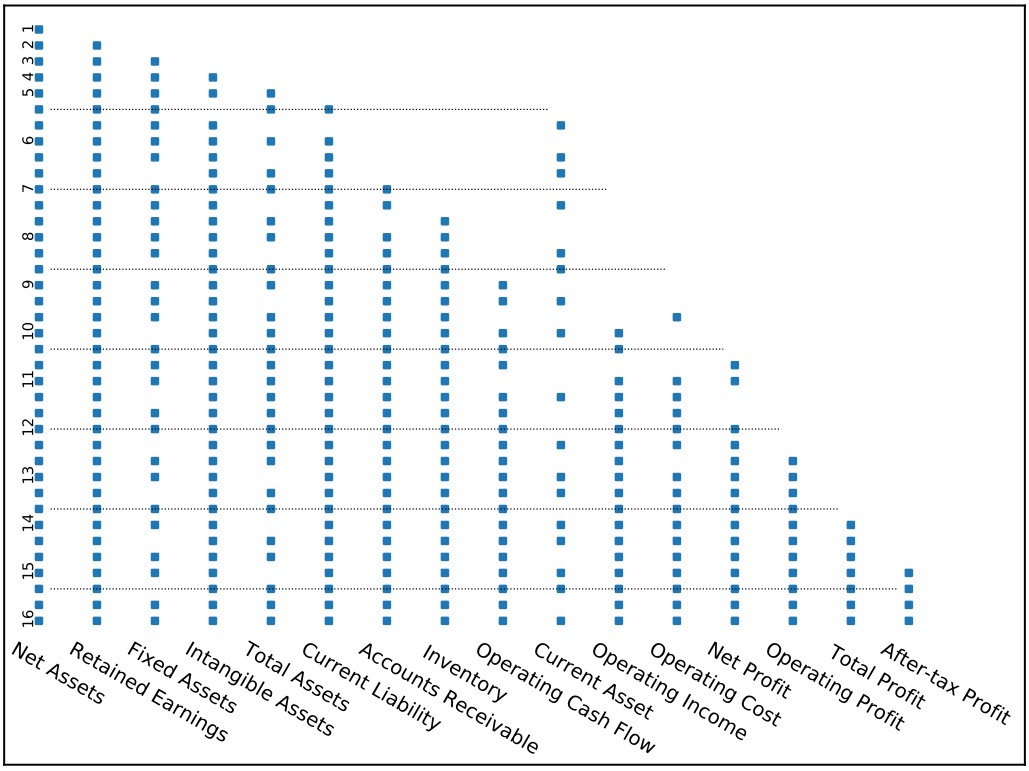
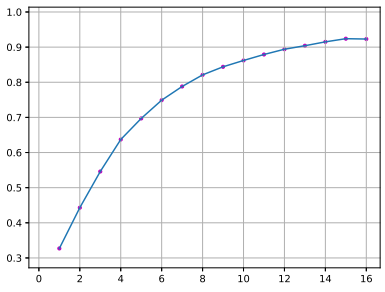
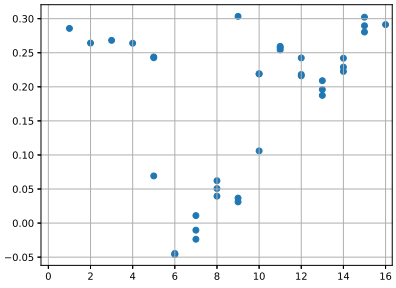


Figure 3.14: Variable selecting results for multi-variable models, order by ascending v (the number of independent variables).

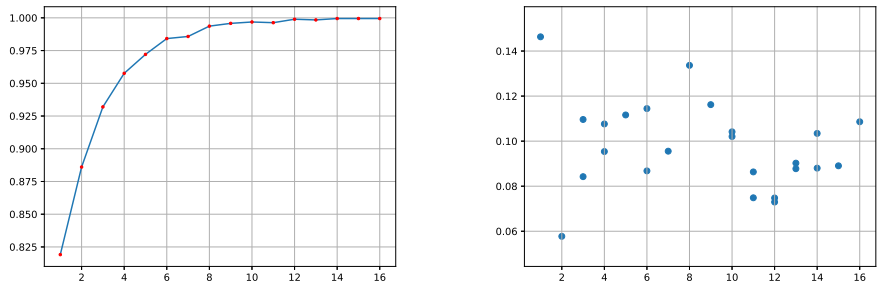


(a) P($R^2 > 0.5$) vs the number of variables.



(b) Mean of beta for net assets vs the number of variables.

Figure 3.15: Statistical breakdown for multi-variable regression models.



(a) $P(R^2 > 0.5)$ vs the number of variables: models with index.

(b) Mean of beta for net assets vs the number of variables: models with index.

Figure 3.16: Statistical breakdown for multi-variable regression models, after adding index into the models.

3.2 Results using quarterly data

The daily data we used in our analysis are generated by filling some values (recall Section 2.2). From 5-variable models (Figure 3.3) to 15-variable models (Figure 3.13), some $P(R^2 > 0.5)$ values are exactly the same. In this section we will explore this issue to see whether the generated data is the reason why some values are the same.

In Table 3.7, we present the results using the raw quarterly data. It is clear that some values are the same. Comparing to the results using the generated daily data (see Figure 3.3 to 3.13), there is no clear evidence that using quarterly data could solve the issue. Obviously the generated daily data is not the reason.

v	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*
$P(R^2 > 0.5)$	0.777	0.810	0.842	0.862	0.883	0.895	0.906	0.908	0.915	0.918	0.924
	0.770	0.810	0.842	0.862	0.883	0.895	0.906	0.908	0.915	0.918	0.924
	0.770	0.810	0.842	0.861	0.883	0.895	0.906	0.908	0.914	0.918	0.924
	0.770	0.810	0.831	0.856	0.882	0.894	0.904	0.908	0.914	0.917	0.921
	0.770	0.801	0.829	0.856	0.882	0.894	0.904	0.908	0.912	0.917	0.920
	0.767	0.801	0.829	0.856	0.882	0.894	0.904	0.907	0.912	0.917	0.919
	0.766	0.801	0.828	0.855	0.881	0.893	0.903	0.907	0.912	0.917	0.917
	0.763	0.798	0.828	0.855	0.881	0.893	0.903	0.907	0.910	0.917	0.917
	0.763	0.798	0.828	0.855	0.881	0.893	0.903	0.906	0.910	0.917	0.916
	0.762	0.798	0.828	0.855	0.880	0.892	0.903	0.906	0.910	0.915	0.916

Table 3.7: $P(R^2 > 0.5)$ values of best 10 v-variable models, using quarterly data.

3.3 Principal component analysis

In this section, we will use principal component analysis (PCA) to reduce the dimensionality of data and retain as much as possible of the data variation. We will introduce 4 variables. PC1 is the normalised first principal component of net assets, retained earnings, fixed assets, intangible assets and total assets data. PC2 is the normalised first principal component of current asset, current liability, accounts receivable and inventory data. PC3 is the normalised first principal

component of net profit, operating profit, total profit and after-tax profit; PC4: operating cash flow and operating income data.

3.3.1 PC1: net assets, retained earnings, fixed assets, intangible assets and total assets

First of all, we try to extract information from the variables of the first 5-variable model and combine the information into one variable (PC1), using principal component analysis technique. Please note that we extract information from the untreated data, before normalise the result into PC1 (recall Section 2.2). We treat this PC1 variable as a normal variable as any others. The five variables are net assets, retained earnings, fixed assets, intangible assets and total assets (recall Figure 3.3). The four of these five variables are clearly asset-based, except retained earnings. The retained earnings is an item placed on income statement, while asset items are placed on balance sheet. However, retained earnings also make positive contribution to assets. In this analysis, since we only focus on assets-liabilities dividing, to simplify the research, we also treat the retained earnings as asset. Then we remove these five variables from our datasets.

We repeat our approach for PC1 and the rest 11 variables. We select the best 1-variable model, then 2-variables model and so on, using the same method as in last section. The PC1 factor is the first one which has been selected. In 1-variable model selecting, model with PC1 has higher $\mathrm{P}(R^2 > 0.5)$ value than any other models (shown in Table 3.8). For the 2-variable models (shown in Table 3.9), the result affirms our finding.

Feature	Probability
PC1	0.297
Current Asset	0.242
Current Liability	0.209
Inventory	0.205
Accounts Receivable	0.172
Net Profit	0.152
Operating Cash Flow	0.079
Operating Profit	0.067
Total Profit	0.066
Operating Income	0.051
Operating Cost	0.042
After-tax Profit	0.038

Table 3.8: $\mathrm{P}(R^2 > 0.5)$: single variable (including PC1) models (order by descending probability).

After that, we keep those selected models and recalculate the probability value of large R^2 and the beta mean for the PC1 factor, under 85% confidence interval for PC1 beta mean. Figure 3.18a shows the $\mathrm{P}(R^2 > 0.5)$. As we expected, the probability values increase very quickly. Finally, for 12-variable model ($v = 12$), the probability values arrive at 0.860. Which means that over 86% companies can be interpreted properly by the 12-variable model. Please note, in the 12-variable model, we have used all 16 variables.

Figure 3.17 shows the variable selecting results. Figure 3.18b shows the beta mean for the PC1 factor. In 1-variable model ($v = 1$), except PC1 factor, there is no other factors. In 2-variable

Feature 1	Feature 2	Probability
PC1	Current Asset	0.418
PC1	Current Liability	0.387
PC1	Inventory	0.379
PC1	Accounts Receivable	0.354
PC1	Net Profit	0.347
PC1	Operating Cash Flow	0.338
PC1	Operating Profit	0.332
PC1	Total Profit	0.326
PC1	Operating Cost	0.323
PC1	After-tax Profit	0.323
PC1	Operating Income	0.321

Table 3.9: P($R^2 > 0.5$): 2-variable (including PC1) models (order by descending probability).

model ($v = 2$), after another factor joined, the PC1 beta mean value drops a lot. At $v = 6$, the PC1 beta mean drops again. It seems that these four factors, the second factor up to the fifth factor, can be place in a group. And from the sixth factor, the direction changes again.

Adding index feature into above models improves P($R^2 > 0.5$) results (Figure 3.19a), as we expected. However it doesn't improve the beta mean results (Figure 3.19b). So firstly, we try to solve the first gap. Then, we check the result again, to make sure both those two values can be convergent.

3.3.2 PC2: current asset, current liability, accounts receivable and inventory

Now we extract another first principal component from the further four variables shown in Figure 3.18b and combine the information into one new variable (PC2), using the same principal component analysis technique as we did before. These four variables are current asset, current liability, accounts receivable and inventory. These four variables are all short-term high-liquidity items. They could change very frequently in an operating period. Additionally, all these four items make positive contributions to assets. The 1-variable models result (Table 3.10), 2-variable models result (Table 3.11) and variable selecting result (Figure 3.20) are presented.

Feature	Probability
PC1	0.323
PC2	0.300
Net Profit	0.152
Operating Cash Flow	0.079
Operating Profit	0.067
Total Profit	0.066
Operating Income	0.051
Operating Cost	0.042
After-tax Profit	0.038

Table 3.10: P($R^2 > 0.5$): single variable (including PC1 and PC2) models (order by descending probability).

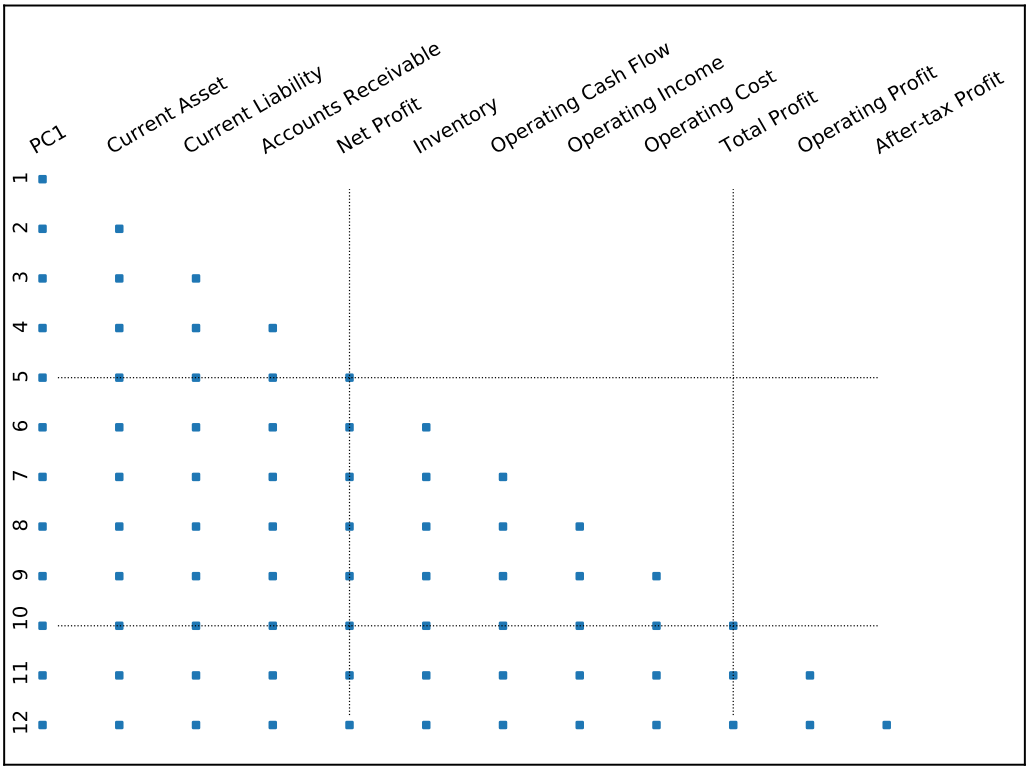
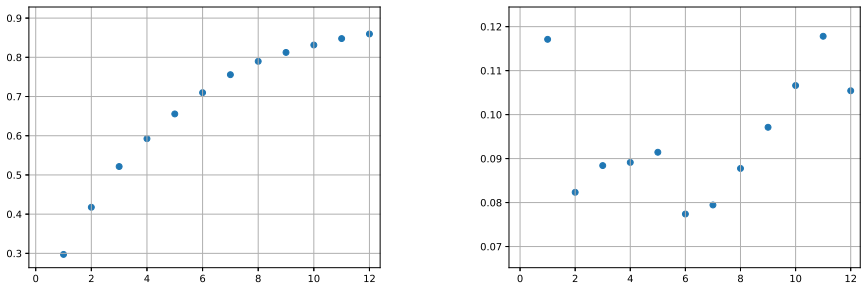


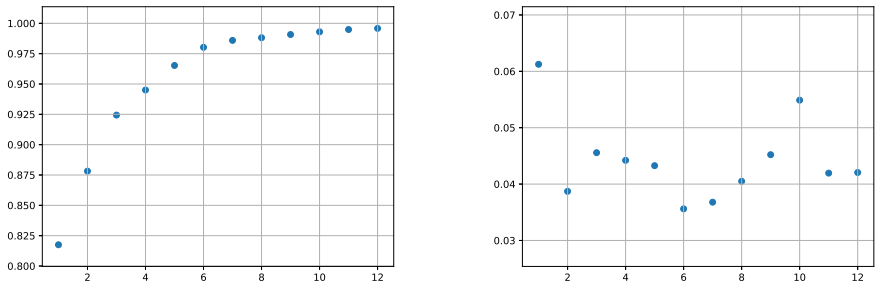
Figure 3.17: Variable selecting results after introducing PC1.



(a) P($R^2 > 0.5$) vs the number of variables, models with PC1.

(b) Beta mean for PC1 vs the number of variables.

Figure 3.18: Statistical breakdown for multi-variable regression model, after introducing PC1 factor.



(a) P($R^2 > 0.5$) vs the number of variables, models with PC1 and index.

(b) Beta mean for PC1 vs the number of variables: models with index.

Figure 3.19: Statistical breakdown for multi-variable regression models: adding index into PC1 models.

Feature 1	Feature 2	Probability
PC1	PC2	0.394
PC1	Net Profit	0.347
PC1	Operating Cash Flow	0.338
PC1	Operating Profit	0.332
PC1	Total Profit	0.326
PC1	Operating Cost	0.323
PC1	After-tax Profit	0.323
PC1	Operating Income	0.321

Table 3.11: P($R^2 > 0.5$): 2-variable (including PC1 and PC2) models (order by descending probability).

Figure 3.21a shows the P($R^2 > 0.5$). In the new 9-variable model (where all 16 variables in the datasets have been used), 74.6% companies data can be explained properly. The corresponding value in the last section is 86%. This means that less companies can be interpreted, after we introduce more PCA factors.

However, the summary of beta mean for PC1 and PC2 in Figure 3.21b shows that the result is still not convergent. There are still some gaps or drops in the figure. For PC1 mean line (the red points and line in Figure 3.21b), the drop between $v = 1$ and $v = 2$ is quite obvious. This is reasonable as we have argued that the PC2 factor is another strong factor. As a result of Table 3.1, we already expected that PC2 is stronger than any remaining individual factors. So we can accept this drop. However, for PC2, the gap between $v = 4$ and $v = 5$ is much more serious. It suggests that the 5th variable (operating income) is stronger than PC2 which contains information from four variables. Additionally the operating income feature can only interpret 5% companies' data on its own (recall Table 3.1). Hence, we decide to use principal component analysis technique again to extract information from the rest of the variables.

After adding index feature, P($R^2 > 0.5$) results are improved a lot (Figure 3.22a). But beta mean results are unchanged (Figure 3.22b).

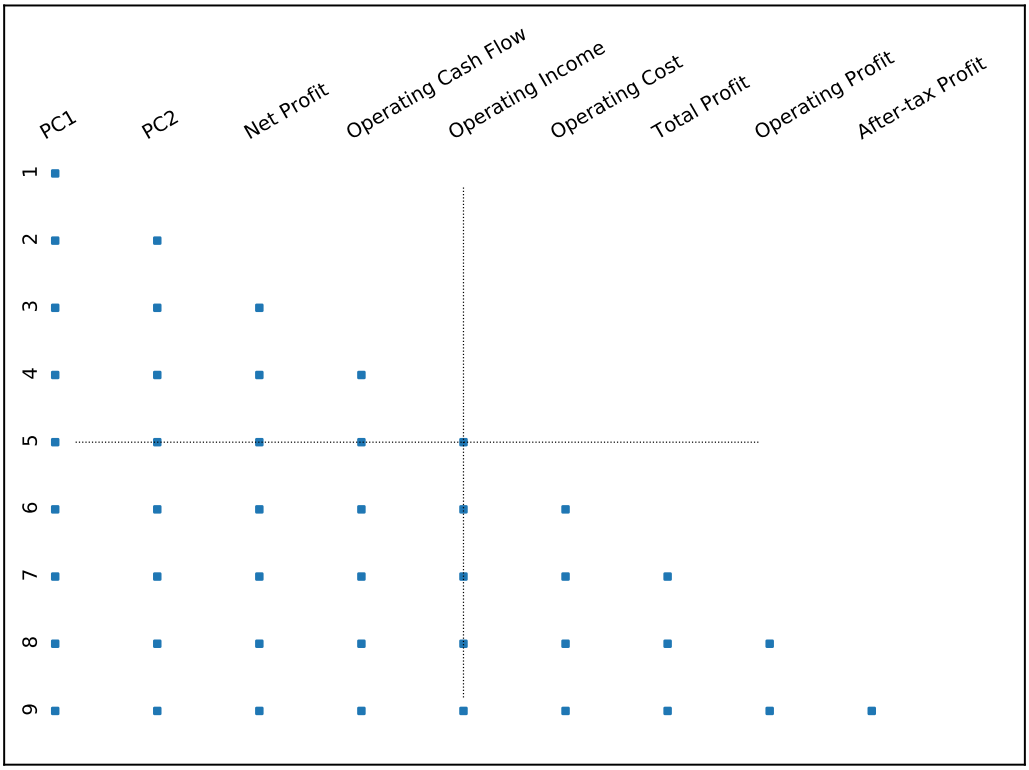
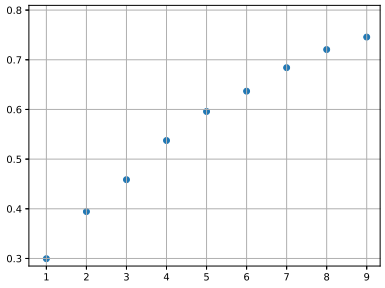
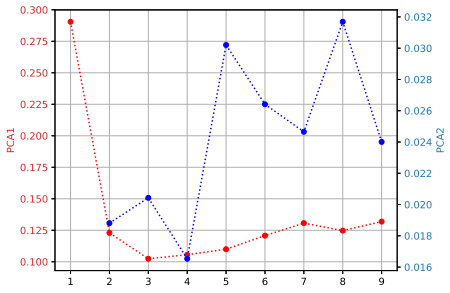


Figure 3.20: Variable selecting results after introducing PC1 and PC2.

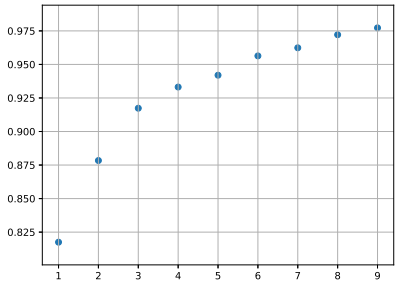


(a) P($R^2 > 0.5$) vs the number of variables, models with PC1 and PC2.

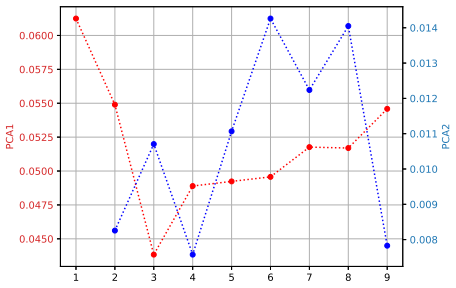


(b) Beta mean for PC1 and PC2 vs the number of variables (PC2 on secondary axis).

Figure 3.21: Statistical breakdown for multi-variable regression models: after introducing PC1 and PC2 factors.



(a) P($R^2 > 0.5$) vs the number of variables, models with PC1, PC2 and index.



(b) Beta mean for PC1 and PC2 vs the number of variables: models with index (PC2 on secondary axis).

Figure 3.22: Statistical breakdown for multi-variable regression models: adding index into PC1-PC2 models.

3.3.3 PC3: net profit, operating profit, total profit and after-tax profit; PC4: operating cash flow and operating income

Instead of the grouping method based on orders we used before, in this subsection, we use a new grouping method based on the underlying economic meaning. We separate the rest seven variables into three groups. The first one is PC3, including net profit, operating profit, total profit and after-tax profit. This group is clearly profit-related. The second group is PC4, including operating cash flow and operating income. This group is operating-related. There is only one variable left, which is operating cost. So we do not need to name it again.

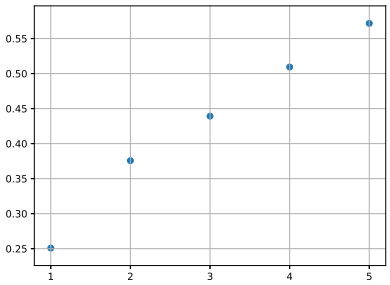
It is anticipated that, in our datasets, there is only one variable could make negative contribution to the assets, which is the operating cost. So we separate it out from other groups.

We use the same technique as we did before to calculate the P($R^2 > 0.5$) as well as beta mean for PC1 and PC2. Figure 3.23a shows the P($R^2 > 0.5$) for models with different number of variables. In 5-variable model (where all 16 variables in our datasets have been used), 57.1% companies data can be explained properly. Figure 3.23b shows the beta mean values for the two PCA factors. Except the big drop between $v = 1$ and $v = 2$ for PC1 factor, the result is acceptable, the curve is much more smooth.

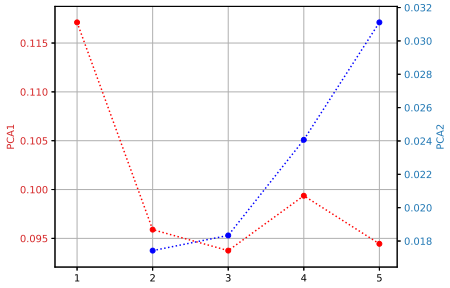
Finally, the beta mean value for operating cost is -0.166. The operating cost only appear on the 5-variable model. So we do not need to plot a figure for it. This shows that our hypothesis, operating cost makes negative contribution to assets, might be true.

So far, we have separated all variables into groups and extract the corresponding information, using principal component analysis technique, from each group.

Combining all results, we can find that PCA factors are strong. As we expected, they do contain lots of information of the corresponding companies. Additionally, the much stronger factor, index feature, could improve the goodness of fit a lot and interpret more companies. All models including PCA models are improved by adding index data into them.

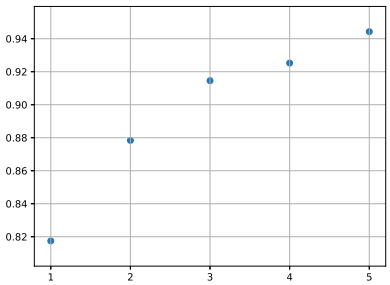


(a) P($R^2 > 0.5$) vs the number of variables, models with PC1, PC2, PC3 and PC4.

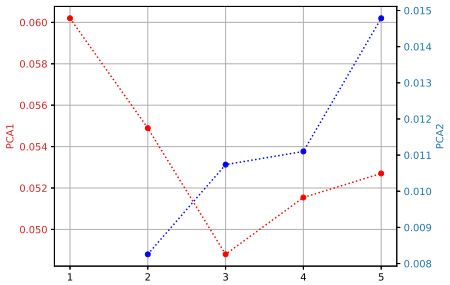


(b) Beta mean for PC1 and PC2 vs the number of variables (PC2 on secondary axis).

Figure 3.23: Statistical breakdown for multi-variable regression models: after introducing PC1, PC2, PC3 and PC4 factors.



(a) P($R^2 > 0.5$) vs the number of variables, models with PC1, PC2, PC3, PC4 and index.



(b) Beta mean for PC1 and PC2 vs the number of variables: models with index (PC2 on secondary axis).

Figure 3.24: Statistical breakdown for multi-variable regression models: adding index into 4-PCA models.

In this part, we analysed the Chinese stock market. We introduced our dataset and verified Shiller's (2005) idea using Chinese data in Chapter 2. We collected 16 quarterly variables from the Chinese financial market between 1990 and 2015. These data are used through the whole thesis. We generated and normalised the daily data and used them to probe the relationship among the variables. To remove the stock split effect, instead of stock prices, we used the market capitalisation values to generate PE values and returns. The simple return-PE model showed unsatisfactory results. Although we used a large set of data and changed the holding length, we found no clear relationship between return and PE.

Following Fama et al. (1969) and Nissim & Penman's (2001) idea, using regression analysis techniques, we carried out a further research about the relationship between market capitalisation and other variables in Chapter 3. In this chapter, we conducted a plain data analysis, instead of ratio analysis. The multi-variable model results showed that more than 92.3% of market capitalisation data can be explained by our model, by using 16 independent variables. However we decided not to accept the models, because of the convergence concern. The mean of beta values jumped several times. We guessed that there are noise data in our dataset.

To reduce the noise, we carried out analyses using quarterly data, and using principal components. We expected that quarterly data could exclude the redundant data that we introduced in the daily data generating process. However, the results showed that there is no clear improvement. Then we separated our variables into 5 groups. The principal component analyses reduced noise and extracted information from each group of datasets.

Finally, the PCA-based model is acceptable. Compared to the return-PE model $\Big($see Figure 2.5, $\mathrm{P}(R^2 > 0.5) < 0.05\Big)$ introduced in Chapter 2, more than 57.2% of market capitalisation data can be explained by this 5-variable model. Compared to the 16-variance model in Section 3.1, this model allays the convergence concern. However, it still requires 16 datasets. It is not practical. To solve this problem, we will view the data from another perspective. We will include options and set up a portfolio of stocks and options. We expect that the portfolio requires much less datasets and reduces even more uncertainty. In the next part, we will introduce our option valuation models.

Part II

Option pricing for assets with restricted underlying price movements

Chapter 4

Review of binomial option pricing models

In the following chapters, the mathematical idea of Cox-Ross-Rubinstein (CRR) binomial options pricing model (Cox, Ross & Rubinstein 1979) will be used to model a special case, where stock prices cannot rise or fall over given thresholds. For example, in the Chinese financial market, prices are capped by how much they can rise or fall. In one single trading day, the maximum gain is limited to 10% of the latest close price, while the maximum loss is limited to -10%. Which means at any time t, the daily return $r_t \in [-0.1, 0.1]$. Some market regulators may impose different boundaries. For example, on the Taiwanese bond market, the daily returns are limited within $[-0.05, 0.05]$. In this thesis, we will introduce a generalised model (the re-distributed model) for regulator-imposed price boundaries. This model is appropriate for any financial markets, if the security prices are limited in an interval proportional to a historical price. We will introduce some preparation information in Chapter 5 and 6, before introducing the re-distributed model (Chapter 7). For the empirical analysis (Section 7.6), because of data availability, we will use Chinese financial market data to examine our models.

4.1 The Black-Scholes model

In finance, a European option gives its owner the right to buy (for a call option) or sell (for a put option) the underlying security at strike price K on the expiry date T. It is well known that the value V of the European option at initial time $t_0 = 0$ is given by the following Black-Scholes partial differential equation (the Black-Scholes equation),

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS\frac{\partial V}{\partial S} - rV = 0, \tag{4.1}$$

where V is a function of time t and stock price S, r is the risk-free rate of interest and σ is the volatility of the underlying security.

To derive the option pricing model, Black & Scholes (1973) assumed the following conditions:

1. There is no transaction costs or taxes in buying or selling the stock or the option.

2. The stock pays no dividends or other distributions.
3. The option is European.
4. There are no penalties to short selling.
5. Security trading is continuous.
6. The risk-free rate of interest r is constant, it is possible to borrow or lend any amount of cash at the risk-free rate.
7. The stock price follows a geometric Brownian motion (GBM), i.e.

$$\ln\Big(\frac{S_T}{S_0}\Big) \sim \mathcal{N}\Big((\mu - \frac{\sigma^2}{2})T, \sigma^2 T\Big). \tag{4.2}$$

8. The volatility σ and drift rate μ are known and constant through time.
9. It is a risk-neutral world, i.e. the expected return of any investment is equal to the risk-free rate. Hence the discount rate is also the risk-free rate.

For European call options with final condition $C(S_T, K) = \max(S_T - K, 0)$, the solution is

Theorem 4.1 (Black-Scholes formula)**:**

$$c = S_0\Phi(d_1) - Ke^{-rT}\Phi(d_2),$$

where

$$d_1 = \frac{\ln(S_0/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}},$$

$$d_2 = \frac{\ln(S_0/K) + (r - \sigma^2/2)T}{\sigma\sqrt{T}}$$

and

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt.$$

4.2 The Cox-Ross-Rubinstein model

In this section, the Cox-Ross-Rubinstein model (CRR model) developed by Cox, Ross & Rubinstein (1979) for European call options will be reviewed. Let n be the number of steps for the binomial tree. For a time point $t = i \times T/n$, where $i \in \mathbb{Z} \cap [0, n)$, it is obvious that there exists a one-step binomial branch over time interval $[t, t+T/n]$. During the time period $[t, t+T/n]$, the continuously compounded return of a risk-free asset is rT/n, where r is the annual risk-free interest rate.

For the one-step binomial tree branch over $[t, t+T/n]$, Cox, Ross & Rubinstein (1979) assumed a stock with price S_t at time $t \in [0, T)$ can either move up to $S_t \times u$ with the upward rate $u > 1$ or move down to $S_t \times d$ with downward rate $d \in (0, 1)$. If the upward movement probability is

assumed to be $p \in (0,1)$, by risk-neutral valuation, the stock will have risk-free return over time interval $[t, t+T/n]$, i.e.

$$p(S_t \times u) + (1-p)(S_t \times d) = S_t \times e^{rT/n}. \tag{4.3}$$

Solving (4.3), the probability of upward movement becomes

$$p = \frac{e^{rT/n} - d}{u - d}. \tag{4.4}$$

Since the volatility is defined as the standard derivation of the percentage change in the stock price in the time interval $[t, t+T/n]$, and the stock prices follow a geometric Brownian motion, the volatility of the stock should be $\sigma\sqrt{T/n}$.

For a random variable X, variance of X equals $E(X^2) - E^2(X)$. So

$$\sigma^2 T/n = \Big[pu^2 + (1-p)d^2\Big] - \Big[pu + (1-p)d\Big]^2. \tag{4.5}$$

The above 3-variable system (4.4) and (4.5) has many solutions for u, d and p. To simplify the analysis, Cox, Ross & Rubinstein (1979) chose

Proposition 4.2:

$$u \times d = 1.$$

Hence combining (4.4), (4.5) and Proposition 4.2, we have

$$u = e^{\sigma\sqrt{T/n}} \tag{4.6}$$

and

$$d = e^{-\sigma\sqrt{T/n}}. \tag{4.7}$$

Therefore, (4.4) becomes

$$p = \frac{e^{rT/n} - e^{-\sigma\sqrt{T/n}}}{e^{\sigma\sqrt{T/n}} - e^{-\sigma\sqrt{T/n}}}. \tag{4.8}$$

If the initial stock price is S_0 at time $t = 0$, for a branch which has i steps going up and $n - i$ steps going down, the price of the stock at the end of this binomial tree branch is

$$S_0 u^i d^{n-i}. \tag{4.9}$$

For the binomial tree with probability going up at p, by the definition of binomial distribution probability mass function, the probability of the branches that have exactly i upwards and $n - i$ downwards steps is

$$\frac{n!}{(n-i)!i!} p^i (1-p)^{n-i}. \tag{4.10}$$

Since for an European call option, its payoff formula is $\max(S_0u^id^{n-i} - K, 0)$, the expected payoff from the European call option is

$$\sum_{i=0}^{n} \frac{n!}{(n-i)!i!} p^i(1-p)^{n-i} \max(S_0u^id^{n-i} - K, 0). \tag{4.11}$$

For an option with life T and an annual risk-free interest rate r, by risk-neutral valuation, the call option price should be (Cox et al. 1979):

Theorem 4.3:

$$c = e^{-rT} \sum_{i=0}^{n} \frac{n!}{(n-i)!i!} p^i(1-p)^{n-i} \max(S_0u^id^{n-i} - K, 0). \tag{4.12}$$

Finally, we present S_0, u, d and p in Figure 4.1. Then we define

Definition 4.1: For fixed n and p, we denote

$$\begin{aligned} b(i, n, p) :=& \frac{n!}{(n-i)!i!} p^i(1-p)^{n-i} \\ =:& b(i) \end{aligned}$$

and

Definition 4.2: For fixed n, S_0, K, u and d, we denote

$$\begin{aligned} f(i, n, S_0, K, u, d) :=& S_0u^id^{n-i} - K \\ =:& f(i). \end{aligned}$$

The above payoff formula can be rewritten as

$$\begin{aligned} & \max(S_0u^id^{n-i} - K, 0) \\ =& \max(f(i), 0) \\ :=& f_{payoff}(i). \end{aligned} \tag{4.13}$$

So we re-write (4.12) as

$$c := e^{-rT} \sum_{i=0}^{n} b(i) f_{payoff}(i). \tag{4.14}$$

4.3 A closed-form solution by Hull (2015)

In this section, we present the Hull's (2015) result which states that if n is large enough, the above CRR model converges to the BS model. In (4.11), it is clear that

Proposition 4.4:

$$\max(S_0u^id^{n-i} - K, 0) = \begin{cases} S_0u^id^{n-i} - K, & \text{if } S_0u^id^{n-i} - K > 0 \\ 0, & \text{if } S_0u^id^{n-i} - K \leqslant 0. \end{cases}$$

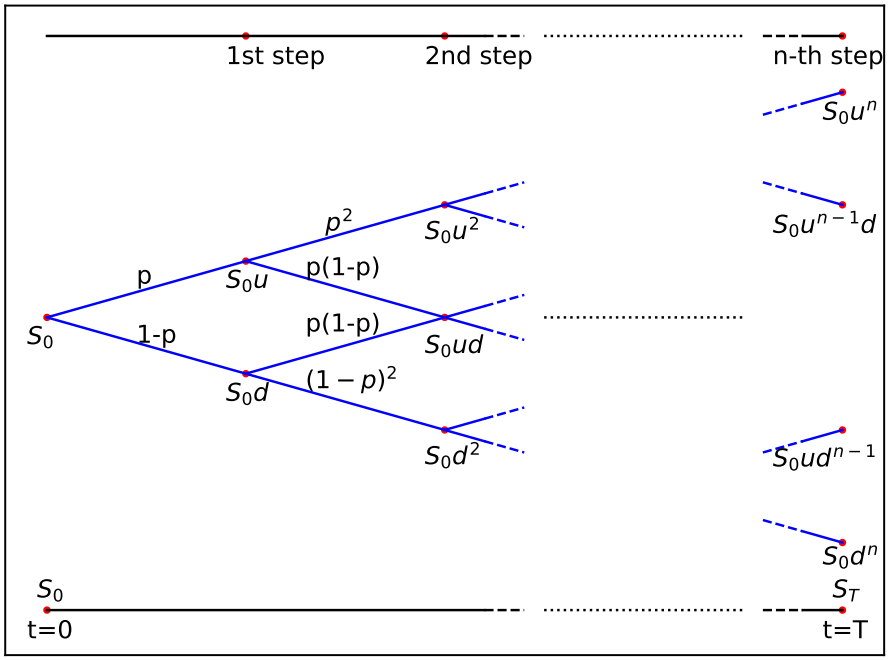


Figure 4.1: Binomial tree model.

We can find that

Lemma 4.5:

$$i \geqslant \frac{1}{2}\Big(\frac{\ln(K/S_0)}{\sigma\sqrt{T/n}} + n\Big),$$

iff

$$\max(S_0u^id^{n-i} - K, 0) = S_0u^id^{n-i} - K.$$

Proof: Since $S_0, K, u, d > 0$,

$$\begin{aligned}
S_0u^id^{n-i} &\geqslant K \\
&\iff \quad S_0/K \geqslant \quad 1/(u^id^{n-i}) \\
&\iff \ln(S_0/K) \geqslant \ln(\frac{1}{u^id^{n-i}}) \\
&\qquad\qquad = -i\ln(u) - (n-i)\ln(d) \\
&\qquad\qquad = i\Big(\ln(d) - \ln(u)\Big) - n \times \ln(d) \\
&\iff \ln(S_0/K) + n \times \ln(d) \geqslant i\Big(\ln(d) - \ln(u)\Big).
\end{aligned}$$

Since $0 < d < 1 < u$, then $\ln(d) - \ln(u) < 0$, so

$$\ln(S_0/K) + n \times \ln(d) \geqslant i\Big(\ln(d) - \ln(u)\Big)$$
$$\iff i \geqslant \frac{\ln(S_0/K) + n \times \ln(d)}{\ln(d) - \ln(u)}.$$

Since $u = e^{\sigma\sqrt{T/n}}$ and $d = e^{-\sigma\sqrt{T/n}}$, then

$$\frac{\ln(S_0/K) + n \times \ln(d)}{\ln(d) - \ln(u)} = \frac{1}{2}\Big(\frac{\ln(K/S_0)}{\sigma\sqrt{T/n}} + n\Big).$$

Therefore

$$i \geqslant \frac{1}{2}\Big(\frac{\ln(K/S_0)}{\sigma\sqrt{T/n}} + n\Big),$$

iff

$$S_0 u^i d^{n-i} \geqslant K,$$

iff

$$\max(S_0 u^i d^{n-i} - K, 0) = S_0 u^i d^{n-i} - K.$$

□

Combining above, we have

Lemma 4.6:

$$\max(S_0 u^i d^{n-i} - K, 0) = \begin{cases} S_0 u^i d^{n-i} - K, & \textit{iff } i \geqslant \alpha \\ 0, & \textit{iff } i < \alpha, \end{cases}$$

where

$$\alpha := \frac{1}{2}\Big(\frac{\ln(K/S_0)}{\sigma\sqrt{T/n}} + n\Big).$$

Then (4.12) becomes

$$c = e^{-rT} \sum_{i \geqslant \alpha}^{n} \frac{n!}{(n-i)!i!} p^i (1-p)^{n-i} (S_0 u^i d^{n-i} - K), \tag{4.15}$$

where

$$\alpha = \frac{1}{2}\Big(\frac{\ln(K/S_0)}{\sigma\sqrt{T/n}} + n\Big). \tag{4.16}$$

By defining

$$U_1 = \sum_{i \geqslant \alpha}^{n} \frac{n!}{(n-i)!i!} p^i (1-p)^{n-i} u^i d^{n-i} \tag{4.17}$$

and

$$U_2 = \sum_{i \geqslant \alpha}^{n} \frac{n!}{(n-i)!i!} p^i (1-p)^{n-i}, \tag{4.18}$$

(4.15) becomes

$$c = e^{-rT}(S_0 U_1 - K U_2). \tag{4.19}$$

We have

Lemma 4.7: *If n is large,*

$$p \approx 0.5.$$

Proof: By Taylor series, since n is large,

$$\begin{aligned}
p &= \frac{e^{\frac{rT}{n}} - e^{-\sigma\sqrt{\frac{T}{n}}}}{e^{\sigma\sqrt{\frac{T}{n}}} - e^{-\sigma\sqrt{\frac{T}{n}}}} \\
&= \frac{1 + \frac{rT}{n} + \cdots - 1 + \sigma\sqrt{\frac{T}{n}} + \cdots}{1 + \sigma\sqrt{\frac{T}{n}} + \cdots - 1 + \sigma\sqrt{\frac{T}{n}} + \cdots} \\
&\approx \frac{\frac{rT}{n} + \sigma\sqrt{\frac{T}{n}}}{2\sigma\sqrt{\frac{T}{n}}} \\
&= \frac{\frac{rT}{\sqrt{n}} + \sigma\sqrt{T}}{2\sigma\sqrt{T}} \\
&\approx \frac{\sigma\sqrt{T}}{2\sigma\sqrt{T}} \\
&= \frac{1}{2}.
\end{aligned}$$

□

It is obvious that

Corollary 4.8: *If n is large,*

$$p(1-p) \approx \frac{1}{4}.$$

Hence if n is large, p is far away from 0 and 1. So $\mathcal{B}\Big(n, p\Big)$ is approximately an $\mathcal{N}\Big(np, np(1-p)\Big)$ distribution. Therefore Hull (2015, pp. 299-300) stated that if n is large,

Lemma 4.9:

$$U_2 \approx \Phi\Big(\frac{\sigma\sqrt{T}\sqrt{n}(2p-1) + \ln(\frac{S_0}{K})}{2\sigma\sqrt{T}\sqrt{p(1-p)}}\Big),$$

where

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt$$

is the cumulative distribution function of standard normal distribution.

Proof:

$$\begin{aligned}
U_2 &= \sum_{i \geqslant \alpha}^{n} \frac{n!}{(n-i)!i!} p^i (1-p)^{n-i} \\
&\approx 1 - \Phi\Big(\frac{\alpha - np}{\sqrt{np(1-p)}}\Big) \\
&= \Phi(\frac{np - \alpha}{\sqrt{np(1-p)}}) \\
&= \Phi\left(\frac{np - \frac{1}{2}\Big(\frac{\ln(K/S_0)}{\sigma\sqrt{T/n}} + n\Big)}{\sqrt{np(1-p)}}\right) \\
&= \Phi\Big(\frac{\sigma\sqrt{T}\sqrt{n}(2p-1) + \ln(\frac{S_0}{K})}{2\sigma\sqrt{T}\sqrt{p(1-p)}}\Big).
\end{aligned}$$

□

Furthermore if n is large,

Lemma 4.10:

$$\sqrt{n}(p - \frac{1}{2}) \approx \frac{(r - \sigma^2/2)\sqrt{T}}{2\sigma}.$$

Proof: By Taylor series, since n is large,

$$\begin{aligned}
&\sqrt{n}(p - \frac{1}{2}) \\
=&\frac{\sqrt{n}}{2} \frac{2e^{\frac{rT}{n}} - e^{-\sigma\sqrt{\frac{T}{n}}} - e^{\sigma\sqrt{\frac{T}{n}}}}{e^{\sigma\sqrt{\frac{T}{n}}} - e^{-\sigma\sqrt{\frac{T}{n}}}} \\
=&\frac{\sqrt{n}}{2} \frac{2(1 + \frac{rT}{n} + \cdots) - (1 - \sigma\sqrt{\frac{T}{n}} + \frac{\sigma^2 T/n}{2!} + \cdots)}{1 + \sigma\sqrt{\frac{T}{n}} + \cdots - (1 - \sigma\sqrt{\frac{T}{n}} + \cdots)} \\
&+ \frac{\sqrt{n}}{2} \frac{-(1 + \sigma\sqrt{\frac{T}{n}} + \frac{\sigma^2 T/n}{2!} + \cdots)}{1 + \sigma\sqrt{\frac{T}{n}} + \cdots - (1 - \sigma\sqrt{\frac{T}{n}} + \cdots)} \\
\approx&\frac{\sqrt{n}}{2} \times \frac{2rT/n - \sigma^2 T/n}{2\sigma\sqrt{\frac{T}{n}}} \\
=&\frac{(r - \sigma^2/2)\sqrt{T}}{2\sigma}.
\end{aligned}$$

□

Hence Hull (2015, pp. 299-300) stated that

Lemma 4.11: *If n is large enough, for*

$$\alpha := \frac{1}{2}\Big(\frac{\ln(K/S_0)}{\sigma\sqrt{T/n}} + n\Big),$$

it is true that

$$\begin{aligned} U_2 :=& \sum_{i\geqslant\alpha}^{n} \frac{n!}{(n-i)!i!} p^i (1-p)^{n-i} \\ \approx& \Phi\Big(\frac{\ln(S_0/K) + (r-\sigma^2/2)T}{\sigma\sqrt{T}}\Big). \end{aligned}$$

Proof: Since n is large enough,

$$\begin{aligned} U_2 \approx& \Phi\Big(\frac{\sigma\sqrt{T}\sqrt{n}(2p-1) + \ln(\frac{S_0}{K})}{2\sigma\sqrt{T}\sqrt{p(1-p)}}\Big) \\ =& \Phi\left(\frac{\sigma\sqrt{T}\Big(\sqrt{n}(2p(n)-1)\Big) + \ln(\frac{S_0}{K})}{2\sigma\sqrt{T}\Big(\sqrt{p(n)(1-p(n))}\Big)}\right) \\ =& \Phi\left(\frac{\sigma\sqrt{T}\Big(\frac{(r-\sigma^2/2)\sqrt{T}}{\sigma}\Big) + \ln(\frac{S_0}{K})}{2\sigma\sqrt{T}\Big(\sqrt{\frac{1}{4}}\Big)}\right) \\ =& \Phi\Big(\frac{\ln(S_0/K) + (r-\sigma^2/2)T}{\sigma\sqrt{T}}\Big). \end{aligned}$$

□

Furthermore, it is clear that

Corollary 4.12: *If n is large enough, for any fixed x and*

$$m := \frac{1}{2}\Big(\frac{\ln(x)}{\sigma\sqrt{T/n}} + n\Big),$$

we have

$$\begin{aligned} &\sum_{i\geqslant m}^{n} \frac{n!}{(n-i)!i!} p^i (1-p)^{n-i} \\ \approx\ & \Phi\Big(\frac{\ln(1/x) + (r-\sigma^2/2)T}{\sigma\sqrt{T}}\Big). \end{aligned}$$

According to (4.17),

$$U_1 = \sum_{i \geqslant \alpha}^{n} \frac{n!}{(n-i)!i!} \Big[pu\Big]^i \Big[(1-p)d\Big]^{n-i}. \tag{4.17b}$$

By setting

$$p' := \frac{pu}{pu + (1-p)d}, \tag{4.20}$$

we have

$$1 - p' = \frac{(1-p)d}{pu + (1-p)d}. \tag{4.21}$$

According to (4.4),

$$pu + (1-p)d = e^{rT/n}, \tag{4.4b}$$

so U_1 becomes

$$\begin{aligned} U_1 &= \Big[pu + (1-p)d\Big]^n \sum_{i>\alpha}^{n} \frac{n!}{(n-i)!i!} (p')^i (1-p')^{n-i} \\ &= e^{rT} \sum_{i>\alpha}^{n} \frac{n!}{(n-i)!i!} (p')^i (1-p')^{n-i}. \end{aligned} \tag{4.22}$$

Then Hull (2015, pp. 299-300) stated that

Lemma 4.13: *If n is large enough, for*

$$\alpha := \frac{1}{2}\Big(\frac{\ln(K/S_0)}{\sigma\sqrt{T/n}} + n\Big),$$

it is true that

$$\begin{aligned} U_1 :=& \sum_{i \geqslant \alpha}^{n} \frac{n!}{(n-i)!i!} p^i (1-p)^{n-i} u^i d^{n-i} \\ \approx& e^{rT} \Phi\Big(\frac{\ln(S_0/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}\Big). \end{aligned}$$

Proof: By Taylor series and above Lemmas, since n is large enough, we have

$$\begin{aligned} U_1 \approx& e^{rT} \Phi\Big(\frac{np' - \alpha}{\sqrt{np'(1-p')}}\Big) \\ =& e^{rT} \Phi\Big(\frac{\sigma\sqrt{T}\sqrt{n}(2p'-1) + \ln(\frac{S_0}{K})}{2\sigma\sqrt{T}\sqrt{p'(1-p')}}\Big) \\ =& e^{rT} \Phi\Big(\frac{\ln(S_0/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}\Big). \end{aligned}$$

□

Furthermore, it is clear that

Corollary 4.14: *If n is large enough, for any fixed x and*

$$m := \frac{1}{2}\Big(\frac{\ln(x)}{\sigma\sqrt{T/n}} + n\Big),$$

we have

$$\begin{aligned}&\sum_{i\geqslant m}^{n} \frac{n!}{(n-i)!i!} p^i (1-p)^{n-i} u^i d^{n-i} \\ &\approx e^{rT}\Phi\Big(\frac{\ln(1/x) + (r+\sigma^2/2)T}{\sigma\sqrt{T}}\Big).\end{aligned}$$

Finally, we can put all results together. A European call option has value

$$\begin{aligned}c =& e^{-rT}(S_0 U_1 - K U_2) \qquad (4.23)\\ \longrightarrow& e^{-rT}\Bigg[S_0 e^{rT}\Phi\Big(\frac{\ln(S_0/K) + (r+\sigma^2/2)T}{\sigma\sqrt{T}}\Big)\\ &- K\Phi\Big(\frac{\ln(S_0/K) + (r-\sigma^2/2)T}{\sigma\sqrt{T}}\Big)\Bigg],\end{aligned}$$

as $n \longrightarrow \infty$. By rearranging the above into the standard form, the CRR model converges to the BS model, if the number of steps in the binomial tree is large enough.

4.4 Option pricing under truncated distributions

As price boundaries are imposed in some financial markets, to evaluate the options on such markets, an intuitive method is to introduce a doubly truncated normal distribution (Amemiya 1973). Under a normal distribution $\mathcal{N}(\mu, \sigma^2)$, the probability density function of doubly truncated normal distribution is defined as (Johnson et al. 1994):

Definition 4.3:

$$\varphi_{truncated\ normal}(x) := \begin{cases} 0, & \text{if } x < a \\ \dfrac{1}{\sigma}\dfrac{\varphi\left(\dfrac{x-\mu}{\sigma}\right)}{\Phi\left(\dfrac{b-\mu}{\sigma}\right) - \Phi\left(\dfrac{a-\mu}{\sigma}\right)}, & \text{if } x \in [a,b] \\ 0, & \text{if } x > b, \end{cases}$$

where the boundaries a and b satisfy $-\infty < a < b < \infty$,

$$\varphi(x) := \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

is the probability density function of standard normal distribution and

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$$

is the cumulative distribution function of standard normal distribution.

So now we can assume that the logarithmic returns of stocks in these markets follow a truncated normal distribution defined above (Friedmann & Sanddorf-Köhle 2007). Following the similar method to the Black & Scholes's (1973) approach, Lin et al. (2015) and Zhu & He (2018) demonstrated some examples of option pricing models under truncated normal distributions. Their approaches rely on the Black-Scholes partial differential equation (4.1). In Chapter 6, we will use a truncated binomial tree to approximate the truncated normal distribution, and derive an option pricing model.

However, the stocks returns usually behave asymmetric and non-mesokurtic (Jarrow & Rudd 1982). One normal or truncated normal distribution is not enough to explain the returns. Different from a truncated normal distribution, another simple approach is to combine several distributions together. For example, Ki et al. (2005) took a linear combination of 2 normal distributions and presented a extended normal distribution model. Empirical results showed that the extended normal distribution has better performance to explains skewness and kurtosis than plain normal distributions (Ki et al. 2005).

However, normal distribution is a bell-shaped distribution. Friedmann & Sanddorf-Köhle (2007) believed that in the price-bounded markets, probabilities are clustered near the boundaries. Furthermore, Arak & Cook (1997) suggested the magnet effect of price boundaries. The price tends to move towards the boundary as it is close enough to a price boundary. The boundary itself seems like a magnet which has a powerful attraction. Using high-frequency data, Cho et al. (2003) proposed an autoregressive model to empirically verify the magnet effect. By analysing the coefficient of dummy variables related to the boundaries, Cho et al. (2003) concluded that if the price is close to the boundaries (within 3%), there is a tendency to accelerate towards the boundaries.

To capture the magnet effect, Friedmann & Sanddorf-Köhle (2007) introduced a mixed beta distribution model, by combine 2 beta distributions. The beta distribution family already contains J-shaped distribution and U-shaped distribution. Furthermore, by combining 2 beta distributions, the mixture family also accepts W-shaped distribution. Moreover, beta distribution itself is supported on a bounded interval. Hence the mixed beta distribution is a good alternative to truncated normal distributions.

To transform those distributions into some discrete frameworks, such as binomial trees, beta distributions require more complicated techniques than the normal distributions that we are familiar with (Ki et al. 2005). In this thesis, we will focus on normal distributions and binomial distributions. To capture the magnet effect, we will use different techniques, instead of the beta distributions.

Kodres (1993) suggested using a censored distribution model. The idea is that if the consensus price is between the price boundaries, the observed price equals the consensus price. And if the consensus price is beyond a boundary, the observed price equals the boundary value. Moreover, we assume that the consensus price follows a geometric Brownian motion. In this thesis, we will follow

the idea from the mixed beta distribution models and shift probability mass to the boundaries. To focus on the normal distributions, the probability mass is clustered at the boundaries. Hence, following Kodres's (1993) idea, we introduce another W-shaped distribution. This distribution will be transformed into discrete binomial trees in Chapter 7. We will introduce a re-distributed model under this distribution.

Chapter 5

A truncated model

In Chapter 4, the standard option pricing model introduced by Black & Scholes (1973) and Cox, Ross & Rubinstein (1979) was recalled. That model serves as a useful reference in the study of the option pricing models in most cases. However, some markets have regulator imposed regulations. In this chapter, we will evaluate call options on the price-bounded financial markets. Same as standard European call options, these options may be exercised only at the expiration time T. However, the payoff formula for these options is defined as

$$\begin{cases} 0, & \text{if } S_t \geqslant S_0 L_u \\ \max(S_t - K, 0), & \text{if } S_t \in (S_0 L_d, S_0 L_u) \\ 0, & \text{if } S_t \leqslant S_0 L_d, \end{cases} \tag{5.1}$$

where $S_0 L_d$ is the lower price bound and $S_0 L_u$ is the upper price bound for the underlying stocks (more details will be explained in Section 5.2 and 5.3). We assume that $0 < L_d < 1 < L_u < \infty$ and those values are given at time $t = 0$. Hence, there are price bounds for S_T. If S_T goes beyond the bounds, option payoff becomes 0. In this chapter, we construct an absorption-vanishing process at the imposed bounds in returns. We assume that the price follows a geometric Brownian motion in the interior. So it is the same as the existing model. If the price S_T hits or exceeds the bounds at time $t = T$, the option payoff vanishes (Skorokhod 1961). This hypothetical model introduces some mathematical preparations. Some lemmas derived in this chapter will be used later. Finally, we can present our boundaries as the blue lines in Figure 5.1. In Figure 5.1, the horizontal axis represents time t and the vertical axis represents security price S.

5.1 Description and preliminary steps to adapt CRR to our truncated model

In our binomial tree, there are a total of n steps. Among these steps, we assume that there are i steps where price moves up and j steps where price moves down. Hence we have $i, j \in \mathbb{Z}^+ \cup \{0\}$, $i + j = n$. So it is obvious that $i \in \mathbb{Z} \cap [0, n]$. In this chapter, we are going to introduce β and γ for the upper and lower price bounds in our truncated model respectively. We will prove that they are

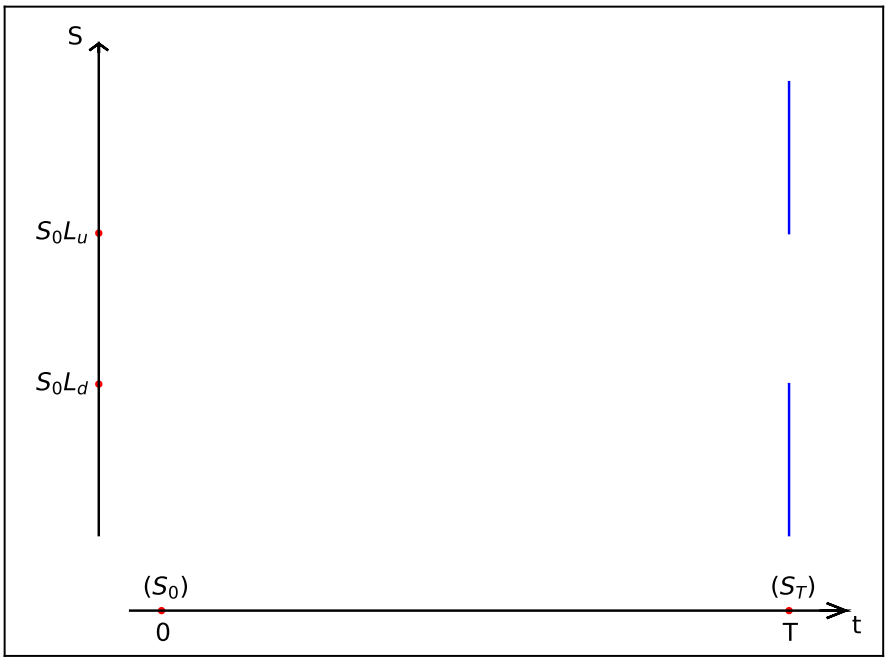


Figure 5.1: The boundaries in our model.

$$\beta = \frac{1}{2}\Big(\frac{\ln(L_u)}{\sigma\sqrt{T/n}} + n\Big) \tag{5.2}$$

and

$$\gamma = \frac{1}{2}\Big(\frac{\ln(L_d)}{\sigma\sqrt{T/n}} + n\Big) \tag{5.3}$$

respectively. Additionally, we introduced (see (4.16))

$$\alpha = \frac{1}{2}\Big(\frac{\ln(K/S_0)}{\sigma\sqrt{T/n}} + n\Big). \tag{5.4}$$

All these numbers will play a role in restricting the range of i in our truncated model. According to Proposition 4.4 and Lemma 4.5, we must have

$$i \geqslant \alpha \tag{5.5}$$

for strike price K, if payoff is positive. In other words, the binomial tree branches must contain at least α steps going upward. It is clear that for given K, S_0, σ and T, when n is sufficiently large, α satisfies $0 \leqslant \alpha \leqslant n$. We are going to limit our discussion to such range, specifically we have

Lemma 5.1: *For any given K, S_0, σ and T, to ensure $0 \leqslant \alpha \leqslant n$, n must satisfy*

$$\sqrt{n} \geqslant \Big|\frac{\ln(K/S_0)}{\sigma\sqrt{T}}\Big|.$$

Proof:

$$\begin{aligned} & 0 \leqslant \frac{1}{2}\Big(\frac{\ln(K/S_0)}{\sigma\sqrt{T/n}} + n\Big) \leqslant n \\ \Longleftrightarrow\ & 0 \leqslant \frac{1}{2}(\frac{\ln(K/S_0)}{\sigma\sqrt{T/n}} + n) \text{ and } \frac{1}{2}(\frac{\ln(K/S_0)}{\sigma\sqrt{T/n}} + n) \leqslant n, \end{aligned}$$

then it is equivalent to

$$\begin{aligned} 0 \leqslant & \frac{1}{2}(\frac{\ln(K/S_0)}{\sigma\sqrt{T/n}} + n) \\ \Longleftrightarrow\ & 0 \leqslant \frac{\ln(K/S_0)}{\sigma\sqrt{T}} + \sqrt{n} \\ \Longleftrightarrow\ & -\frac{\ln(K/S_0)}{\sigma\sqrt{T}} \leqslant \sqrt{n} \end{aligned}$$

and

$$\begin{aligned} & \frac{1}{2}(\frac{\ln(K/S_0)}{\sigma\sqrt{T/n}} + n) \leqslant n \\ \Longleftrightarrow\ & \frac{\ln(K/S_0)}{\sigma\sqrt{T/n}} \leqslant n \\ \Longleftrightarrow\ & \frac{\ln(K/S_0)}{\sigma\sqrt{T}} \leqslant \sqrt{n}. \end{aligned}$$

Hence combining the above together, we have

$$\begin{aligned} & 0 \leqslant \frac{1}{2}\Big(\frac{\ln(K/S_0)}{\sigma\sqrt{T/n}} + n\Big) \leqslant n \\ \Longleftrightarrow\ & \Big|\frac{\ln(K/S_0)}{\sigma\sqrt{T}}\Big| \leqslant \sqrt{n}. \end{aligned}$$

□

For β and γ, it is obvious that

Corollary 5.2: *For any given L_u, σ and T, to ensure $0 \leqslant \beta \leqslant n$, n must satisfy*

$$\sqrt{n} \geqslant \Big|\frac{\ln(L_u)}{\sigma\sqrt{T}}\Big|$$

and

Corollary 5.3: *For any given L_d, σ and T, to ensure $0 \leqslant \gamma \leqslant n$, n must satisfy*

$$\sqrt{n} \geqslant \Big|\frac{\ln(L_d)}{\sigma\sqrt{T}}\Big|.$$

Since S_0, K, L_d, L_u, σ and T all have constant values, the above solutions show the restrictions for the range of n, the number of steps in a binomial tree. Therefore combining the above, we must have

Theorem 5.4: *For any given L_d, σ and T, to ensure $0 \leqslant \alpha, \gamma, \beta \leqslant n$, where*

$$\alpha = \frac{1}{2}\Big(\frac{\ln(K/S_0)}{\sigma\sqrt{T/n}} + n\Big), \tag{5.6}$$

$$\beta = \frac{1}{2}\Big(\frac{\ln(L_u)}{\sigma\sqrt{T/n}} + n\Big) \tag{5.7}$$

and

$$\gamma = \frac{1}{2}\Big(\frac{\ln(L_d)}{\sigma\sqrt{T/n}} + n\Big). \tag{5.8}$$

n must satisfy

$$\sqrt{n} \geqslant \max\left(\left|\frac{\ln(L_d)}{\sigma\sqrt{T}}\right|, \left|\frac{\ln(L_u)}{\sigma\sqrt{T}}\right|, \left|\frac{\ln(K/S_0)}{\sigma\sqrt{T}}\right|\right).$$

5.2 Upper price bound

In this section, an extension of the CRR formula where price movement is bounded from the above is introduced. If there is an upper bound $1 < L_u \ll \infty$ on the price movement, then we have

Hypothesis 5.1:

$$S_0 u^i d^j \leqslant S_0 L_u,$$

$$\forall i, j \in \mathbb{Z}^+ \cup \{0\}, i + j = n,$$

where u is the upward rate, d is the downward rate at each step and S_0 is the initial price.

Now we show that the above is equivalent to

Lemma 5.5:

$$i \leqslant \frac{1}{2}\Big(\frac{\ln(L_u)}{\sigma\sqrt{T/n}} + n\Big),$$

i.e. $\beta := \frac{1}{2}\Big(\frac{\ln(L_u)}{\sigma\sqrt{T/n}} + n\Big)$ is an upper bound for i.

Proof: According to Proposition 4.2, since $S_0 u^i d^j \leqslant S_0 L_u$,

$$\begin{aligned} u^i d^j =& u^{i-j} u^j d^j \\ =& u^{i-j} \leqslant L_u \\ \iff& (i-j)\ln(u) \leqslant \ln(L_u) \\ \iff& i - j \leqslant \frac{\ln(L_u)}{\sigma\sqrt{T/n}}. \end{aligned}$$

Since $i + j = n$,

$$2i \leqslant \frac{\ln(L_u)}{\sigma\sqrt{T/n}} + n.$$

Hence

$$i \leqslant \beta := \frac{1}{2}\Big(\frac{\ln(L_u)}{\sigma\sqrt{T/n}} + n\Big).$$

□

Furthermore we have

Lemma 5.6: $\beta := \frac{1}{2}\Big(\frac{\ln(L_u)}{\sigma\sqrt{T/n}} + n\Big)$ *is the least upper bound.*

Proof: Proof by contradiction. We assume that there exists an arbitrarily small positive quantity ε, such that the least upper bound is $\beta - \varepsilon$. So at time T,

$$\begin{aligned} S_T/S_0 :=& u^{\beta-\varepsilon}d^{n-\beta+\varepsilon} \\ =& u^{2\beta-n-2\varepsilon} \\ =& e^{\sigma\sqrt{T/n}\times(2\beta-n-2\varepsilon)} \\ =& e^{(2\beta-n)\sigma\sqrt{T/n}}/e^{2\varepsilon\sigma\sqrt{T/n}} \\ =& L_u/e^{2\varepsilon\sigma\sqrt{T/n}}. \end{aligned}$$

Since $2\varepsilon\sigma\sqrt{T/n} > 0$, then $S_T < S_0L_u$. It is a contradiction to our assumption. □

Then using β as the upper bound for i, we re-write (4.12) into

$$\begin{aligned} c =& e^{-rT}\sum_{i=0}^{\lfloor\beta\rfloor}\frac{n!}{(n-i)!i!}p^i(1-p)^{n-i}\max(S_0u^id^{n-i}-K,0) \\ =& e^{-rT}\sum_{i=0}^{\lfloor\beta\rfloor}b(i)f_{payoff}(i), \end{aligned} \tag{5.9}$$

where floor of β, $\lfloor\beta\rfloor := \max\{m \in \mathbb{Z} | m \leqslant \beta\}$.

According to Proposition 4.4 and Lemma 4.5, it is clear that

$$\sum_{i=0}^{n} f_{payoff}(i) = \sum_{i\geqslant\alpha}^{n} f(i), \tag{5.10}$$

where

$$\alpha = \frac{1}{2}\Big(\frac{\ln(K/S_0)}{\sigma\sqrt{T/n}} + n\Big)$$

and

$$f(i) := S_0 u^i d^{n-i} - K.$$

Now we split our discussion into the following cases:

1. $K \leqslant S_0 L_u$.
2. $K > S_0 L_u$.

For Case 2, it is obvious that for any S_T in $(0, S_0 L_u]$, S_T is less than K. According to Proposition 4.4 and Lemma 4.5, call option value is zero for such S_T. For Case 1, we have

Lemma 5.7:

$$\alpha \leqslant \beta \iff K \leqslant S_0 L_u.$$

Proof:

$$\begin{aligned}
&\alpha \leqslant \beta \\
\iff &\frac{1}{2}\left(\frac{\ln(K/S_0)}{\sigma\sqrt{T/n}} + n\right) \leqslant \frac{1}{2}\left(\frac{\ln(L_u)}{\sigma\sqrt{T/n}} + n\right) \\
\iff &\frac{K}{S_0} \leqslant L_u \\
\iff &K \leqslant S_0 L_u.
\end{aligned}$$

□

As $\alpha \leqslant \beta$, according to Proposition 4.4 and Lemma 4.5, we also have

$$\sum_{i \geqslant \lfloor \beta \rfloor}^{n} f_{payoff}(i) = \sum_{i \geqslant \lfloor \beta \rfloor}^{n} f(i). \tag{5.11}$$

Then (4.12) becomes

$$\begin{aligned}
c :=& e^{-rT} \sum_{i=0}^{\lfloor \beta \rfloor} b(i) f_{payoff}(i) \\
=& e^{-rT} \Big(\sum_{i=0}^{n} b(i) f_{payoff}(i) - \sum_{i \geqslant \lfloor \beta \rfloor}^{n} b(i) f_{payoff}(i) \Big) \\
=& e^{-rT} \sum_{i=0}^{n} b(i) f_{payoff}(i) - e^{-rT} \sum_{i \geqslant \lfloor \beta \rfloor}^{n} b(i) f_{payoff}(i) \\
=:& e^{-rT} \sum_{i \geqslant \alpha}^{n} b(i) f(i) - c'.
\end{aligned} \tag{5.12}$$

For c', we have

$$
\begin{aligned}
c' :=& e^{-rT} \sum_{i \geqslant \lfloor \beta \rfloor}^{n} b(i) f(i) \\
=& e^{-rT} \sum_{i' \geqslant \lfloor \beta \rfloor}^{n} \frac{n!}{(n-i')!i'!} p^{i'} (1-p)^{n-i'} (S_0 u^{i'} d^{n-i'} - K) \\
=& e^{-rT} \Big(S_0 \sum_{i' \geqslant \lfloor \beta \rfloor}^{n} \frac{n!}{(n-i')!i'!} p^{i'} (1-p)^{n-i'} u^{i'} d^{n-i'} \\
& -K \sum_{i' \geqslant \lfloor \beta \rfloor}^{n} \frac{n!}{(n-i')!i'!} p^{i'} (1-p)^{n-i'} \Big)
\end{aligned} \tag{5.13}
$$

and

$$
\beta = \frac{1}{2} \Big(\frac{\ln(L_u)}{\sigma \sqrt{T/n}} + n \Big). \tag{5.14}
$$

Then we can define

$$
U_1' = \sum_{i' \geqslant \lfloor \beta \rfloor}^{n} \frac{n!}{(n-i')!i'!} p^{i'} (1-p)^{n-i'} u^{i'} d^{n-i'} \tag{5.15}
$$

and

$$
U_2' = \sum_{i' \geqslant \lfloor \beta \rfloor}^{n} \frac{n!}{(n-i')!i'!} p^{i'} (1-p)^{n-i'}. \tag{5.16}
$$

So (5.13) can be re-written as

$$
c' = e^{-rT} (S_0 U_1' - K U_2'). \tag{5.17}
$$

According to Corollary 4.14 and Corollary 4.12, if n is large enough, (5.15) and (5.16) become

$$
U_1' \approx e^{rT} \Phi \Big(\frac{\ln(1/L_u) + (r + \sigma^2/2)T}{\sigma \sqrt{T}} \Big) \tag{5.18}
$$

and

$$
U_2' \approx \Phi \Big(\frac{\ln(1/L_u) + (r - \sigma^2/2)T}{\sigma \sqrt{T}} \Big). \tag{5.19}
$$

Then (5.13) becomes

$$
\begin{aligned}
c' \longrightarrow e^{-rT} \Bigg(& S_0 e^{rT} \Phi \Big(\frac{\ln(1/L_u) + (r + \sigma^2/2)T}{\sigma \sqrt{T}} \Big) \\
& -K \Phi \Big(\frac{\ln(1/L_u) + (r - \sigma^2/2)T}{\sigma \sqrt{T}} \Big) \Bigg),
\end{aligned} \tag{5.20}
$$

as $n \longrightarrow \infty$. By combining the above results together, since n is large enough, we obtain

Theorem 5.8: *If $K \leqslant S_0 L_u$ (Case 1),*

$$c = S_0 \Phi(d_1) - K e^{-rT} \Phi(d_2) - S_0 \Phi(d_3) + K e^{-rT} \Phi(d_4),$$

where

$$d_1 = \frac{\ln(S_0/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}},$$

$$d_2 = \frac{\ln(S_0/K) + (r - \sigma^2/2)T}{\sigma\sqrt{T}},$$

$$d_3 = \frac{\ln(1/L_u) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}$$

and

$$d_4 = \frac{\ln(1/L_u) + (r - \sigma^2/2)T}{\sigma\sqrt{T}}.$$

If $K > S_0 L_u$ (Case 2),

$$c = 0.$$

Proof: If $K > S_0 L_u$ (Case 2), then $\alpha > \beta$. Since the lower bound of summation is greater than the upper bound of summation in $\sum_{i \geqslant \alpha}^{\beta} f(i)$, we have

$$\sum_{i \geqslant \alpha}^{\beta} f(i) = 0,$$

hence $c = 0$. If $K \leqslant S_0 L_u$ (Case 1), according to Hull's (2015) results and Theorem 4.1, it is clear that

$$e^{-rT} \sum_{i \geqslant \alpha}^{n} b(i) f(i) = S_0 \Phi(d_1) - K e^{-rT} \Phi(d_2),$$

where

$$d_1 = \frac{\ln(S_0/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}$$

and

$$d_2 = \frac{\ln(S_0/K) + (r - \sigma^2/2)T}{\sigma\sqrt{T}}.$$

According to (5.13), if n is large enough,

$$c' = S_0 \Phi(d_3) - K e^{-rT} \Phi(d_4),$$

where

$$d_3 = \frac{\ln(1/L_u) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}$$

and

$$d_4 = \frac{\ln(1/L_u) + (r - \sigma^2/2)T}{\sigma\sqrt{T}}.$$

Hence

$$\begin{aligned} c =& e^{-rT}\sum_{i=0}^{n} f(i) - c' \\ =& S_0\Phi(d_1) - Ke^{-rT}\Phi(d_2) - S_0\Phi(d_3) + Ke^{-rT}\Phi(d_4). \end{aligned}$$

□

5.3 Lower price bound

In this section, we follow the same procedure as the previous section and introduce another extension of the CRR formula where price movement is bounded from below. We assume that there is a lower bound $L_d \in (0, 1)$ at time $t = T$, then we have

Hypothesis 5.2:

$$S_0 L_d \leqslant S_0 u^i d^j,$$

$$\forall i, j \in \mathbb{Z}^+ \cup \{0\}, i + j = n.$$

The above is equivalent to

Lemma 5.9:

$$\gamma := \frac{1}{2}\Big(\frac{\ln(L_d)}{\sigma\sqrt{T/n}} + n\Big)$$

is a lower bound.

Proof: According to Proposition 4.2, since $S_0 L_d \leqslant S_0 u^i d^j$, we could get

$$\begin{aligned} & u^{i-j} \geqslant L_d \\ \Longleftrightarrow & (i-j)\ln(u) \geqslant \ln(L_d) \\ \Longleftrightarrow & i - j \geqslant \frac{\ln(L_d)}{\sigma\sqrt{T/n}}. \end{aligned} \tag{5.21}$$

Since $i + j = n$, we have

$$i \geqslant \frac{1}{2}\Big(\frac{\ln(L_d)}{\sigma\sqrt{T/n}} + n\Big).$$

□

Furthermore, we have

Lemma 5.10: $\gamma = \frac{1}{2}\Big(\frac{\ln(L_d)}{\sigma\sqrt{T/n}} + n\Big)$ *is the greatest lower bound.*

Proof: Proof by contradiction. We assume that there exists an arbitrarily small positive quantity ε, such that the greatest lower bound is $\gamma + \varepsilon$. So at time T,

$$\begin{aligned} S_T/S_0 :=& u^{\gamma+\varepsilon} d^{n-\gamma-\varepsilon} \\ =& u^{2\gamma-n+2\varepsilon} \\ =& e^{\sigma\sqrt{T/n}\times(2\gamma-n+2\varepsilon)} \\ =& e^{(2\gamma-n)\sigma\sqrt{T/n}} \times e^{2\varepsilon\sigma\sqrt{T/n}} \\ =& L_d \times e^{2\varepsilon\sigma\sqrt{T/n}}. \end{aligned}$$

Since $2\varepsilon\sigma\sqrt{T/n} > 0$, then $S_T > S_0 L_d$. It is a contradiction to our assumption. □

Since γ is a lower bound for i, we must have that for any $i \leqslant \gamma$, the value of our payoff is zero. To satisfy this condition, we can define

$$f_{\text{truncated payoff}}(i) := \begin{cases} f_{payoff}(i) & \text{, if } i > \gamma \\ 0 & \text{, if } i \leqslant \gamma. \end{cases} \tag{5.22}$$

Now we split our discussion into the following cases:

1. $K < S_0 L_d$.
2. $K \geqslant S_0 L_d$.

For Case 1, we have

Lemma 5.11:

$$\alpha < \gamma \iff K < S_0 L_d.$$

Proof:

$$\begin{aligned} & \alpha < \gamma \\ \iff & \frac{1}{2}(\frac{\ln(K/S_0)}{\sigma\sqrt{T/n}} + n) < \frac{1}{2}\Big(\frac{\ln(L_d)}{\sigma\sqrt{T/n}} + n\Big) \\ \iff & \frac{K}{S_0} < L_d \\ \iff & K < S_0 L_d. \end{aligned}$$

□

So $i \geqslant \lceil\gamma\rceil \Longrightarrow i > \alpha$, where ceiling of γ, $\lceil\gamma\rceil := \min\{m \in \mathbb{Z} | m \geqslant \gamma\}$. Hence if $i \geqslant \lceil\gamma\rceil$,

$$f_{\text{truncated payoff}}(i) = f_{payoff}(i) = f(i). \tag{5.23}$$

According to (5.22), for any $i \leqslant \lceil \gamma \rceil$, $f_{\text{truncated payoff}}(i)$ is zero. Then (4.12) becomes

$$\begin{aligned} c =& e^{-rT} \sum_{i=0}^{n} b(i) f_{\text{truncated payoff}}(i) \\ =& e^{-rT} \sum_{i=0}^{\alpha} b(i) f_{\text{truncated payoff}}(i) + e^{-rT} \sum_{i \geqslant \alpha}^{\lceil \gamma \rceil} b(i) f_{\text{truncated payoff}}(i) \\ &+ e^{-rT} \sum_{i \geqslant \lceil \gamma \rceil}^{n} b(i) f_{\text{truncated payoff}}(i) \\ =& 0 + 0 + e^{-rT} \sum_{i \geqslant \lceil \gamma \rceil}^{n} b(i) f(i) \\ =:& c'. \end{aligned} \tag{5.24}$$

For c', we have

$$\begin{aligned} c' :=& e^{-rT} \sum_{i \geqslant \lceil \gamma \rceil}^{n} b(i) f(i) \\ =& e^{-rT} \Big(S_0 \sum_{i' \geqslant \lceil \gamma \rceil}^{n} \frac{n!}{(n-i')!i'!} p^{i'} (1-p)^{n-i'} u^{i'} d^{n-i'} \\ & - K \sum_{i' \geqslant \lceil \gamma \rceil}^{n} \frac{n!}{(n-i')!i'!} p^{i'} (1-p)^{n-i'} \Big) \end{aligned} \tag{5.25}$$

and

$$\gamma = \frac{1}{2} \Big(\frac{\ln(L_d)}{\sigma \sqrt{T/n}} + n \Big). \tag{5.26}$$

We can define

$$U_1' = \sum_{i' \geqslant \lceil \gamma \rceil}^{n} \frac{n!}{(n-i')!i'!} p^{i'} (1-p)^{n-i'} u^{i'} d^{n-i'} \tag{5.27}$$

and

$$U_2' = \sum_{i' \geqslant \lceil \gamma \rceil}^{n} \frac{n!}{(n-i')!i'!} p^{i'} (1-p)^{n-i'}. \tag{5.28}$$

then re-arrange the above as

$$c' = e^{-rT} (S_0 U_1' - K U_2'). \tag{5.29}$$

According to Corollary 4.14 and Corollary 4.12, if n is large enough, (5.27) and (5.28) become

$$U_1' \longrightarrow e^{rT}\Phi\Big(\frac{\ln(1/L_d)+(r+\sigma^2/2)T}{\sigma\sqrt{T}}\Big) \tag{5.30}$$

and

$$U_2' \longrightarrow \Phi\Big(\frac{\ln(1/L_d)+(r-\sigma^2/2)T}{\sigma\sqrt{T}}\Big), \tag{5.31}$$

as $n \longrightarrow \infty$.

For Case 2, $K \geqslant S_0L_d$, it is obvious that if $S_T \in [K,\infty)$, S_T is greater than S_0L_d. Hence $f_{\text{truncated payoff}}(i) = f_{payoff}(i)$ for any i in this case. And again $S_T \geqslant K$, then $f_{\text{truncated payoff}}(i) = f(i)$. By combining the above results together, since n is large enough, we obtain

Theorem 5.12: *If $K < S_0L_d$ (Case 1),*

$$c' = S_0\Phi(d_1) - Ke^{-rT}\Phi(d_2),$$

where

$$d_1 = \frac{\ln(1/L_d)+(r+\sigma^2/2)T}{\sigma\sqrt{T}}$$

and

$$d_2 = \frac{\ln(1/L_d)+(r-\sigma^2/2)T}{\sigma\sqrt{T}}.$$

If $S_0L_d \leqslant K$ (Case 2), the original CRR model should be used, i.e.

$$c' = S_0\Phi(d_1) - Ke^{-rT}\Phi(d_2),$$

where

$$d_1 = \frac{\ln(S_0/K)+(r+\sigma^2/2)T}{\sigma\sqrt{T}}$$

and

$$d_2 = \frac{\ln(S_0/K)+(r-\sigma^2/2)T}{\sigma\sqrt{T}}.$$

Proof: If $K < S_0L_d$ (Case 1),

$$\begin{aligned} c' =& e^{-rT}(S_0U_1' - KU_2') \\ =& e^{-rT}\Bigg(S_0e^{rT}\Phi\Big(\frac{\ln(1/L_d)+(r+\sigma^2/2)T}{\sigma\sqrt{T}}\Big) \\ & -K\Phi\Big(\frac{\ln(1/L_d)+(r-\sigma^2/2)T}{\sigma\sqrt{T}}\Big)\Bigg) \\ =& S_0\Phi(d_1) - Ke^{-rT}\Phi(d_2), \end{aligned}$$

where

$$d_1 = \frac{\ln(1/L_d) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}$$

and

$$d_2 = \frac{\ln(1/L_d) + (r - \sigma^2/2)T}{\sigma\sqrt{T}}.$$

If $S_0 L_d \leqslant K$ (Case 2), according to Proposition 4.4 and Lemma 4.5, it is clear that for any $i \in [\gamma, \alpha]$, we have $\max(S_0 u^{i'} d^{n-i'} - K, 0) = 0$, so

$$\sum_{i \geqslant \gamma}^{\alpha} f(i) = 0.$$

Hence according to Hull's (2015) results and Theorem 4.1, it is clear that

$$c' = S_0 \Phi(d_1) - K e^{-rT} \Phi(d_2),$$

where

$$d_1 = \frac{\ln(S_0/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}$$

and

$$d_2 = \frac{\ln(S_0/K) + (r - \sigma^2/2)T}{\sigma\sqrt{T}}.$$

□

5.4 Combine the upper and lower bounds

In this section, a stock market where both upper and lower bounds exist for the underlying security price movements will be analysed. For the price bounds $S_0 L_d$ and $S_0 L_u$, since $0 < L_d < 1 < L_u \ll \infty$, we always have $S_0 L_d < S_0 L_u$. So we could have three possible situations for strike price K:

1. $K < S_0 L_d$ (upper bound Case 1 and lower bound Case 1).
2. $S_0 L_d \leqslant K \leqslant S_0 L_u$ (upper bound Case 1 and lower bound Case 2).
3. $S_0 L_u < K$ (upper bound Case 2 and lower bound Case 2).

These situations will be discussed separately. For Situation 1, we can combine upper bound Case 1 result and lower bound Case 1 result, so the payoff formula becomes Figure 5.2.

Hence we have

Theorem 5.13: *If $K < S_0 L_d$ (Situation 1),*

$$c = S_0 \Phi(d_1) - K e^{-rT} \Phi(d_2) - S_0 \Phi(d_3) + K e^{-rT} \Phi(d_4),$$

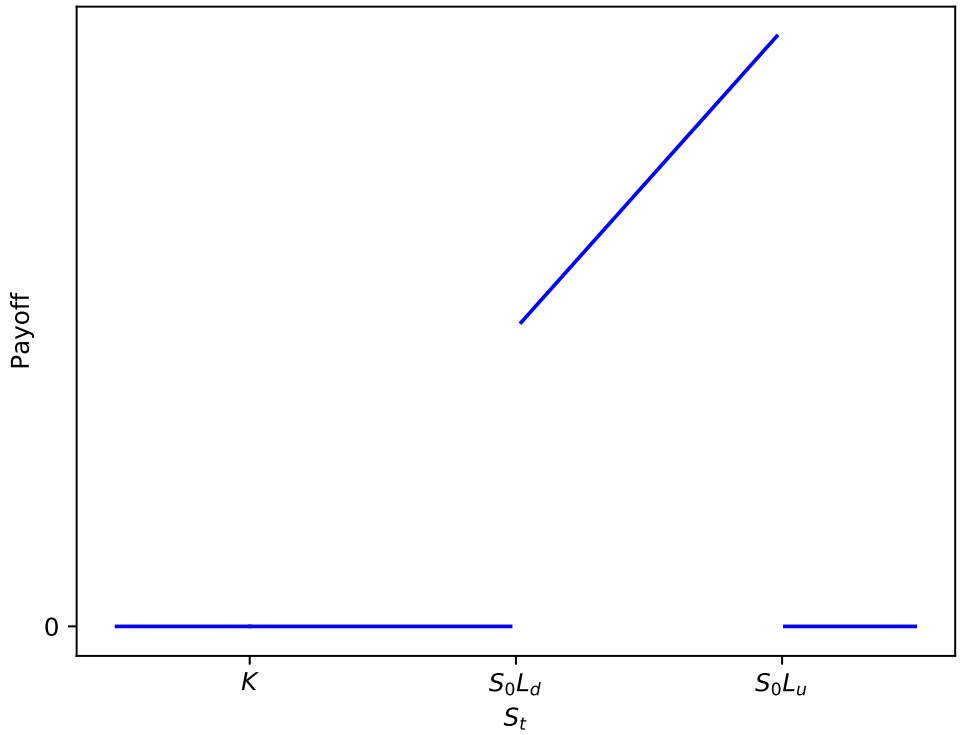


Figure 5.2: Payoff: $K < S_0L_d$.

where

$$d_1 = \frac{\ln(1/L_d) + (r + \sigma^2/2)T}{\sigma\sqrt{T}},$$

$$d_2 = \frac{\ln(1/L_d) + (r - \sigma^2/2)T}{\sigma\sqrt{T}},$$

$$d_3 = \frac{\ln(1/L_u) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}$$

and

$$d_4 = \frac{\ln(1/L_u) + (r - \sigma^2/2)T}{\sigma\sqrt{T}}.$$

Proof: $K < S_0L_d \Longrightarrow \alpha < \gamma$, according to (5.22),

$$\sum_{i \geqslant \alpha}^{\gamma} f_{\text{truncated payoff}}(i) = 0.$$

Hence according to Theorem 5.8 Case 1 result and Theorem 5.12 Case 1 result, (4.12) becomes

$$\begin{aligned}
c =& e^{-rT}\sum_{i=0}^{n} b(i) f_{\text{truncated payoff}}(i) \\
=& e^{-rT}\sum_{i\geqslant\lceil\gamma\rceil}^{\lfloor\beta\rfloor} f(i) \\
=& e^{-rT}\sum_{i\geqslant\lceil\gamma\rceil}^{n} f(i) - e^{-rT}\sum_{i\geqslant\lfloor\beta\rfloor}^{n} f(i) \\
=& S_0\Phi(d_1) - Ke^{-rT}\Phi(d_2) - S_0\Phi(d_3) + Ke^{-rT}\Phi(d_4),
\end{aligned}$$

where

$$d_1 = \frac{\ln(1/L_d) + (r+\sigma^2/2)T}{\sigma\sqrt{T}},$$

$$d_2 = \frac{\ln(1/L_d) + (r-\sigma^2/2)T}{\sigma\sqrt{T}},$$

$$d_3 = \frac{\ln(1/L_u) + (r+\sigma^2/2)T}{\sigma\sqrt{T}}$$

and

$$d_4 = \frac{\ln(1/L_u) + (r-\sigma^2/2)T}{\sigma\sqrt{T}}.$$

□

For Situation 3, we can combine upper bound Case 2 result and lower bound Case 2 result, so we have

Theorem 5.14: *If $S_0L_u < K$ (Situation 3),*

$$c = 0.$$

Proof: $S_0L_u < K \Longrightarrow \beta < \alpha$, since the lower bound of summation is greater than the upper bound of summation in $\sum_{i\geqslant\alpha}^{\beta} f(i)$, we have

$$\sum_{i\geqslant\alpha}^{\beta} f(i) = 0,$$

hence $c = 0$. □

And for Situation 2, we can combine upper bound Case 1 result and lower bound Case 2 result, so the amount of payoff is plotted in Figure 5.3.

Hence we have

Theorem 5.15: *If $S_0L_d \leqslant K \leqslant S_0L_u$ (Situation 2),*

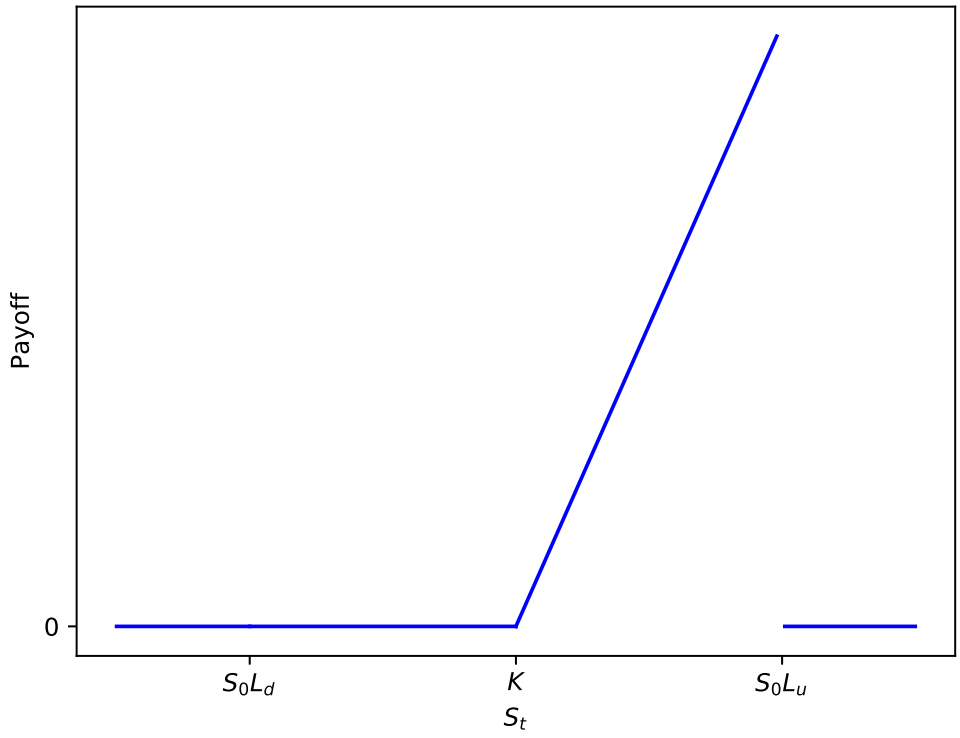


Figure 5.3: Payoff: $S_0L_d \leqslant K \leqslant S_0L_u$.

$$c = S_0\Phi(d_1) - Ke^{-rT}\Phi(d_2) - S_0\Phi(d_3) + Ke^{-rT}\Phi(d_4),$$

where

$$d_1 = \frac{\ln(S_0/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}},$$

$$d_2 = \frac{\ln(S_0/K) + (r - \sigma^2/2)T}{\sigma\sqrt{T}},$$

$$d_3 = \frac{\ln(1/L_u) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}$$

and

$$d_4 = \frac{\ln(1/L_u) + (r - \sigma^2/2)T}{\sigma\sqrt{T}}.$$

Proof: $S_0L_d \leqslant K \leqslant S_0L_u \Longrightarrow \gamma \leqslant \alpha \leqslant \beta$, so $i \leqslant \gamma \Longrightarrow i \leqslant \alpha$. Then according to Proposition 4.4 and Lemma 4.5, we have that for any $i < \alpha$, $f_{payoff}(i) = 0$. So

$$\sum_{i \geqslant \gamma}^{\alpha} b(i) f_{payoff}(i) = 0.$$

Hence according to Theorem 5.8 Case 1 result and Theorem 5.12 Case 2 result, (4.12) becomes

$$\begin{aligned} c =& e^{-rT} \sum_{i=0}^{n} b(i) f_{\text{truncated payoff}}(i) \\ =& e^{-rT} \sum_{i \geqslant \alpha}^{n} b(i) f(i) - e^{-rT} \sum_{i \geqslant \beta}^{n} b(i) f(i) \\ =& S_0 \Phi(d_1) - K e^{-rT} \Phi(d_2) - S_0 \Phi(d_3) + K e^{-rT} \Phi(d_4), \end{aligned}$$

where

$$d_1 = \frac{\ln(S_0/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}},$$

$$d_2 = \frac{\ln(S_0/K) + (r - \sigma^2/2)T}{\sigma\sqrt{T}},$$

$$d_3 = \frac{\ln(1/L_u) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}$$

and

$$d_4 = \frac{\ln(1/L_u) + (r - \sigma^2/2)T}{\sigma\sqrt{T}}.$$

□

Chapter 6

Re-weighting model

6.1 Introduction

In this chapter, we will discuss another type of European call option. Similar to the discussions made in Chapter 5, the underlying asset for this option has boundaries imposed on its price movements: S_0L_d and S_0L_u, where $L_d < L_u$. Suppose that the market regulator dictates that if the underlying stock price S_T at time $t = T$ goes beyond the stated boundaries, the buyers of this option will get a refund, and the option contract is annulled. The refund value equals the future value at time T of its initial cost c. In other words, this option has a payoff formula which is defined as

$$\begin{cases} e^{rT}c, & \text{if } S_T \geqslant S_0L_u \\ \max(S_T - K, 0), & \text{if } S_T \in (S_0L_d, S_0L_u) \\ e^{rT}c, & \text{if } S_T \leqslant S_0L_d, \end{cases} \tag{6.1}$$

where c is the European call option initial cost at time $t = 0$. In this chapter, our payoff formula is an implicit formula. Right now, we don't know the exact value of c. Now let me introduce some technical preparations here. The following lemmas will be used later. For our n-step binomial tree defined in Chapter 4, we introduce that for S_0L_u and $S_0u^id^{n-i}$ (which equals S_T),

Lemma 6.1: $S_0u^id^{n-i} > S_0L_u \iff i > \beta$,
where

$$\beta := \frac{1}{2}\Big(\frac{\ln(L_u)}{\sigma\sqrt{T/n}} + n\Big).$$

Proof: According to Lemma 5.5,

$$\begin{aligned} & S_0u^id^{n-i} > S_0L_u \\ \iff\ & i > \frac{1}{2}\Big(\frac{\ln(L_u)}{\sigma\sqrt{T/n}} + n\Big). \end{aligned}$$

□

For K and $S_0u^id^{n-i}$,

Lemma 6.2: $S_0u^id^{n-i} > K \iff i > \alpha$,
where

$$\alpha := \frac{1}{2}\Big(\frac{\ln(K/S_0)}{\sigma\sqrt{T/n}} + n\Big).$$

Proof: According to Lemma 4.5,

$$\begin{aligned} & S_0u^id^{n-i} > K \\ \iff & i > \frac{1}{2}\Big(\frac{\ln(K/S_0)}{\sigma\sqrt{T/n}} + n\Big). \end{aligned}$$

□

And for S_0L_d and $S_0u^id^{n-i}$,

Lemma 6.3: $S_0u^id^{n-i} > S_0L_d \iff i > \gamma$,
where

$$\gamma := \frac{1}{2}\Big(\frac{\ln(L_d)}{\sigma\sqrt{T/n}} + n\Big).$$

Proof: According to Lemma 5.9,

$$\begin{aligned} & S_0u^id^{n-i} > S_0L_d \\ \iff & i > \frac{1}{2}\Big(\frac{\ln(L_d)}{\sigma\sqrt{T/n}} + n\Big). \end{aligned}$$

□

Now we use these lemmas to re-write (6.1) into three cases:

- $K \in (S_0L_d, S_0L_u)$, i.e. $\alpha \in [\gamma, \beta]$.
- $K < S_0L_d$, i.e. $\alpha < \gamma$.
- $K > S_0L_u$, i.e. $\alpha > \beta$.

We will discuss them respectively.

6.2 If $K \in (S_0L_d, S_0L_u)$

If $K \in (S_0L_d, S_0L_u)$, it is obvious that (S_0L_d, S_0L_u) can be divided into two intervals, $(S_0L_d, K]$ and (K, S_0L_u). If $S_0u^id^{n-i} \in (S_0L_d, K]$, then

$$S_0u^id^{n-i} - K \leqslant 0, \tag{6.2}$$

hence

$$\max(S_0u^id^{n-i} - K, 0) = 0. \tag{6.3}$$

If $S_0u^id^{n-i} \in (K, S_0L_u)$, then

$$S_0u^id^{n-i} - K > 0, \tag{6.4}$$

hence

$$\max(S_0u^id^{n-i} - K, 0) = S_0u^id^{n-i} - K. \tag{6.5}$$

Therefore if $K \in (S_0L_d, S_0L_u)$, (6.1) becomes

$$\begin{cases} e^{rT}c, & \text{if } S_0u^id^{n-i} \geqslant S_0L_u \\ S_0u^id^{n-i} - K := f(i), & \text{if } S_0u^id^{n-i} \in (K, S_0L_u) \\ 0, & \text{if } S_0u^id^{n-i} \in (S_0L_d, K] \\ e^{rT}c, & \text{if } S_0u^id^{n-i} \leqslant S_0L_d. \end{cases} \tag{6.6}$$

According to Lemmas 6.1, 6.2 and 6.3, the above payoff formula (also shown in Figure 6.1) can be re-written as

$$f_{\text{truncated payoff}}(i) := \begin{cases} e^{rT}c, & \text{if } i \geqslant \beta \\ f(i), & \text{if } i \in (\alpha, \beta) \\ 0, & \text{if } i \in (\gamma, \alpha] \\ e^{rT}c, & \text{if } i \leqslant \gamma. \end{cases} \tag{6.7}$$

According to CRR model, our call option is priced as

$$\begin{aligned} c &= e^{-rT} \sum_{i=0}^{n} b(i) f_{\text{truncated payoff}}(i) \qquad (6.8) \\ &= e^{-rT} \Big[\sum_{i=0}^{\lfloor\gamma\rfloor} b(i) f_{\text{truncated payoff}}(i) \\ &\qquad + \sum_{i=\lfloor\gamma\rfloor}^{\lceil\alpha\rceil} b(i) f_{\text{truncated payoff}}(i) \\ &\qquad + \sum_{i=\lceil\alpha\rceil}^{\lfloor\beta\rfloor} b(i) f_{\text{truncated payoff}}(i) \\ &\qquad + \sum_{i=\lfloor\beta\rfloor}^{n} b(i) f_{\text{truncated payoff}}(i) \Big] \\ &= e^{-rT} \Big[\sum_{i=0}^{\lfloor\gamma\rfloor} b(i) e^{rT}c + \sum_{i=\lceil\alpha\rceil}^{\lfloor\beta\rfloor} b(i) f(i) + \sum_{i=\lfloor\beta\rfloor}^{n} b(i) e^{rT}c \Big] \\ &= e^{-rT} \sum_{i=\lceil\alpha\rceil}^{\lfloor\beta\rfloor} b(i) f(i) + e^{-rT} \Big[\sum_{i=0}^{\lfloor\gamma\rfloor} b(i) + \sum_{i=\lfloor\beta\rfloor}^{n} b(i) \Big] e^{rT}c \end{aligned}$$

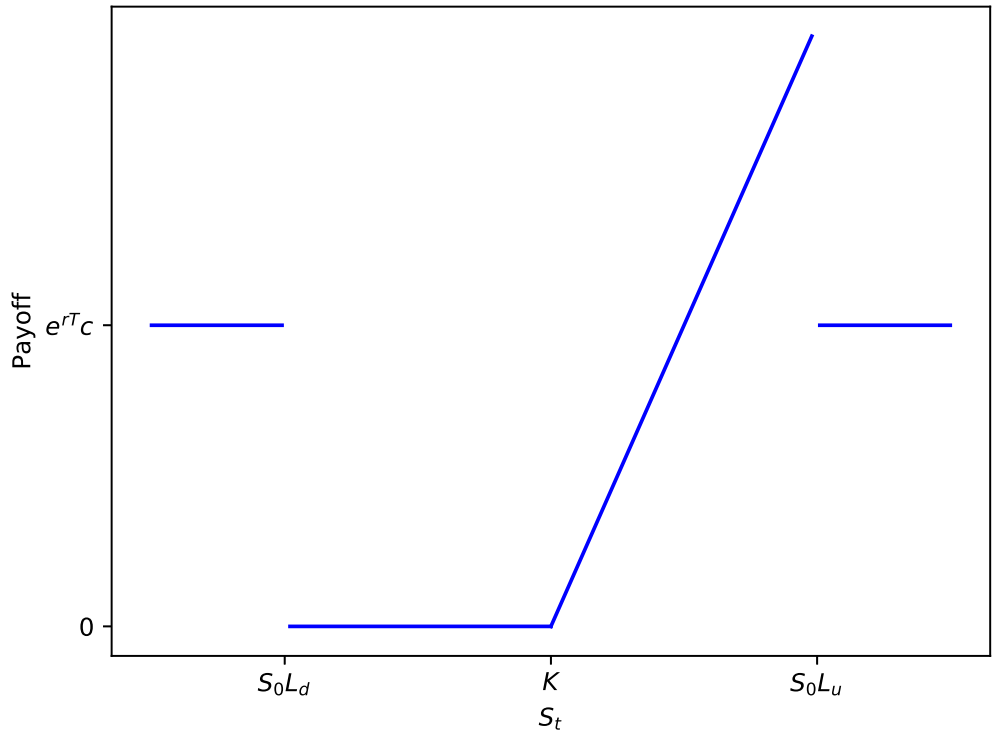


Figure 6.1: Payoff: $K \in (S_0L_d, S_0L_u)$.

$$= e^{-rT} \sum_{i=\lceil\alpha\rceil}^{\lfloor\beta\rfloor} b(i)f(i) + \Big[\sum_{i=0}^{\lfloor\gamma\rfloor} b(i) + \sum_{i=\lfloor\beta\rfloor}^{n} b(i)\Big]c.$$

Rearranging the above, we have

$$c\Big[1 - \sum_{i=0}^{\lfloor\gamma\rfloor} b(i) - \sum_{i=\lfloor\beta\rfloor}^{n} b(i)\Big] = e^{-rT} \sum_{i=\lceil\alpha\rceil}^{\lfloor\beta\rfloor} b(i)f(i). \tag{6.9}$$

Now our model becomes explicit for the option cost c. Since for binomial distribution probability mass function

$$b(i) := \frac{n!}{(n-i)!i!}p^i(1-p)^{n-i},$$

we have

$$\sum_{i=0}^{n} b(i) \tag{6.10}$$

$$
\begin{aligned}
&:=\sum_{i=0}^{n} \frac{n!}{(n-i)!i!} p^i (1-p)^{n-i} \\
&=1.
\end{aligned}
$$

Hence in the left hand side of (6.9),

$$1 - \sum_{i=0}^{\lfloor \gamma \rfloor} b(i) - \sum_{i=\lfloor \beta \rfloor}^{n} b(i) = \sum_{i=\lfloor \gamma \rfloor}^{\lfloor \beta \rfloor} b(i). \tag{6.11}$$

Therefore (6.9) becomes

$$c \sum_{i=\lfloor \gamma \rfloor}^{\lfloor \beta \rfloor} b(i) = e^{-rT} \sum_{i=\lceil \alpha \rceil}^{\lfloor \beta \rfloor} b(i) f(i). \tag{6.12}$$

For the left hand side of (6.12), we have

Lemma 6.4:

$$\sum_{i=\lfloor \gamma \rfloor}^{\lfloor \beta \rfloor} b(i) = \varrho,$$

where

$$\varrho :\approx \Phi\Big(\frac{\ln(1/L_d) + (r - \sigma^2/2)T}{\sigma\sqrt{T}}\Big) - \Phi\Big(\frac{\ln(1/L_u) + (r - \sigma^2/2)T}{\sigma\sqrt{T}}\Big).$$

Proof: According to (5.31),

$$
\begin{aligned}
&\sum_{i=\lfloor \gamma \rfloor}^{n} b(i) \\
\approx &\Phi\Big(\frac{\ln(1/L_d) + (r - \sigma^2/2)T}{\sigma\sqrt{T}}\Big).
\end{aligned}
$$

According to (5.19),

$$
\begin{aligned}
&\sum_{i=\lfloor \beta \rfloor}^{n} b(i) \\
\approx &\Phi\Big(\frac{\ln(1/L_u) + (r - \sigma^2/2)T}{\sigma\sqrt{T}}\Big).
\end{aligned}
$$

Hence

$$\begin{aligned}
&\sum_{i=\lfloor\gamma\rfloor}^{\lfloor\beta\rfloor} b(i) \\
=&\sum_{i=\lfloor\gamma\rfloor}^{n} b(i) - \sum_{i=\lfloor\beta\rfloor}^{n} b(i) \\
\approx&\Phi\Big(\frac{\ln(1/L_d)+(r-\sigma^2/2)T}{\sigma\sqrt{T}}\Big) - \Phi\Big(\frac{\ln(1/L_u)+(r-\sigma^2/2)T}{\sigma\sqrt{T}}\Big).
\end{aligned}$$

□

For the right hand side of (6.12), according to Theorem 5.15,

$$\begin{aligned}
&e^{-rT}\sum_{i=\lceil\alpha\rceil}^{\lfloor\beta\rfloor} b(i)f(i) \\
\approx&S_0\Phi(d_1) - Ke^{-rT}\Phi(d_2) - S_0\Phi(d_5) + Ke^{-rT}\Phi(d_6),
\end{aligned} \tag{6.13}$$

where

$$d_1 = \frac{\ln(S_0/K)+(r+\sigma^2/2)T}{\sigma\sqrt{T}},$$

$$d_2 = \frac{\ln(S_0/K)+(r-\sigma^2/2)T}{\sigma\sqrt{T}},$$

$$d_5 = \frac{\ln(1/L_u)+(r+\sigma^2/2)T}{\sigma\sqrt{T}}$$

and

$$d_6 = \frac{\ln(1/L_u)+(r-\sigma^2/2)T}{\sigma\sqrt{T}}.$$

Therefore combining the above, since n is large enough, we have

Theorem 6.5: *If $K \in (S_0L_d, S_0L_u)$, our European call option is priced as*

$$c = \frac{1}{\varrho}\Big(S_0\Phi(d_1) - Ke^{-rT}\Phi(d_2) - S_0\Phi(d_5) + Ke^{-rT}\Phi(d_6)\Big),$$

where

$$d_1 = \frac{\ln(S_0/K)+(r+\sigma^2/2)T}{\sigma\sqrt{T}},$$

$$d_2 = \frac{\ln(S_0/K)+(r-\sigma^2/2)T}{\sigma\sqrt{T}},$$

$$d_4 = \frac{\ln(1/L_d) + (r - \sigma^2/2)T}{\sigma\sqrt{T}},$$

$$d_5 = \frac{\ln(1/L_u) + (r + \sigma^2/2)T}{\sigma\sqrt{T}},$$

$$d_6 = \frac{\ln(1/L_u) + (r - \sigma^2/2)T}{\sigma\sqrt{T}}$$

and

$$\varrho = \Phi(d_4) - \Phi(d_6).$$

Proof: According to (6.12),

$$\begin{aligned} c =& 1\Big/\Big(\sum_{i=\lfloor\gamma\rfloor}^{\lfloor\beta\rfloor} b(i)\Big) \times e^{-rT} \sum_{i=\lceil\alpha\rceil}^{\lfloor\beta\rfloor} b(i) f(i) \\ =& \frac{1}{\varrho}\Big(S_0\Phi(d_1) - Ke^{-rT}\Phi(d_2) - S_0\Phi(d_5) + Ke^{-rT}\Phi(d_6)\Big). \end{aligned}$$

□

6.3 If $K \leqslant S_0 L_d$

If $K \leqslant S_0 L_d$, then for any $S_T \in (S_0 L_d, S_0 L_u)$, S_T is always greater than K. Hence if $S_T \in (S_0 L_d, S_0 L_u)$, then

$$\max(S_T - K, 0) = S_T - K. \tag{6.14}$$

Therefore if $K \leqslant S_0 L_d$, (6.1) becomes

$$\begin{cases} e^{rT}c, & \text{if } S_0 u^i d^{n-i} \geqslant S_0 L_u \\ f(i), & \text{if } S_0 u^i d^{n-i} \in (S_0 L_d, S_0 L_u) \\ e^{rT}c, & \text{if } S_0 u^i d^{n-i} \leqslant S_0 L_d. \end{cases} \tag{6.15}$$

According to Lemma 6.1, 6.2 and 6.3, the above payoff formula (also shown in Figure 6.2) can be re-written as

$$f_{\text{truncated payoff}}(i) := \begin{cases} e^{rT}c, & \text{if } i \geqslant \beta \\ f(i), & \text{if } i \in (\gamma, \beta) \\ e^{rT}c, & \text{if } i \leqslant \gamma. \end{cases} \tag{6.16}$$

According to CRR model, our call option is priced as

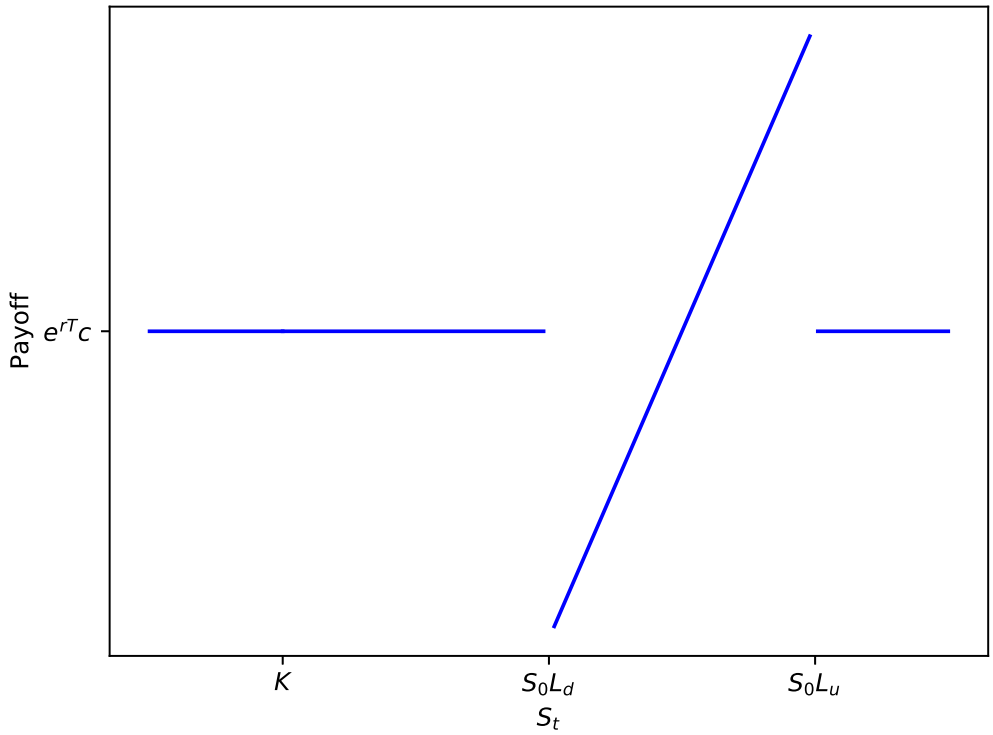


Figure 6.2: Payoff: $K \leqslant S_0 L_d$.

$$
\begin{aligned}
c &= e^{-rT}\ b(i) f_{\text{truncated payoff}}(i) \qquad (6.17)\\
&= e^{-rT}\Big[\sum_{i=0}^{\lfloor\gamma\rfloor} b(i) f_{\text{truncated payoff}}(i)\\
&\qquad + \sum_{i=\lfloor\gamma\rfloor}^{\lfloor\beta\rfloor} b(i) f_{\text{truncated payoff}}(i)\\
&\qquad + \sum_{i=\lfloor\beta\rfloor}^{n} b(i) f_{\text{truncated payoff}}(i)\Big]\\
&= e^{-rT}\Big[\sum_{i=0}^{\lfloor\gamma\rfloor} b(i) e^{rT} c + \sum_{i=\lfloor\gamma\rfloor}^{\lfloor\beta\rfloor} b(i) f(i) + \sum_{i=\lfloor\beta\rfloor}^{n} b(i) e^{rT} c\Big]\\
&= e^{-rT} \sum_{i=\lfloor\gamma\rfloor}^{\lfloor\beta\rfloor} b(i) f(i) + e^{-rT}\Big[\sum_{i=0}^{\lfloor\gamma\rfloor} b(i) + \sum_{i=\lfloor\beta\rfloor}^{n} b(i)\Big] e^{rT} c\\
&= e^{-rT} \sum_{i=\lfloor\gamma\rfloor}^{\lfloor\beta\rfloor} b(i) f(i) + \Big[\sum_{i=0}^{\lfloor\gamma\rfloor} b(i) + \sum_{i=\lfloor\beta\rfloor}^{n} b(i)\Big] c.
\end{aligned}
$$

According to (6.11), (6.17) becomes

$$c\sum_{i=\lfloor\gamma\rfloor}^{\lfloor\beta\rfloor} b(i) = e^{-rT}\sum_{i=\lfloor\gamma\rfloor}^{\lfloor\beta\rfloor} b(i)f(i). \tag{6.18}$$

Now our model becomes explicit for the option cost c. For the left hand side of (6.18), according to Lemma 6.4,

$$\sum_{i=\lfloor\gamma\rfloor}^{\lfloor\beta\rfloor} b(i) = \varrho. \tag{6.19}$$

For the right hand side of (6.18), according to Theorem 5.13,

$$\begin{aligned} & e^{-rT}\sum_{i=\lfloor\gamma\rfloor}^{\lfloor\beta\rfloor} b(i)f(i) \\ \approx & S_0\Phi(d_3) - Ke^{-rT}\Phi(d_4) - S_0\Phi(d_5) + Ke^{-rT}\Phi(d_6), \end{aligned} \tag{6.20}$$

where

$$d_3 = \frac{\ln(1/L_d) + (r+\sigma^2/2)T}{\sigma\sqrt{T}},$$

$$d_4 = \frac{\ln(1/L_d) + (r-\sigma^2/2)T}{\sigma\sqrt{T}},$$

$$d_5 = \frac{\ln(1/L_u) + (r+\sigma^2/2)T}{\sigma\sqrt{T}}$$

and

$$d_6 = \frac{\ln(1/L_u) + (r-\sigma^2/2)T}{\sigma\sqrt{T}}.$$

Therefore, since n is large enough, we have

Theorem 6.6: *If $K \leqslant S_0L_d$, our European call option is priced as*

$$c = \frac{1}{\varrho}\Big(S_0\Phi(d_3) - Ke^{-rT}\Phi(d_4) - S_0\Phi(d_5) + Ke^{-rT}\Phi(d_6)\Big),$$

where

$$d_3 = \frac{\ln(1/L_d) + (r+\sigma^2/2)T}{\sigma\sqrt{T}},$$

$$d_4 = \frac{\ln(1/L_d) + (r-\sigma^2/2)T}{\sigma\sqrt{T}},$$

$$d_5 = \frac{\ln(1/L_u) + (r + \sigma^2/2)T}{\sigma\sqrt{T}},$$

$$d_6 = \frac{\ln(1/L_u) + (r - \sigma^2/2)T}{\sigma\sqrt{T}}$$

and

$$\varrho = \Phi(d_4) - \Phi(d_6).$$

Proof: According to (6.18),

$$\begin{aligned} c =& 1\Big/\Big(\sum_{i=\lfloor\gamma\rfloor}^{\lfloor\beta\rfloor} b(i)\Big) \times e^{-rT} \sum_{i=\lfloor\gamma\rfloor}^{\lfloor\beta\rfloor} b(i) f(i) \\ =& \frac{1}{\varrho}\Big(S_0\Phi(d_3) - Ke^{-rT}\Phi(d_4) - S_0\Phi(d_5) + Ke^{-rT}\Phi(d_6)\Big). \end{aligned}$$

□

6.4 If $K \geqslant S_0 L_u$

If $K \geqslant S_0 L_u$, then for any $S_T \in (S_0 L_d, S_0 L_u)$, we always have

$$\max(S_T - K, 0) = 0. \tag{6.21}$$

Therefore according to Lemma 6.1, 6.2 and 6.3, our payoff formula (also shown in Figure 6.3) can be re-written as

$$f_{\text{truncated payoff}}(i) := \begin{cases} e^{rT}c, & \text{if } i \geqslant \beta \\ 0, & \text{if } i \in (\gamma, \beta) \\ e^{rT}c, & \text{if } i \leqslant \gamma. \end{cases} \tag{6.22}$$

It is obvious that

Theorem 6.7: *If $K \geqslant S_0 L_u$, our European call option has value*

$$c = 0.$$

Proof: According to CRR model, our call option is priced as

$$\begin{aligned} c =& e^{-rT} b(i) f_{\text{truncated payoff}}(i) \\ =& e^{-rT}\Big[\sum_{i=0}^{\lfloor\gamma\rfloor} b(i) f_{\text{truncated payoff}}(i) \\ & + \sum_{i=\lfloor\gamma\rfloor}^{\lfloor\beta\rfloor} b(i) f_{\text{truncated payoff}}(i) \end{aligned}$$

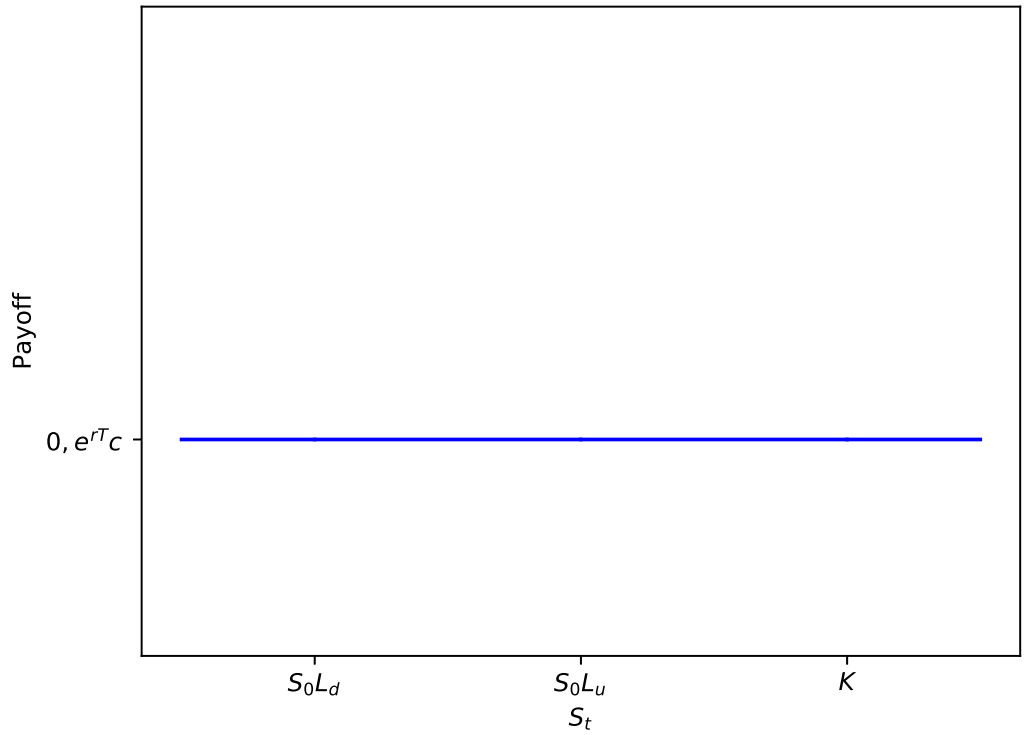


Figure 6.3: Payoff: $K \geqslant S_0L_u$.

$$
\begin{aligned}
&+\sum_{i=\lfloor\beta\rfloor}^{n} b(i) f_{\text{truncated payoff}}(i)\Big] \\
=&e^{-rT}\Big[\sum_{i=0}^{\lfloor\gamma\rfloor} b(i)e^{rT}c+\sum_{i=\lfloor\gamma\rfloor}^{\lfloor\beta\rfloor} b(i)\times 0+\sum_{i=\lfloor\beta\rfloor}^{n} b(i)e^{rT}c\Big] \\
=&0+\Big[\sum_{i=0}^{\lfloor\gamma\rfloor} b(i)+\sum_{i=\lfloor\beta\rfloor}^{n} b(i)\Big]c.
\end{aligned}
$$

Hence according to (6.11) and Lemma 6.4, we have

$$c \times \varrho = 0.$$

Since $L_d < L_u$, according to Lemma 6.4, it is clear that $\varrho > 0$. Therefore $c = 0$. □

6.5 Summary

Finally, we summarise our findings together. For a European call option whose payoff formula satisfies (6.1), we have

Theorem 6.8: *If $K \leqslant S_0 L_d$, the European call option is priced as*

$$c = \frac{1}{\varrho}\Big(S_0\Phi(d_3) - Ke^{-rT}\Phi(d_4) - S_0\Phi(d_5) + Ke^{-rT}\Phi(d_6)\Big).$$

If $K \in (S_0 L_d, S_0 L_u)$, the European call option is priced as

$$c = \frac{1}{\varrho}\Big(S_0\Phi(d_1) - Ke^{-rT}\Phi(d_2) - S_0\Phi(d_5) + Ke^{-rT}\Phi(d_6)\Big).$$

If $K \geqslant S_0 L_u$, the European call option has value

$$c = 0,$$

where

$$d_1 = \frac{\ln(S_0/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}},$$

$$d_2 = \frac{\ln(S_0/K) + (r - \sigma^2/2)T}{\sigma\sqrt{T}},$$

$$d_3 = \frac{\ln(1/L_d) + (r + \sigma^2/2)T}{\sigma\sqrt{T}},$$

$$d_4 = \frac{\ln(1/L_d) + (r - \sigma^2/2)T}{\sigma\sqrt{T}},$$

$$d_5 = \frac{\ln(1/L_u) + (r + \sigma^2/2)T}{\sigma\sqrt{T}},$$

$$d_6 = \frac{\ln(1/L_u) + (r - \sigma^2/2)T}{\sigma\sqrt{T}}$$

and

$$\varrho = \Phi(d_4) - \Phi(d_6).$$

Chapter 7

Re-distributed model

7.1 Introduction

In this chapter, we assume that the stock economic value $V_t := S_0 u^i d^j$ follows a geometric Brownian motion for any time $t \in [0, T]$. However the market regulator imposes price bounds $S_0 L_d$ and $S_0 L_u$ for the market price P_t, where $0 < L_d < 1 < L_u$. Hence combining the above, we set

Hypothesis 7.1: *At time $t = T$, the market price*

$$P_T = \begin{cases} S_0 L_u, & \text{if } V_T := S_0 u^i d^{n-i} > S_0 L_u \\ V_T, & \text{if } V_T \in [S_0 L_d, S_0 L_u] \\ S_0 L_d, & \text{if } V_T < S_0 L_d. \end{cases}$$

According to Lemmas 6.1, 6.2 and 6.3, Hypothesis 7.1 is equivalent to

$$P_T = \begin{cases} S_0 L_u, & \text{if } i > \lfloor \beta \rfloor \\ S_0 u^i d^{n-i}, & \text{if } i \in \left[\lceil \gamma \rceil, \lfloor \beta \rfloor \right] \\ S_0 L_d, & \text{if } i < \lceil \gamma \rceil. \end{cases} \tag{7.1}$$

Hence the stock's logarithmic returns follow the following probability distribution

$$b_{truncated}(i) = \begin{cases} 0, & \text{if } V_T > S_0 L_u \\ \sum_{j=\lfloor \beta \rfloor + 1}^{n} b(j), & \text{if } V_T = S_0 L_u \\ b(i), & \text{if } V_T \in [S_0 L_d, S_0 L_u] \\ \sum_{j=0}^{\lceil \gamma \rceil - 1} b(j), & \text{if } V_T = S_0 L_d \\ 0, & \text{if } V_T < S_0 L_d, \end{cases} \tag{7.2}$$

where

$$b(i) := \frac{n!}{(n-i)!i!} p^i (1-p)^{n-i}.$$

Using Lemmas 6.1, 6.2 and 6.3, the above (also shown in Figure 7.1) can be re-written as

$$b_{truncated}(i) = \begin{cases} 0, & \text{if } i \geqslant \lfloor\beta\rfloor + 2 \\ \sum_{j=\lfloor\beta\rfloor+1}^{n} b(j), & \text{if } i = \lfloor\beta\rfloor + 1 \\ b(i), & \text{if } i \in \left[\lceil\gamma\rceil, \lfloor\beta\rfloor\right] \\ \sum_{j=0}^{\lceil\gamma\rceil-1} b(j), & \text{if } i = \lceil\gamma\rceil - 1 \\ 0, & \text{if } i \leqslant \lceil\gamma\rceil - 2. \end{cases} \tag{7.3}$$

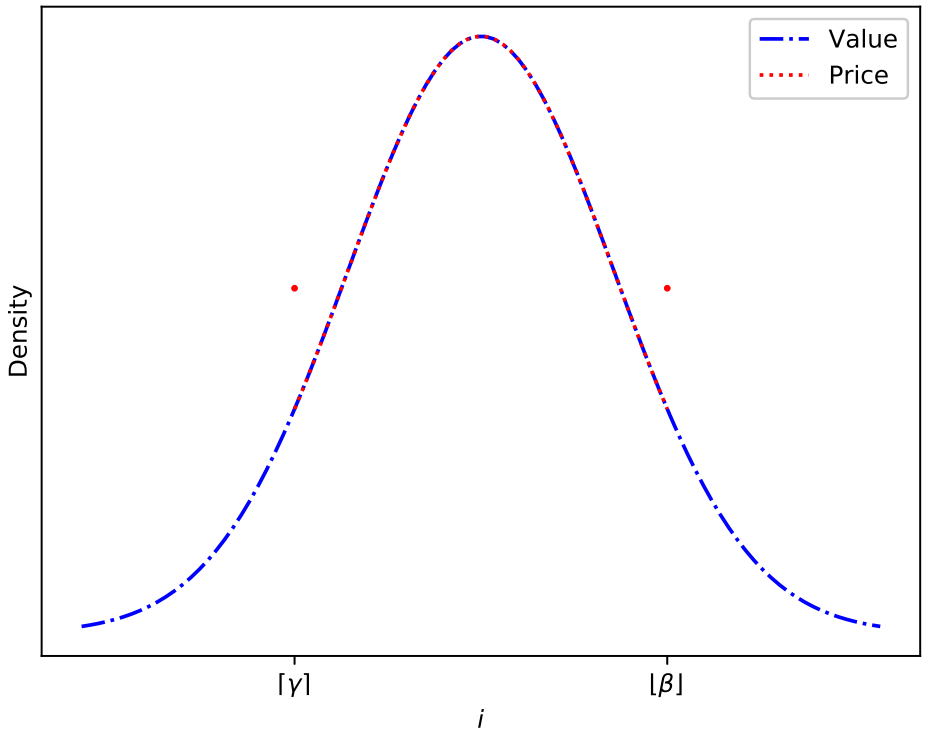


Figure 7.1: Probability density: economic value vs market price.

In this chapter, we will study a plain European call option for this kind of stocks. The option payoff formula is defined as

$$f_{\text{truncated payoff}}(i) := \begin{cases} f(i) := P_T - K, & \text{if } P_T > K \\ 0, & \text{if } P_T \leqslant K. \end{cases} \tag{7.4}$$

In the following sections, we will explore three possible situations for the strike price K.

7.2 If $K \in (S_0L_d, S_0L_u)$

According to CRR model, our call option is priced as

$$c = e^{-rT} \sum_{i=0}^{n} b_{truncated}(i) f_{\text{truncated payoff}}(i). \tag{7.5}$$

If $S_0 L_d < K < S_0 L_u$, it is obvious that

$$\frac{1}{2}\Big(\frac{\ln(L_d)}{\sigma\sqrt{T/n}} + n\Big) < \frac{1}{2}\Big(\frac{\ln(K/S_0)}{\sigma\sqrt{T/n}} + n\Big) < \frac{1}{2}\Big(\frac{\ln(L_u)}{\sigma\sqrt{T/n}} + n\Big). \tag{7.6}$$

According to (4.16), Lemmas 5.5 and 5.9, the above becomes

$$\gamma < \alpha < \beta. \tag{7.7}$$

Hence (7.5) becomes

$$\begin{aligned} c = e^{-rT}\Big[& \sum_{i=0}^{\lceil\gamma\rceil-2} b_{truncated}(i) f_{\text{truncated payoff}}(i) \\ & + b_{truncated}\big(\lceil\gamma\rceil - 1\big) f_{\text{truncated payoff}}\big(\lceil\gamma\rceil - 1\big) \\ & + \sum_{i=\lceil\gamma\rceil}^{\lfloor\alpha\rfloor} b_{truncated}(i) f_{\text{truncated payoff}}(i) \\ & + \sum_{i=\lfloor\alpha\rfloor+1}^{\lfloor\beta\rfloor} b_{truncated}(i) f_{\text{truncated payoff}}(i) \\ & + b_{truncated}\big(\lfloor\beta\rfloor + 1\big) f_{\text{truncated payoff}}\big(\lfloor\beta\rfloor + 1\big) \\ & + \sum_{i=\lfloor\beta\rfloor+2}^{n} b_{truncated}(i) f_{\text{truncated payoff}}(i)\Big]. \end{aligned} \tag{7.8}$$

Now we discuss these terms one by one. Firstly for any $i \in \big[0, \lceil\gamma\rceil - 2\big]$, according to (7.3),

$$b_{truncated}(i) = 0. \tag{7.9}$$

Hence

$$\sum_{i=0}^{\lceil\gamma\rceil-2} b_{truncated}(i) f_{\text{truncated payoff}}(i) = 0. \tag{7.10}$$

Secondly for any $i \in \big[\lceil\gamma\rceil - 1, \lfloor\alpha\rfloor\big]$, according to Lemmas 6.2 and 6.3,

$$V_T \in [S_0 L_d, K]. \tag{7.11}$$

Moreover, since $K \in (S_0 L_d, S_0 L_u)$, and according to Hypothesis 7.1,

$$P_T = V_T \in [S_0 L_d, K]. \tag{7.12}$$

Furthermore, according to (7.4),

$$f_{\text{truncated payoff}}(i) = 0. \tag{7.13}$$

Hence

$$b_{truncated}\big(\lceil\gamma\rceil - 1\big) f_{\text{truncated payoff}}\big(\lceil\gamma\rceil - 1\big) = 0 \tag{7.14}$$

and

$$\sum_{i=\lceil\gamma\rceil}^{\lfloor\alpha\rfloor} b_{truncated}(i) f_{\text{truncated payoff}}(i) = 0. \tag{7.15}$$

Thirdly for any $i \in \big[\lfloor\alpha\rfloor + 1, \lfloor\beta\rfloor\big]$, according to (7.3),

$$b_{truncated}(i) = b(i). \tag{7.16}$$

On the other hand, according to Lemmas 6.1 and 6.2,

$$V_T \in (K, S_0 L_u). \tag{7.17}$$

Since $K \in (S_0 L_d, S_0 L_u)$, and according to Hypothesis 7.1,

$$P_T = V_T \in (K, S_0 L_u). \tag{7.18}$$

In addition, according to (7.4),

$$f_{\text{truncated payoff}}(i) = f(i). \tag{7.19}$$

Hence

$$\sum_{i=\lfloor\alpha\rfloor+1}^{\lfloor\beta\rfloor} b_{truncated}(i) f_{\text{truncated payoff}}(i) = \sum_{i=\lfloor\alpha\rfloor+1}^{\lfloor\beta\rfloor} b(i) f(i). \tag{7.20}$$

Fourthly for $i = \lfloor\beta\rfloor + 1$, according to (7.3),

$$b_{truncated}(i) = \sum_{j=\lfloor\beta\rfloor+1}^{n} b(j). \tag{7.21}$$

On the other hand, according to Lemma 6.1,

$$V_T = S_0 L_u. \tag{7.22}$$

Since $K \in (S_0 L_d, S_0 L_u)$, and according to Hypothesis 7.1,

$$P_T = S_0 L_u. \tag{7.23}$$

Furthermore, according to (7.4),

$$f_{\text{truncated payoff}}(i) = S_0 L_u - K. \tag{7.24}$$

Hence

$$b_{truncated}\big(\lfloor\beta\rfloor+1\big)f_{\text{truncated payoff}}\big(\lfloor\beta\rfloor+1\big) = (S_0L_u-K)\times\sum_{j=\lfloor\beta\rfloor+1}^{n}b(j). \tag{7.25}$$

Fifthly for any $i\in\big[\lfloor\beta\rfloor+2,n\big]$, according to (7.3),

$$b_{truncated}(i)=0. \tag{7.26}$$

Hence

$$\sum_{i=\lfloor\beta\rfloor+2}^{n}b_{truncated}(i)f_{\text{truncated payoff}}(i)=0. \tag{7.27}$$

Combining all of the above, since n is large enough, we have

Theorem 7.1: *If $K\in(S_0L_d,S_0L_u)$, our option is priced as*

$$c=S_0\Phi(d_1)-Ke^{-rT}\Phi(d_2)-S_0\Phi(d_5)+e^{-rT}S_0L_u\Phi(d_6),$$

where

$$d_1=\frac{\ln(S_0/K)+(r+\sigma^2/2)T}{\sigma\sqrt{T}},$$

$$d_2=\frac{\ln(S_0/K)+(r-\sigma^2/2)T}{\sigma\sqrt{T}},$$

$$d_5=\frac{\ln(1/L_u)+(r+\sigma^2/2)T}{\sigma\sqrt{T}}$$

and

$$d_6=\frac{\ln(1/L_u)+(r-\sigma^2/2)T}{\sigma\sqrt{T}}.$$

Proof: According to Theorem 5.15,

$$\begin{aligned}&e^{-rT}\sum_{i=\lfloor\alpha\rfloor}^{\lfloor\beta\rfloor}b(i)f(i)\\ \longrightarrow&S_0\Phi(d_1)-Ke^{-rT}\Phi(d_2)-S_0\Phi(d_5)+Ke^{-rT}\Phi(d_6),\end{aligned}$$

as $n\longrightarrow\infty$, where

$$d_1=\frac{\ln(S_0/K)+(r+\sigma^2/2)T}{\sigma\sqrt{T}},$$

$$d_2 = \frac{\ln(S_0/K) + (r - \sigma^2/2)T}{\sigma\sqrt{T}},$$

$$d_5 = \frac{\ln(1/L_u) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}$$

and

$$d_6 = \frac{\ln(1/L_u) + (r - \sigma^2/2)T}{\sigma\sqrt{T}}.$$

Additionally, according to (5.19),

$$\sum_{j=\lfloor\beta\rfloor+1}^{n} b(j) \longrightarrow \Phi(d_6),$$

as $n \longrightarrow \infty$.

Hence (7.8) becomes

$$c = e^{-rT} \sum_{i=\lfloor\alpha\rfloor+1}^{\lfloor\beta\rfloor} b(i)f(i) + e^{-rT}(S_0L_u - K) \times \sum_{j=\lfloor\beta\rfloor+1}^{n} b(j).$$

According to (7.4) and Lemma 6.2, we obtain that

$$f\big(\lfloor\alpha\rfloor\big) = 0.$$

Hence

$$\begin{aligned} c \longrightarrow & e^{-rT} \sum_{i=\lfloor\alpha\rfloor}^{\lfloor\beta\rfloor} b(i)f(i) + e^{-rT}(S_0L_u - K) \sum_{j=\lfloor\beta\rfloor+1}^{n} b(j) \\ = & S_0\Phi(d_1) - Ke^{-rT}\Phi(d_2) - S_0\Phi(d_5) + Ke^{-rT}\Phi(d_6) \\ & + e^{-rT}(S_0L_u - K)\Phi(d_6) \\ = & S_0\Phi(d_1) - Ke^{-rT}\Phi(d_2) - S_0\Phi(d_5) + e^{-rT}S_0L_u\Phi(d_6), \end{aligned}$$

as $n \longrightarrow \infty$. □

7.3 If $K \leqslant S_0L_d$

If $K \leqslant S_0L_d$, and since $S_0L_d < S_0L_u$ it is obvious that

$$\frac{1}{2}\Big(\frac{\ln(K/S_0)}{\sigma\sqrt{T/n}} + n\Big) \leqslant \frac{1}{2}\Big(\frac{\ln(L_d)}{\sigma\sqrt{T/n}} + n\Big) < \frac{1}{2}\Big(\frac{\ln(L_u)}{\sigma\sqrt{T/n}} + n\Big). \tag{7.28}$$

According to (4.16), Lemmas 5.5 and 5.9, the above becomes

$$\alpha \leqslant \gamma < \beta. \tag{7.29}$$

As a result, according to CRR model, our call option is priced as

$$\begin{aligned} c &= e^{-rT} \sum_{i=0}^{n} b_{truncated}(i) f_{\text{truncated payoff}}(i) \qquad (7.30)\\ &= e^{-rT}\Big[\sum_{i=0}^{\lfloor\alpha\rfloor} b_{truncated}(i) f_{\text{truncated payoff}}(i) \\ &\quad + \sum_{i=\lfloor\alpha\rfloor+1}^{\lceil\gamma\rceil-2} b_{truncated}(i) f_{\text{truncated payoff}}(i) \\ &\quad + b_{truncated}\big(\lceil\gamma\rceil-1\big) f_{\text{truncated payoff}}\big(\lceil\gamma\rceil-1\big) \\ &\quad + \sum_{i=\lceil\gamma\rceil}^{\lfloor\beta\rfloor} b_{truncated}(i) f_{\text{truncated payoff}}(i) \\ &\quad + b_{truncated}\big(\lfloor\beta\rfloor+1\big) f_{\text{truncated payoff}}\big(\lfloor\beta\rfloor+1\big) \\ &\quad + \sum_{i=\lfloor\beta\rfloor+2}^{n} b_{truncated}(i) f_{\text{truncated payoff}}(i)\Big]. \end{aligned}$$

Now we discuss these terms one by one. Firstly for any $i < \lceil\gamma\rceil - 1$, according to (7.3),

$$b_{truncated}(i) = 0, \tag{7.31}$$

and since $\alpha \leqslant \gamma$, then

$$\sum_{i=0}^{\lfloor\alpha\rfloor} b_{truncated}(i) f_{\text{truncated payoff}}(i) = 0 \tag{7.32}$$

and

$$\sum_{i=\lfloor\alpha\rfloor+1}^{\lceil\gamma\rceil-2} b_{truncated}(i) f_{\text{truncated payoff}}(i) = 0. \tag{7.33}$$

Secondly for $i = \lceil\gamma\rceil - 1$, according to (7.3),

$$b_{truncated}(i) = \sum_{j=0}^{\lceil\gamma\rceil-1} b(j). \tag{7.34}$$

And according to Lemma 6.3,

$$V_T = S_0 L_d. \tag{7.35}$$

Since $K \leqslant S_0 L_d$, and according to Hypothesis 7.1,

$$P_T = S_0 L_d. \tag{7.36}$$

Furthermore, according to (7.4),

$$f_{\text{truncated payoff}}(i) = S_0 L_d - K. \tag{7.37}$$

Hence

$$\begin{aligned} & b_{truncated}\big(\lceil\gamma\rceil - 1\big) f_{\text{truncated payoff}}\big(\lceil\gamma\rceil - 1\big) \\ &= (S_0 L_d - K) \times \sum_{j=0}^{\lceil\gamma\rceil - 1} b(j). \end{aligned} \tag{7.38}$$

Thirdly for any $i \in \big[\lceil\gamma\rceil, \lfloor\beta\rfloor\big]$, according to (7.3),

$$b_{truncated}(i) = b(i). \tag{7.39}$$

On the other hand, according to Lemmas 6.1 and 6.3,

$$V_T \in (S_0 L_d, S_0 L_u). \tag{7.40}$$

Since $K \leqslant S_0 L_d$, and according to Hypothesis 7.1,

$$P_T = V_T \in (S_0 L_d, S_0 L_u). \tag{7.41}$$

Furthermore, since $P_T > S_0 L_d \geqslant K$, according to (7.4),

$$f_{\text{truncated payoff}}(i) = f(i). \tag{7.42}$$

Hence

$$\sum_{i=\lceil\gamma\rceil}^{\lfloor\beta\rfloor} b_{truncated}(i) f_{\text{truncated payoff}}(i) = \sum_{i=\lceil\gamma\rceil}^{\lfloor\beta\rfloor} b(i) f(i). \tag{7.43}$$

Fourthly For $i = \lfloor\beta\rfloor + 1$, according to (7.3),

$$b_{truncated}(i) = \sum_{j=\lfloor\beta\rfloor+1}^{n} b(j). \tag{7.44}$$

In addition, according to Lemma 6.1,

$$V_T = S_0 L_u. \tag{7.45}$$

According to Hypothesis 7.1,

$$P_T = S_0 L_u. \tag{7.46}$$

Furthermore, since $K \leqslant S_0 L_d < S_0 L_u = P_T$, and according to (7.4),

$$f_{\text{truncated payoff}}(i) = S_0 L_u - K. \tag{7.47}$$

Hence

$$\begin{aligned} & b_{truncated}\big(\lfloor\beta\rfloor\big) f_{\text{truncated payoff}}\big(\lfloor\beta\rfloor\big) \\ &= (S_0 L_u - K) \times \sum_{j=\lfloor\beta\rfloor+1}^{n} b(j). \end{aligned} \tag{7.48}$$

Fifthly for any $i > \lfloor\beta\rfloor + 1$, according (7.3),

$$b_{truncated}(i) = 0. \tag{7.49}$$

Hence

$$\sum_{i=\lfloor\beta\rfloor+2}^{n} b_{truncated}(i) f_{\text{truncated payoff}}(i) = 0. \tag{7.50}$$

All things considered, since n is large enough, therefore

Theorem 7.2: *If $K \leqslant S_0 L_d$, our option is priced as*

$$\begin{aligned} c =& e^{-rT}(S_0 L_d - K) - e^{-rT} S_0 L_d \Phi(d_4) \\ &+ S_0 \Phi(d_3) - S_0 \Phi(d_5) \\ &+ e^{-rT} S_0 L_u \Phi(d_6), \end{aligned}$$

where

$$d_3 = \frac{\ln(1/L_d) + (r + \sigma^2/2)T}{\sigma\sqrt{T}},$$

$$d_4 = \frac{\ln(1/L_d) + (r - \sigma^2/2)T}{\sigma\sqrt{T}},$$

$$d_5 = \frac{\ln(1/L_u) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}$$

and

$$d_6 = \frac{\ln(1/L_u) + (r - \sigma^2/2)T}{\sigma\sqrt{T}}.$$

Proof: According to Theorem 5.13,

$$\begin{aligned} & e^{-rT} \sum_{i=\lceil\gamma\rceil}^{\lfloor\beta\rfloor} b(i) f(i) \\ \longrightarrow & S_0 \Phi(d_3) - K e^{-rT} \Phi(d_4) - S_0 \Phi(d_5) + K e^{-rT} \Phi(d_6), \end{aligned}$$

as $n \longrightarrow \infty$, where

$$d_3 = \frac{\ln(1/L_d) + (r + \sigma^2/2)T}{\sigma\sqrt{T}},$$

$$d_4 = \frac{\ln(1/L_d) + (r - \sigma^2/2)T}{\sigma\sqrt{T}},$$

$$d_5 = \frac{\ln(1/L_u) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}$$

and

$$d_6 = \frac{\ln(1/L_u) + (r - \sigma^2/2)T}{\sigma\sqrt{T}}.$$

Moreover, according to (5.19),

$$\sum_{j=\lfloor\beta\rfloor+1}^{n} b(j) = \Phi(d_6).$$

And according to (5.31),

$$\sum_{j=0}^{\lceil\gamma\rceil-1} b(j) = 1 - \Phi(d_4). \tag{7.51}$$

Hence (7.30) becomes

$$\begin{aligned}
c =& e^{-rT}(S_0 L_d - K) \sum_{j=0}^{\lceil\gamma\rceil-1} b(j) \\
&+ e^{-rT} \sum_{i=\lceil\gamma\rceil}^{\lfloor\beta\rfloor} b(i) f(i) \\
&+ e^{-rT}(S_0 L_u - K) \sum_{j=\lfloor\beta\rfloor+1}^{n} b(j) \\
\longrightarrow& e^{-rT}(S_0 L_d - K)(1 - \Phi(d_4)) \\
&+ S_0\Phi(d_3) - Ke^{-rT}\Phi(d_4) - S_0\Phi(d_5) + Ke^{-rT}\Phi(d_6) \\
&+ e^{-rT}(S_0 L_u - K)\Phi(d_6) \\
=& e^{-rT}(S_0 L_d - K) - e^{-rT} S_0 L_d \Phi(d_4) \\
&+ S_0\Phi(d_3) - S_0\Phi(d_5) \\
&+ e^{-rT} S_0 L_u \Phi(d_6),
\end{aligned} \tag{7.52}$$

as $n \longrightarrow \infty$. □

7.4 If $K \geqslant S_0 L_u$

According to CRR model, our call option is priced as

$$c = e^{-rT} \sum_{i=0}^{n} b_{truncated}(i) f_{\text{truncated payoff}}(i). \tag{7.53}$$

If $K \geqslant S_0 L_u$, and since $S_0 L_d < S_0 L_u$ it is obvious that

$$\frac{1}{2}\Big(\frac{\ln(L_d)}{\sigma\sqrt{T/n}} + n\Big) < \frac{1}{2}\Big(\frac{\ln(L_u)}{\sigma\sqrt{T/n}} + n\Big) \leqslant \frac{1}{2}\Big(\frac{\ln(K/S_0)}{\sigma\sqrt{T/n}} + n\Big). \tag{7.54}$$

According to (4.16), Lemmas 5.5 and 5.9, the above becomes

$$\gamma < \beta \leqslant \alpha. \tag{7.55}$$

Furthermore, according to Hypothesis 7.1,

$$P_T \leqslant S_0 L_u. \tag{7.56}$$

Since $K \geqslant S_0 L_u$, then

$$f_{\text{truncated payoff}}(i) = 0. \tag{7.57}$$

Therefore, if $K \geqslant S_0 L_u$, then the option has value

$$c = 0. \tag{7.58}$$

7.5 Summary of re-distributed model

Finally, we summarise our findings for a call option which satisfies (7.3), (7.4) and Hypothesis 7.1 as follows:

Theorem 7.3: *If $K \in (S_0 L_d, S_0 L_u)$, the option is priced as*

$$c = S_0 \Phi(d_1) - K e^{-rT} \Phi(d_2) - S_0 \Phi(d_5) + e^{-rT} S_0 L_u \Phi(d_6).$$

If $K \leqslant S_0 L_d$, the option is priced as

$$\begin{aligned} c =& e^{-rT}(S_0 L_d - K) - e^{-rT} S_0 L_d \Phi(d_4) \\ &+ S_0 \Phi(d_3) - S_0 \Phi(d_5) \\ &+ e^{-rT} S_0 L_u \Phi(d_6). \end{aligned}$$

If $K \geqslant S_0 L_u$, then the option has value

$$c = 0,$$

where

$$d_1 = \frac{\ln(S_0/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}},$$

$$d_2 = \frac{\ln(S_0/K) + (r - \sigma^2/2)T}{\sigma\sqrt{T}},$$

$$d_3 = \frac{\ln(1/L_d) + (r + \sigma^2/2)T}{\sigma\sqrt{T}},$$

$$d_4 = \frac{\ln(1/L_d) + (r - \sigma^2/2)T}{\sigma\sqrt{T}},$$

$$d_5 = \frac{\ln(1/L_u) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}$$

and

$$d_6 = \frac{\ln(1/L_u) + (r - \sigma^2/2)T}{\sigma\sqrt{T}}.$$

7.6 Delta hedging portfolios

In finance, investors could use delta hedging strategy to set up a portfolio by purchasing Δ shares and short selling 1 corresponding call option contract. The portfolio value should remain unchanged if the underlying stock price changes a little (Hull 2015). The delta Δ here is defined as

$$\Delta := \frac{\partial c}{\partial S_0}. \tag{7.59}$$

For BS model, Δ equals $\Phi(d_1)$ (Hull 2015, p. 404).
For our re-distributed model, it is clear that if $K \leqslant S_0 L_d$,

$$\begin{aligned}\Delta :=& \frac{\partial c}{\partial S_0} \\ =& e^{-rT} L_d - e^{-rT} L_d \Phi(d_4) + \Phi(d_3) - \Phi(d_5) + e^{-rT} L_u \Phi(d_6).\end{aligned}$$

If $K \in (S_0 L_d, S_0 L_u)$,

$$\begin{aligned}\Delta :=& \frac{\partial c}{\partial S_0} \\ =& \Phi(d_1) - \Phi(d_5) + e^{-rT} L_u \Phi(d_6).\end{aligned}$$

If $K \geqslant S_0 L_u$,

$$\Delta = 0.$$

According to GBM assumption (see Section 4.1), the logarithmic return of the portfolio satisfies

$$\ln\Big(\frac{S_T}{S_0}\Big) \sim \mathcal{N}\Big((r - \frac{\sigma^2}{2})T, \sigma^2 T\Big). \tag{7.60}$$

If the stock price follows a geometric Brownian motion, it is clear that BS model can be used here. For BS model, the corresponding risk-free rate r_{bs} and volatility σ_{bs} should satisfy

$$\begin{cases} r_{bs} - \dfrac{\sigma_{bs}^2}{2} = \mu_{historical} \\ \sigma_{bs} = \sigma_{historical}, \end{cases} \tag{7.61}$$

where the actual historical drift rate $\mu_{historical}$ and volatility $\sigma_{historical}$ are observable.

Solving the above, we have

$$\begin{cases} r_{bs} = \mu_{historical} + \dfrac{\sigma_{historical}^2}{2} \\ \sigma_{bs} = \sigma_{historical}. \end{cases} \tag{7.62}$$

We solve the formulas to obtain risk-free rate, instead of collecting the risk-free rate data. Hence what we get here is the implied risk-free rate.

If the stock satisfies the assumptions of the re-distributed model, our RD model should be used here. For RD model, using Newton's method (Burden & Faires 2011, pp. 638-644), we can solve the values of

$$\begin{cases} \mu_{rd}(\mu, \sigma) = \mu_{historical} \\ \sigma_{rd}^2(\mu, \sigma) = \sigma_{historical}^2, \end{cases} \tag{7.63}$$

where $\mu_{underlying}$ is the mean and $\sigma_{underlying}^2$ is the variance of an underlying normal distribution for

$$\begin{aligned} &\mu_{rd}(\mu_{underlying}, \sigma_{underlying}) \\ =&\mu_{underlying}\left[\Phi\left(\frac{\ln(1.1) - \mu_{underlying}}{\sigma_{underlying}}\right) - \Phi\left(\frac{\ln(0.9) - \mu_{underlying}}{\sigma_{underlying}}\right)\right] \\ &+\sigma_{underlying}\left[\varphi\left(\frac{\ln(0.9) - \mu_{underlying}}{\sigma_{underlying}}\right) - \varphi\left(\frac{\ln(1.1) - \mu_{underlying}}{\sigma_{underlying}}\right)\right] \\ &+\Phi\left(\frac{\ln(0.9) - \mu_{underlying}}{\sigma_{underlying}}\right)\ln(0.9) \\ &+\left[1 - \Phi\left(\frac{\ln(1.1) - \mu_{underlying}}{\sigma_{underlying}}\right)\right]\ln(1.1) \end{aligned} \tag{7.64}$$

and

$$
\begin{aligned}
&\sigma_{rd}^2(\mu_{underlying}, \sigma_{underlying}) \qquad (7.65)\\
=&\sigma_{underlying}^2 \left[\frac{\ln(0.9) - \mu_{underlying}}{\sigma_{underlying}} \cdot \varphi\left(\frac{\ln(0.9) - \mu_{underlying}}{\sigma_{underlying}} \right) \right.\\
&\quad - \frac{\ln(1.1) - \mu_{underlying}}{\sigma_{underlying}} \cdot \varphi\left(\frac{\ln(1.1) - \mu_{underlying}}{\sigma_{underlying}} \right)\\
&\quad \left. + \Phi\left(\frac{\ln(1.1) - \mu_{underlying}}{\sigma_{underlying}} \right) - \Phi\left(\frac{\ln(0.9) - \mu_{underlying}}{\sigma_{underlying}} \right) \right]\\
&+ 2\mu_{underlying}\sigma_{underlying} \left[\varphi\left(\frac{\ln(0.9) - \mu_{underlying}}{\sigma_{underlying}} \right) \right.\\
&\quad \left. - \varphi\left(\frac{\ln(1.1) - \mu_{underlying}}{\sigma_{underlying}} \right) \right]\\
&+ \mu_{underlying}^2 \left[\Phi\left(\frac{\ln(1.1) - \mu_{underlying}}{\sigma_{underlying}} \right) - \Phi\left(\frac{\ln(0.9) - \mu_{underlying}}{\sigma_{underlying}} \right) \right]\\
&+ [\ln(0.9)]^2 \Phi\left(\frac{\ln(0.9) - \mu_{underlying}}{\sigma_{underlying}} \right)\\
&+ [\ln(1.1)]^2 \left[1 - \Phi\left(\frac{\ln(1.1) - \mu_{underlying}}{\sigma_{underlying}} \right) \right]\\
&- \left[\mu_{underlying} \left[\Phi\left(\frac{\ln(1.1) - \mu_{underlying}}{\sigma_{underlying}} \right) - \Phi\left(\frac{\ln(0.9) - \mu_{underlying}}{\sigma_{underlying}} \right) \right] \right.\\
&\quad + \sigma_{underlying} \left[\varphi\left(\frac{\ln(0.9) - \mu_{underlying}}{\sigma_{underlying}} \right) - \varphi\left(\frac{\ln(1.1) - \mu_{underlying}}{\sigma_{underlying}} \right) \right]\\
&\quad + \Phi\left(\frac{\ln(0.9) - \mu_{underlying}}{\sigma_{underlying}} \right) \ln(0.9)\\
&\left. + \left[1 - \Phi\left(\frac{\ln(1.1) - \mu_{underlying}}{\sigma_{underlying}} \right) \right] \ln(1.1) \right]^2.
\end{aligned}
$$

The above μ_{rd} and σ_{rd} will be explained in Chapter 9. Then solving $r_{underlying}$, we have

$$r_{underlying} = \mu_{underlying} + \frac{\sigma_{underlying}^2}{2}. \qquad (7.66)$$

Now we have r_{bs} and σ_{bs} for BS model as well as $r_{underlying}$ and $\sigma_{underlying}$ for our RD model. We can set up two portfolios (called BS hedging portfolio and RD hedging portfolio respectively) using delta hedging strategy. These two portfolios satisfy different assumptions. We compare these portfolios, using the raw data we collected in Part I. On rolling base, we analyse 1110048 day trading cases for 186 companies whose listing history is longer than 20 years. On each day, we use historical 20 trading days' data to calculate $\mu_{historical}$ and $\sigma_{historical}$. Since our RD model is a 1-phase model (more details will be explained in Section 8.1), we hold every portfolios for 1 day. Now we run a one-sample t-test for the two portfolios separately. We use the hypotheses

$$\begin{cases} H_0: & \text{expected return of the portfolio equals risk-free rate} \\ H_1: & \text{expected return of the portfolio doesn't equal risk-free rate,} \end{cases}$$

at alpha level of 0.05. Since the risk-free return is not constant, we subtract the corresponding risk-free rate from the portfolio return on each case. So in our t-tests, we test that portfolio return minus risk-free rate equals zero. The results are broke down into Table 7.1.

	expected value of difference of portfolio return minus risk-free rate	p-value
BS hedging portfolio	5.38×10^{-4}	0.073
RD hedging portfolio	5.20×10^{-4}	0.083

Table 7.1: Delta hedging t-test results (comparing RD model with BS model).

For RD hedging portfolio, since the p-value is larger than 0.05, we cannot reject H_0. Therefore using our RD model, the delta hedging technique is effective. The value of RD hedging portfolio remains unchanged. Furthermore the expect value of difference of portfolio return minus risk-free rate for RD hedging portfolio is less than the one for BS hedging portfolio. And the p-value for RD hedging portfolio test is greater than the p-value for BS hedging portfolio test. Therefore our RD model has better performance to keep the portfolio value unchanged.

Chapter 8

Multi-phase model

In previous chapters, we have assumed that there are price bounds (lower bound S_0L_d and upper bound S_0L_u on the price movements of the underlying asset) during time $t \in [0, T]$ and introduced the re-distributed model for the options whose life time is T. Since the unchanged boundaries are applied from 0 to T, the re-distributed model can be considered as a single-phase model (more details will be explained in Section 8.1). In this chapter, we assume that there are multiple phases. For an m-phase case, there are price bounds $S_{j-1}L_d$ and $S_{j-1}L_u$ on the j-th phase, where S_{j-1} is the stock price at the end of the $(j-1)$-th phase and $j \in \mathbb{Z}^+ \cap [1, m]$. Then we set up a n-step binomial tree for each phase.

For the binomial tree on the j-th phase, according to Hypothesis 7.1,

Hypothesis 8.1: *the stock prices at the end of the phases satisfy*

$$\frac{S_j(i_j; u_j, d_j, n, \gamma, \beta)}{S_{j-1}} := \begin{cases} u_j^{\lceil\gamma\rceil} d_j^{n-\lceil\gamma\rceil}, & \text{if } i_j < \lceil\gamma\rceil \\ u_j^{i_j} d_j^{n-i_j}, & \text{if } i_j \in [\lceil\gamma\rceil, \lfloor\beta\rfloor] \\ u_j^{\lfloor\beta\rfloor} d_j^{n-\lfloor\beta\rfloor}, & \text{if } i_j > \lfloor\beta\rfloor. \end{cases}$$

According to (4.10),

Hypothesis 8.2: *the corresponding probabilities satisfy*

$$P_j(i_j; p_j, n) := \frac{n}{(n-i_j)i_j} p_j^{i_j} (1-p_j)^{n-i_j},$$

where $i_j \in \mathbb{Z} \cap [0, n]$ is the number of upwards steps in the j-th phase, $j \in \mathbb{Z}^+ \cap [1, m]$,

$$\gamma := \frac{1}{2}\Big(\frac{\ln(L_d)}{\sigma\sqrt{T_j/n}} + n\Big),$$

$$\beta := \frac{1}{2}\Big(\frac{\ln(L_u)}{\sigma\sqrt{T_j/n}} + n\Big)$$

and u_j, d_j, p_j, T_j are the corresponding variables for each phase. Since the initial price S_0 is given, then the probability at the beginning is $P_0 = 1$.

Hence according to the above recursive definitions, at end of the m-th phase, the stock price is

$$S_m(i_1,\cdots,i_m)=\frac{S_m(i_m)}{S_{m-1}(i_{m-1})}\times\cdots\times\frac{S_1(i_1)}{S_0}\times S_0. \tag{8.1}$$

And the corresponding probability is

$$\begin{aligned}P_m(i_1,\cdots,i_m)=&\frac{n!}{(n-i_1)!i_1!}p^{i_1}(1-p)^{n-i_1}\times\\&\cdots\times\frac{n!}{(n-i_m)!i_m!}p^{i_m}(1-p)^{n-i_m}.\end{aligned} \tag{8.2}$$

To simplify the analysis, we assume that the risk-free interest rate is unchanged during $[0,T]$, i.e.

$$r_1=\cdots=r_m=r \tag{8.3}$$

and the time lengths for each phase are the same, i.e.

$$T_1=\cdots=T_m=\frac{T}{m}. \tag{8.4}$$

According to (4.12), a plain European call option whose payoff formula satisfies

$$\begin{cases}S_m-K, & \text{if } S_m>K\\ 0, & \text{if } S_m\leqslant K,\end{cases} \tag{8.5}$$

is priced as

$$\begin{aligned}c:=&e^{-rT}\sum_{i_1=0}^{n}\cdots\sum_{i_m=0}^{n}\Big[P_m(i_1,\cdots,i_m)\times\max\big(S_m(i_1,\cdots,i_m)-K\big)\Big] \\ =&e^{-rT}\sum_{i_1=0}^{n}\cdots\sum_{i_m=0}^{n}\Bigg[\frac{n!}{(n-i_1)!i_1!}p^{i_1}(1-p)^{n-i_1}\times\\&\qquad\cdots\times\frac{n!}{(n-i_m)!i_m!}p^{i_m}(1-p)^{n-i_m}\\&\qquad\times\max\left(S_0\times\Big[\frac{S_m(i_m)}{S_{m-1}(i_{m-1})}\Big]\times\cdots\times\Big[\frac{S_1(i_1)}{S_0}\Big]-K,0\right)\Bigg].\end{aligned} \tag{8.6}$$

It is clear that the stock price at the end of m-th phase is

$$S_m=S_0\times\Big[\frac{S_m}{S_{m-1}}\Big]\times\cdots\times\Big[\frac{S_1}{S_0}\Big] \tag{8.7}$$

in (8.6). Since the initial stock price S_0 is given and $\dfrac{S_j}{S_{j-1}}$, where $j\in\mathbb{Z}^+\cap[1,m]$ is defined, S_m is achieved recursively.

8.1 1-phase model: $m = 1$

If $m = 1$, it is a single-phase case. Then (8.6) becomes

$$c = e^{-rT} \sum_{i_1=0}^{n} \Bigg[\frac{n!}{(n-i_1)!i_1!} p^{i_1} (1-p)^{n-i_1} \times \max\Big(S_0 \times \frac{S_1(i_1)}{S_0} - K, 0 \Big) \Bigg]. \tag{8.8}$$

According to Hypothesis 8.1, the following 5 intervals should be analysed:

1. $S_1 < S_0 u_1^{\lceil \gamma \rceil} d_1^{n-\lceil \gamma \rceil}$.
2. $S_1 = S_0 u_1^{\lceil \gamma \rceil} d_1^{n-\lceil \gamma \rceil}$.
3. $S_1 \in \big(S_0 u_1^{\lceil \gamma \rceil} d_1^{n-\lceil \gamma \rceil}, S_0 u_1^{\lfloor \beta \rfloor} d_1^{n-\lfloor \beta \rfloor} \big)$.
4. $S_1 = S_0 u_1^{\lfloor \beta \rfloor} d_1^{n-\lfloor \beta \rfloor}$.
5. $S_1 > S_0 u_1^{\lfloor \beta \rfloor} d_1^{n-\lfloor \beta \rfloor}$,

where

$$\gamma := \frac{1}{2} \Big(\frac{\ln(L_d)}{\sigma \sqrt{T_j/n}} + n \Big), \tag{8.9}$$

and

$$\beta := \frac{1}{2} \Big(\frac{\ln(L_u)}{\sigma \sqrt{T_j/n}} + n \Big). \tag{8.10}$$

For the 1-phase case, it is clear that if n is large enough, then

$$S_0 u_1^{\lceil \gamma \rceil} d_1^{n-\lceil \gamma \rceil} \approx S_0 L_d \tag{8.11}$$

and

$$S_0 u_1^{\lfloor \beta \rfloor} d_1^{n-\lfloor \beta \rfloor} \approx S_0 L_u. \tag{8.12}$$

Furthermore since $L_d < L_u$, then for $S_0 L_d$, $S_0 L_u$ and K, there are 3 possible situations:

- $K < S_0 L_d$.
- $K \in [S_0 L_d, S_0 L_u]$.
- $K > S_0 L_u$.

Now the 1-phase model becomes the re-distributed model. A closed-form solution is derived in Chapter 7.

8.2 2-phase model: $m = 2$

If $m \geqslant 2$, some very complex closed formulas can be obtained but loses practical sense (see Appendix A). Hence recursive models are applied and we have to compute (8.6) numerically. For $m = 2$, (8.6) becomes

$$c = e^{-rT} \sum_{i_1=0}^{n} \sum_{i_2=0}^{n} \left[\frac{n!}{(n-i_1)!i_1!} p^{i_1} (1-p)^{n-i_1} \times \frac{n!}{(n-i_2)!i_2!} p^{i_2} (1-p)^{n-i_2} \right.$$
$$\left. \times \max\left(S_0 \times \frac{S_1(i_1)}{S_0} \times \frac{S_2(i_2)}{S_1(i_1)} - K, 0 \right) \right].$$

We explore a simple case here, where $S_0 = 1$, $K = 1$, $r = 0.03$, $\sigma = 0.3$, $T = 1$, $L_d = 0.9$ and $L_u = 1.1$. According to Theorem 5.4, to set up a binomial tree, the number of steps for each phase (n in this chapter) must satisfy

$$\sqrt{n} \geqslant \max\left(\left| \frac{\ln(L_d)}{\sigma\sqrt{T/m}} \right|, \left| \frac{\ln(L_u)}{\sigma\sqrt{T/m}} \right|, \left| \frac{\ln(K/S_0)}{\sigma\sqrt{T/m}} \right| \right). \tag{8.13}$$

For our case, we must have $n \geqslant 1$. For different n, the time (in minutes) taken by the calculation is plotted in Figure 8.1. The computation time curve suggests that a small n should be used for fast calculation, as the computation time raises very quickly.

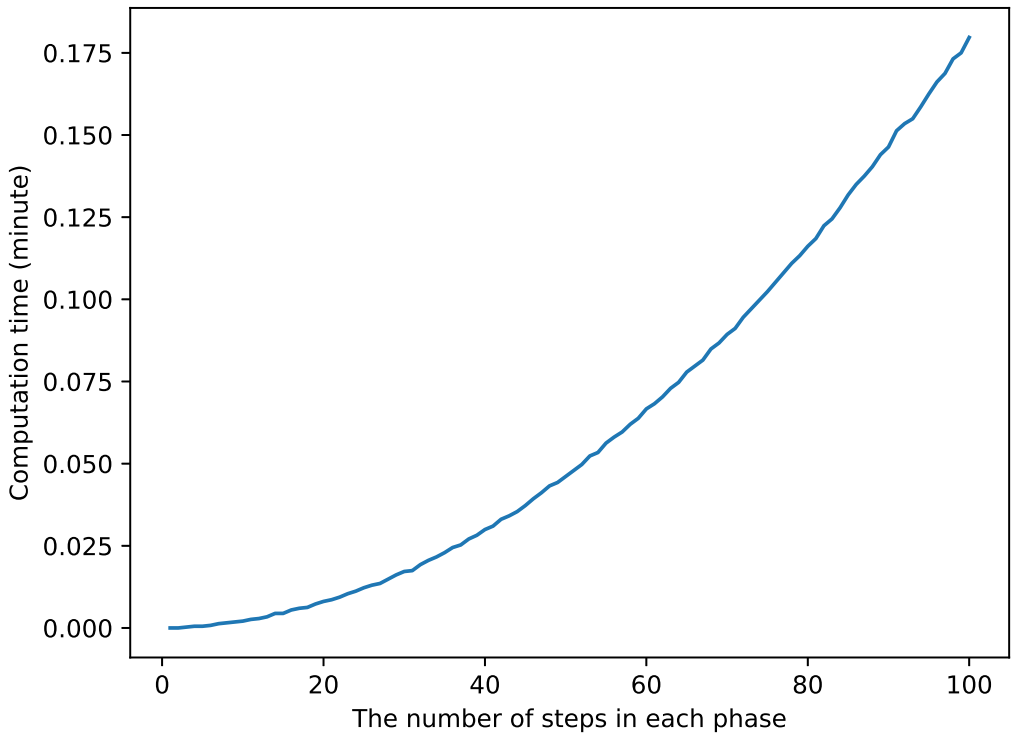


Figure 8.1: 2-phase model results for different number of steps: the computation time.

8.3 3- and 4-phase model: $m \in \{3, 4\}$

It is clear that for $m = 3$, (8.6) becomes

$$
\begin{aligned}
c = e^{-rT} \sum_{i_1=0}^{n} \sum_{i_2=0}^{n} \sum_{i_3=0}^{n} \Bigg[& \frac{n!}{(n-i_1)!i_1!} p^{i_1} (1-p)^{n-i_1} \\
& \times \frac{n!}{(n-i_2)!i_2!} p^{i_2} (1-p)^{n-i_2} \\
& \times \frac{n!}{(n-i_3)!i_3!} p^{i_3} (1-p)^{n-i_3} \\
& \times \max\left(S_0 \times \frac{S_1(i_1)}{S_0} \times \frac{S_2(i_2)}{S_1(i_1)} \times \frac{S_3(i_3)}{S_2(i_2)} - K, 0 \right) \Bigg].
\end{aligned}
\tag{8.14}
$$

And for $m = 4$, (8.6) becomes

$$
\begin{aligned}
c = e^{-rT} \sum_{i_1=0}^{n} \sum_{i_2=0}^{n} \sum_{i_3=0}^{n} \sum_{i_4=0}^{n} \Bigg[& \frac{n!}{(n-i_1)!i_1!} p^{i_1} (1-p)^{n-i_1} \\
& \times \frac{n!}{(n-i_2)!i_2!} p^{i_2} (1-p)^{n-i_2} \\
& \times \frac{n!}{(n-i_3)!i_3!} p^{i_3} (1-p)^{n-i_3} \\
& \times \frac{n!}{(n-i_4)!i_4!} p^{i_4} (1-p)^{n-i_4} \\
& \times \max\left(S_0 \times \frac{S_1(i_1)}{S_0} \times \frac{S_2(i_2)}{S_1(i_1)} \right. \\
& \left. \times \frac{S_3(i_3)}{S_2(i_2)} \times \frac{S_4(i_4)}{S_3(i_3)} - K, 0 \right) \Bigg].
\end{aligned}
\tag{8.15}
$$

The computation time results are represented in Figure 8.2 and 8.3 respectively. Comparing the two computation time curves, it suggests that larger m (the number of phases) implies rapidly increasing time for computation. For the 3-phase model, 100 steps (approximately 33 steps per phase) can be finished in less than 1 minute. But for the 4-phase model, an 100-step (25 steps per phase) calculation requires more than 10 minutes. There is clearly a huge difference. It is obvious that the factor m contributes much more than other factors. We will explore the relationship between the time required and the number of phases in the following section.

Since we don't have a practical closed-form formula for $m > 2$, we are not able to find the exact mean and variance values. Hence no BS model can be used in the analysis for comparative purposes. However we do have a corresponding BS formula for the single-phase model, we will use it to introduce a recursive BS model in the next chapter to serve as a comparative model.

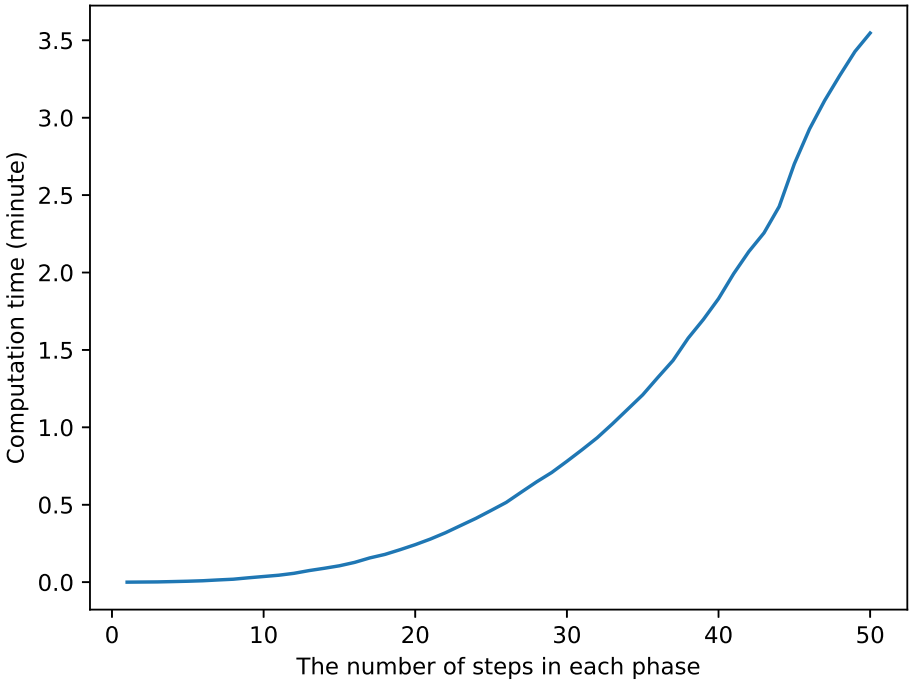


Figure 8.2: 3-phase model results for different number of steps: the computation time.

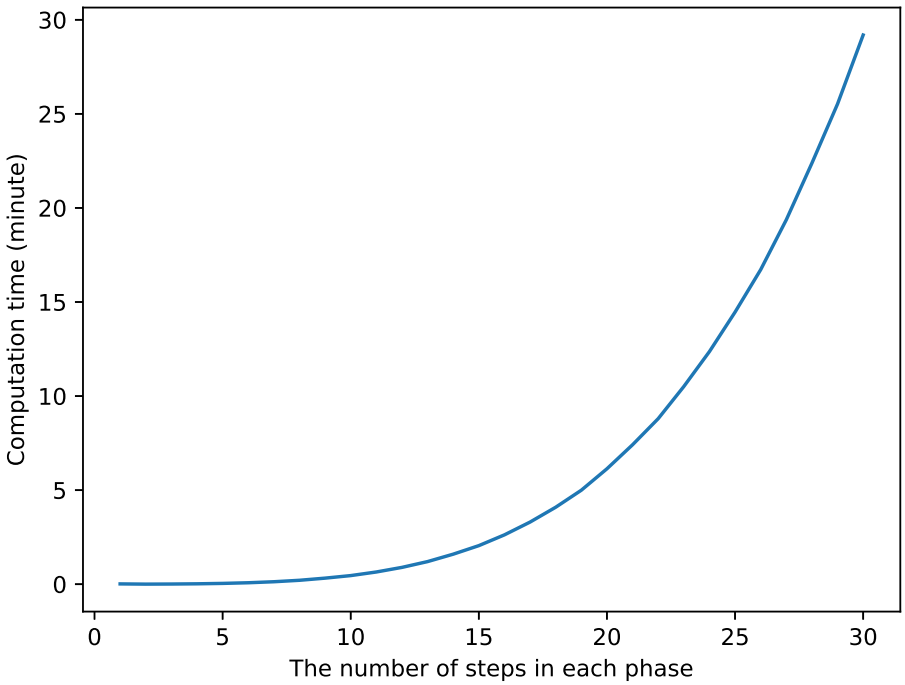


Figure 8.3: 4-phase model results for different number of steps: the computation time.

8.4 The computation time for an m-phase model: 2-step-per-phase models

In this section, we observe the computation times for different number of phases. We fix the number of steps per phase at $n = 2$ as an example. Of course, for computational purposes, it is obvious that we should have chosen $n > 20$.

Figure 8.4 suggests that even for a small n ($n = 2$ in our case), the recursive model requires a very long computation time. The computation time goes up rapidly after $m = 5$. With this in mind, for real-world analyses, m is much larger. For example, to evaluate a 3-month option on the Chinese financial market, because of the daily price bounds on each of the approximately 60 trading days, m is about 60. Therefore this recursive model will pose challenges when used to evaluate long-term options.

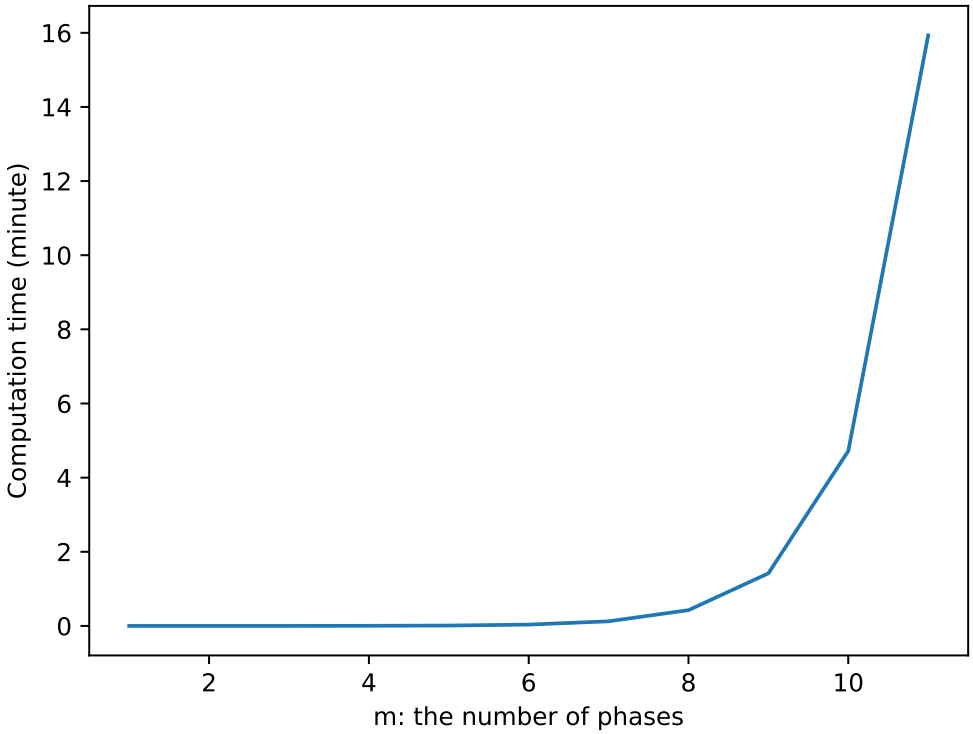


Figure 8.4: The time taken by the m-phase model: the number of steps per phase $n = 2$.

Our study suggests that a usable recursive model requires low n values as well as low m values. However, since both larger m and larger n are desired in modelling real-world situations, we will make a tentative discussion to look at possible alternatives.

Chapter 9

An m-phase recursive Black-Scholes model

In this chapter, to reduce computation time, we will introduce a recursive Black-Scholes (BS) model, and then compare the multi-phase model defined in Chapter 8 with it. In the plain BS model approach introduced in Chapter 4, we assumed that the stock price S_t follows a geometric Brownian motion. Hence its logarithmic return follows a normal distribution, i.e.

$$\ln\Big(\frac{S_T}{S_0}\Big) \sim \mathcal{N}\Big((r-\frac{\sigma^2}{2})T, \sigma^2 T\Big), \tag{9.1}$$

where initial stock price S_0, risk-free rate r, volatility σ and option life time T are given. As we explained in Chapter 5, $\mathcal{N}\Big((r-\frac{\sigma^2}{2})T, \sigma^2 T\Big)$ is also the underlying normal distribution of the re-distributed model (defined in Chapter 7).

For an m-phase BS model, according to (8.4), the time lengths on each phase are equivalent, i.e.

$$T_1 = \cdots = T_m = \frac{T}{m}, \tag{9.2}$$

then within each phase, the logarithmic return is

$$\ln\Big(\frac{S_j}{S_{j-1}}\Big) \sim \mathcal{N}\Big((r-\frac{\sigma^2}{2})\frac{T}{m}, \sigma^2\frac{T}{m}\Big), \tag{9.3}$$

where S_j is the stock price at the end of the j-th phase, S_{j-1} equals the stock price at the beginning of the j-th phase and $j \in \mathbb{Z}^+ \cap [1, m]$.

To understand more about the truncated models, in the following sections, we will analyse a part of above normal distribution as well as the mean and variance for the truncated distribution.

9.1 Calculation of $\int_a^b x\frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right)dx$

In this section, we will explore definite integrals of normal distribution probability density function over an interval $[a, b]$. For $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$ and $a, b \in \mathbb{R}$, $a < b$, to simplify the expressions in this

chapter, we set

- $\alpha := \dfrac{a-\mu}{\sigma}$.
- $\beta := \dfrac{b-\mu}{\sigma}$.
- $f(t) := t\sigma + \mu$.

Then it is clear that $a = \alpha\sigma + \mu =: f(\alpha)$ and $b = \beta\sigma + \mu =: f(\beta)$. We also set the notation for the area over $[a, b]$ as

Definition 9.1: The area of $\mathcal{N}(\mu, \sigma^2)$ over the interval $[a, b]$ is

$$\begin{aligned} Z :=& \int_a^b \frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right) dx \\ =& \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right), \end{aligned}$$

where

$$\varphi(x) := \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

is the probability density function of standard normal distribution.

Since

Lemma 9.1: *The indefinite integral of the function $x\varphi(x)$ is:*

$$\int x\varphi(x)dx = -\varphi(x) + C,$$

where C is the constant of integration.

Proof:

$$\begin{aligned} & \int x\varphi(x)dx \\ =& \int x\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \\ =& \frac{1}{\sqrt{2\pi}} \int x e^{-\frac{x^2}{2}} dx. \end{aligned}$$

Substituting $u := -\dfrac{x^2}{2}$ yields $du = -xdx$ and

$$\begin{aligned}
&\frac{1}{\sqrt{2\pi}}\int xe^{-\frac{x^2}{2}}dx\\
=&\frac{1}{\sqrt{2\pi}}\int -e^u du\\
=&-\frac{1}{\sqrt{2\pi}}e^u + C\\
=&-\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} + C\\
=&-\varphi(x) + C.
\end{aligned}$$

□

and

Corollary 9.2: *the definite integral of the* $x\varphi(x)$ *over* $[a,b]$ *is*

$$\begin{aligned}
&\int_\alpha^\beta x\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx\\
=&-\left[\varphi(x)\right]_\alpha^\beta\\
=&\varphi(\alpha)-\varphi(\beta)\\
=&\varphi\left(\frac{a-\mu}{\sigma}\right)-\varphi\left(\frac{b-\mu}{\sigma}\right).
\end{aligned}$$

we have

Lemma 9.3:

$$\int_a^b \frac{x-\mu}{\sigma}\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx = \sigma\left[\varphi\left(\frac{a-\mu}{\sigma}\right)-\varphi\left(\frac{b-\mu}{\sigma}\right)\right].$$

Proof: Substituting $\alpha := \frac{a-\mu}{\sigma}$, $\beta := \frac{b-\mu}{\sigma}$ and $f(t) := t\sigma+\mu$ yield

$$\begin{aligned}
&\int_a^b \frac{x-\mu}{\sigma}\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx\\
=&\int_{\alpha\sigma+\mu}^{\beta\sigma+\mu}\frac{x-\mu}{\sigma}\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx\\
=&\int_\alpha^\beta t\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}\sigma dt\\
=&\sigma\int_\alpha^\beta t\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt.
\end{aligned}$$

According to Corollary 9.2, the above becomes

$$\sigma\left[\varphi\left(\frac{a-\mu}{\sigma}\right)-\varphi\left(\frac{b-\mu}{\sigma}\right)\right].$$

□

Therefore we have

Theorem 9.4: *Let* $X\sim\mathcal{N}\left(\mu,\sigma^2\right)$, $a,b\in\mathbb{R}$, $a\leqslant b$,

$$\begin{aligned}&\int_a^b x\frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right)dx\\ =&\mu\left[\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)\right]\\ &+\sigma\left[\varphi\left(\frac{a-\mu}{\sigma}\right)-\varphi\left(\frac{b-\mu}{\sigma}\right)\right].\end{aligned}$$

Proof:

$$\begin{aligned}=&\int_a^b x\frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right)dx\\ =&\int_a^b x\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx\\ =&\int_a^b \frac{x-\mu}{\sigma}\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx\\ &+\int_a^b \frac{\mu}{\sigma}\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx\\ =&\int_a^b \frac{x-\mu}{\sigma}\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx\\ &+\mu\int_a^b \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx\\ =&\int_a^b \frac{x-\mu}{\sigma}\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx+\mu Z.\end{aligned}$$

According to Lemma 9.3, the above becomes

$$\begin{aligned}&\mu\left[\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)\right]\\ &+\sigma\left[\varphi\left(\frac{a-\mu}{\sigma}\right)-\varphi\left(\frac{b-\mu}{\sigma}\right)\right].\end{aligned}$$

□

9.2 Calculation of $\int_a^b x^2 \frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right)dx$

In this section, to derive $\int_a^b x^2 \frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right)dx$, we introduce two more lemmas. The first lemma is

Lemma 9.5:

$$\begin{aligned}\frac{1}{\sqrt{2\pi}}\int_\alpha^\beta t^2 e^{-\frac{t^2}{2}}dt =&\left(\frac{a-\mu}{\sigma}\right)\varphi\left(\frac{a-\mu}{\sigma}\right)-\left(\frac{b-\mu}{\sigma}\right)\varphi\left(\frac{b-\mu}{\sigma}\right)\\ &+\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right).\end{aligned}$$

Proof: Let $u := t$ and $v = -e^{-\frac{t^2}{2}}$, then $du = dt$ and $dv := te^{-\frac{t^2}{2}}dt$.

$$\begin{aligned}&\frac{1}{\sqrt{2\pi}}\int_\alpha^\beta t^2 e^{-\frac{t^2}{2}}dt\\ =&\frac{1}{\sqrt{2\pi}}\int_\alpha^\beta udv.\end{aligned}$$

Integrating it by parts, the above becomes

$$\begin{aligned}&\frac{1}{\sqrt{2\pi}}\left([uv]_\alpha^\beta-\int_\alpha^\beta vdu\right)\\ =&\frac{1}{\sqrt{2\pi}}\left(\left[-te^{-\frac{t^2}{2}}\right]_\alpha^\beta-\int_\alpha^\beta -e^{-\frac{t^2}{2}}dt\right)\\ =&-\left[t\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}\right]_\alpha^\beta+\int_\alpha^\beta \frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt\\ =&-\Big[t\varphi(t)\Big]_\alpha^\beta+Z\\ =&\left(\frac{a-\mu}{\sigma}\right)\varphi\left(\frac{a-\mu}{\sigma}\right)-\left(\frac{b-\mu}{\sigma}\right)\varphi\left(\frac{b-\mu}{\sigma}\right)\\ &+\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right).\end{aligned}$$

□

The other lemma is

Lemma 9.6:

$$\begin{aligned}
&\sigma \int_a^b \left(\frac{x-\mu}{\sigma}\right)^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\
=&\sigma^2 \left[\left(\frac{a-\mu}{\sigma}\right) \varphi\left(\frac{a-\mu}{\sigma}\right) - \left(\frac{b-\mu}{\sigma}\right) \varphi\left(\frac{b-\mu}{\sigma}\right) \right. \\
&\left. + \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) \right].
\end{aligned}$$

Proof: Substituting $\alpha := \dfrac{a-\mu}{\sigma}$, $\beta := \dfrac{b-\mu}{\sigma}$ and $f(t) := t\sigma + \mu$ yield

$$\begin{aligned}
&\sigma \int_a^b \left(\frac{x-\mu}{\sigma}\right)^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\
=&\sigma \int_{\alpha\sigma+\mu}^{\beta\sigma+\mu} \left(\frac{x-\mu}{\sigma}\right)^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\
=&\sigma \int_\alpha^\beta t^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \sigma dt \\
=&\sigma^2 \frac{1}{\sqrt{2\pi}} \int_\alpha^\beta t^2 e^{-\frac{t^2}{2}} dt.
\end{aligned}$$

According to Lemma 9.5, the above becomes

$$\begin{aligned}
=&\sigma^2 \left[\left(\frac{a-\mu}{\sigma}\right) \varphi\left(\frac{a-\mu}{\sigma}\right) - \left(\frac{b-\mu}{\sigma}\right) \varphi\left(\frac{b-\mu}{\sigma}\right) \right. \\
&\left. + \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) \right].
\end{aligned}$$

□

Therefore, from these two lemmas, we have the following theorem:

Theorem 9.7: *Let* $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$, $a, b \in \mathbb{R}$, $a \leqslant b$,

$$\begin{aligned}
&\int_a^b x^2 \frac{1}{\sigma} \varphi\left(\frac{x-\mu}{\sigma}\right) dx \\
=&\sigma^2 \left[\left(\frac{a-\mu}{\sigma}\right) \varphi\left(\frac{a-\mu}{\sigma}\right) - \left(\frac{b-\mu}{\sigma}\right) \varphi\left(\frac{b-\mu}{\sigma}\right) \right. \\
&\left. + \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) \right] \\
&+ 2\mu\sigma \left[\varphi\left(\frac{a-\mu}{\sigma}\right) - \varphi\left(\frac{b-\mu}{\sigma}\right) \right]
\end{aligned}$$

$$+\mu^2\left[\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)\right].$$

Proof:

$$\begin{aligned}
&\int_a^b x^2\frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right)dx\\
=&\int_a^b x^2\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx\\
=&\sigma\int_a^b\left(\frac{x^2-2\mu x+\mu^2}{\sigma^2}\right)\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx\\
&+\sigma\int_a^b\left(\frac{2\mu x-\mu^2}{\sigma^2}\right)\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx\\
=&\sigma\int_a^b\left(\frac{x-\mu}{\sigma}\right)^2\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx\\
&+2\mu\int_a^b x\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx\\
&-\mu^2\int_a^b\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx\\
=&\sigma\int_a^b\left(\frac{x-\mu}{\sigma}\right)^2\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx\\
&+2\mu\left(\int_a^b x\frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right)dx\right)-\mu^2Z.
\end{aligned}$$

According to Lemma 9.6, the above becomes

$$\begin{aligned}
&\sigma^2\left[\left(\frac{a-\mu}{\sigma}\right)\varphi\left(\frac{a-\mu}{\sigma}\right)-\left(\frac{b-\mu}{\sigma}\right)\varphi\left(\frac{b-\mu}{\sigma}\right)+Z\right]\\
&+2\mu\left(\int_a^b x\frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right)dx\right)-\mu^2Z.
\end{aligned}$$

According to Theorem 9.4, the above becomes

$$\begin{aligned}
&\sigma^2\left[\left(\frac{a-\mu}{\sigma}\right)\varphi\left(\frac{a-\mu}{\sigma}\right)-\left(\frac{b-\mu}{\sigma}\right)\varphi\left(\frac{b-\mu}{\sigma}\right)+Z\right]\\
&+2\mu\Bigg(\mu\left[\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)\right]\\
&\qquad+\sigma\left[\varphi\left(\frac{a-\mu}{\sigma}\right)-\varphi\left(\frac{b-\mu}{\sigma}\right)\right]\Bigg)\\
&-\mu^2\left[\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)\right]
\end{aligned}$$

$$
\begin{aligned}
=&\sigma^2\left[\left(\frac{a-\mu}{\sigma}\right)\varphi\left(\frac{a-\mu}{\sigma}\right)-\left(\frac{b-\mu}{\sigma}\right)\varphi\left(\frac{b-\mu}{\sigma}\right)+Z\right]\\
&+2\mu\sigma\left[\varphi\left(\frac{a-\mu}{\sigma}\right)-\varphi\left(\frac{b-\mu}{\sigma}\right)\right]\\
&+\mu^2\left[\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)\right]\\
=&\sigma^2\Bigg[\left(\frac{a-\mu}{\sigma}\right)\varphi\left(\frac{a-\mu}{\sigma}\right)-\left(\frac{b-\mu}{\sigma}\right)\varphi\left(\frac{b-\mu}{\sigma}\right)\\
&\qquad+\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)\Bigg]\\
&+2\mu\sigma\left[\varphi\left(\frac{a-\mu}{\sigma}\right)-\varphi\left(\frac{b-\mu}{\sigma}\right)\right]\\
&+\mu^2\left[\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)\right].
\end{aligned}
$$

□

9.3 Re-distributed normal distribution

For the re-distributed case defined in Chapter 7, if the area of normal distribution above the upper bound b is re-distributed at the upper bound point b and the area below the lower bound a is re-distributed at the lower bound point a, we get a re-distributed normal (RD) distribution. The new probability density function based on $\mathcal{N}(\mu,\sigma^2)$ is now defined as

Definition 9.2:

$$
\varphi_{rd}(x;\mu,\sigma,a,b):=\begin{cases}
0 & \text{, if } x<a\\
\frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right)+\Phi\left(\frac{a-\mu}{\sigma}\right) & \text{, if } x=a\\
\frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right) & \text{, if } x\in(a,b)\\
\frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right)+\left[1-\Phi\left(\frac{b-\mu}{\sigma}\right)\right] & \text{, if } x=b\\
0 & \text{, if } x>b.
\end{cases}
$$

$a,b\in\mathbb{R}$ are boundaries, $a\leqslant b$.

For the RD distribution, the mean can be derived as follows,

Theorem 9.8: *Mean of RD distribution is*

$$
\begin{aligned}
\mu_{rd}(\mu,\sigma;a,b)=&\mu\left[\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)\right]\\
&+\sigma\left[\varphi\left(\frac{a-\mu}{\sigma}\right)-\varphi\left(\frac{b-\mu}{\sigma}\right)\right]\\
&+\Phi\left(\frac{a-\mu}{\sigma}\right)a+\left[1-\Phi\left(\frac{b-\mu}{\sigma}\right)\right]b.
\end{aligned}
$$

Proof:

$$\begin{aligned}\mu_{rd}(\mu,\sigma;a,b) :=& \int_{-\infty}^{\infty} x\cdot\varphi_{rd}(x)dx\\ =& \int_{-\infty}^{\infty} x\frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right)dx\\ &+\Phi\left(\frac{a-\mu}{\sigma}\right)a+\left[1-\Phi\left(\frac{b-\mu}{\sigma}\right)\right]b\\ =& \int_{a}^{b} x\frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right)dx\\ &+\Phi\left(\frac{a-\mu}{\sigma}\right)a+\left[1-\Phi\left(\frac{b-\mu}{\sigma}\right)\right]b.\end{aligned}$$

According to Theorem 9.4, the above becomes

$$\begin{aligned}&\mu\left[\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)\right]\\ &+\sigma\left[\varphi\left(\frac{a-\mu}{\sigma}\right)-\varphi\left(\frac{b-\mu}{\sigma}\right)\right]\\ &+\Phi\left(\frac{a-\mu}{\sigma}\right)a+\left[1-\Phi\left(\frac{b-\mu}{\sigma}\right)\right]b.\end{aligned}$$

□

To get variance, we need to find the $E[X^2]$ first.

Lemma 9.9: *$E[X^2]$ of RD distribution is*

$$\begin{aligned}&\int_{-\infty}^{\infty} x^2\varphi_{rd}(x)dx\\ =&\sigma^2\left[\frac{a-\mu}{\sigma}\cdot\varphi\left(\frac{a-\mu}{\sigma}\right)-\frac{b-\mu}{\sigma}\cdot\varphi\left(\frac{b-\mu}{\sigma}\right)\right.\\ &\left.+\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)\right]\\ &+2\mu\sigma\left[\varphi\left(\frac{a-\mu}{\sigma}\right)-\varphi\left(\frac{b-\mu}{\sigma}\right)\right]\\ &+\mu^2\left[\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)\right]\\ &+a^2\Phi\left(\frac{a-\mu}{\sigma}\right)+b^2\left[1-\Phi\left(\frac{b-\mu}{\sigma}\right)\right].\end{aligned}$$

Proof:

$$\begin{aligned}&\int_{-\infty}^{\infty} x^2\varphi_{rd}(x)dx\\ =&\int_{-\infty}^{\infty} x^2\frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right)dx\\ &+a^2\Phi\left(\frac{a-\mu}{\sigma}\right)+b^2\left[1-\Phi\left(\frac{b-\mu}{\sigma}\right)\right]\end{aligned}$$

$$= \int_a^b x^2 \frac{1}{\sigma} \varphi\left(\frac{x-\mu}{\sigma}\right) dx + a^2 \Phi(\alpha) + b^2 \left[1 - \Phi(\beta)\right].$$

According to Theorem 9.7, the above becomes

$$\begin{aligned}
&\sigma^2 \left[\left(\frac{a-\mu}{\sigma}\right) \varphi\left(\frac{a-\mu}{\sigma}\right) - \left(\frac{b-\mu}{\sigma}\right) \varphi\left(\frac{b-\mu}{\sigma}\right) + Z \right] \\
&+2\mu\sigma \left[\varphi\left(\frac{a-\mu}{\sigma}\right) - \varphi\left(\frac{b-\mu}{\sigma}\right) \right] \\
&+\mu^2 Z + a^2 \Phi(\alpha) + b^2 \left[1 - \Phi(\beta)\right] \\
=&\sigma^2 \left[\frac{a-\mu}{\sigma} \cdot \varphi\left(\frac{a-\mu}{\sigma}\right) - \frac{b-\mu}{\sigma} \cdot \varphi\left(\frac{b-\mu}{\sigma}\right) \right. \\
&\qquad \left. +\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) \right] \\
&+2\mu\sigma \left[\varphi\left(\frac{a-\mu}{\sigma}\right) - \varphi\left(\frac{b-\mu}{\sigma}\right) \right] \\
&+\mu^2 \left[\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) \right] \\
&+a^2 \Phi\left(\frac{a-\mu}{\sigma}\right) + b^2 \left[1 - \Phi\left(\frac{b-\mu}{\sigma}\right)\right].
\end{aligned}$$

□

Then we derive the variance of RD distribution as follows

Theorem 9.10: *Variance of RD distribution is*

$$\begin{aligned}
\sigma_{rd}^2(\mu, \sigma; a, b) =&\sigma^2 \left[\frac{a-\mu}{\sigma} \cdot \varphi\left(\frac{a-\mu}{\sigma}\right) - \frac{b-\mu}{\sigma} \cdot \varphi\left(\frac{b-\mu}{\sigma}\right) \right. \\
&\qquad \left. +\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) \right] \\
&+2\mu\sigma \left[\varphi\left(\frac{a-\mu}{\sigma}\right) - \varphi\left(\frac{b-\mu}{\sigma}\right) \right] \\
&+\mu^2 \left[\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) \right] \\
&+a^2 \Phi\left(\frac{a-\mu}{\sigma}\right) + b^2 \left[1 - \Phi\left(\frac{b-\mu}{\sigma}\right)\right] \\
&- \left[\mu \left[\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) \right] \right. \\
&\quad +\sigma \left[\varphi\left(\frac{a-\mu}{\sigma}\right) - \varphi\left(\frac{b-\mu}{\sigma}\right) \right] \\
&\quad \left. +\Phi\left(\frac{a-\mu}{\sigma}\right) a + \left[1 - \Phi\left(\frac{b-\mu}{\sigma}\right)\right] b \right]^2.
\end{aligned}$$

Proof:

$$\begin{aligned}
\sigma_{rd}^2(\mu,\sigma;a,b) :=& \int_{-\infty}^{\infty} x^2\varphi_{rd}(x)dx - (\mu_{rd})^2 \\
=& \sigma^2\left[\alpha\cdot\varphi(\alpha) - \beta\cdot\varphi(\beta) + Z\right] \\
&+2\mu\sigma\left[\varphi(\alpha)-\varphi(\beta)\right] + \mu^2 Z \\
&+a^2\Phi(\alpha) + b^2\left[1-\Phi(\beta)\right] \\
&-\Big(\mu Z + \sigma\left[\varphi(\alpha)-\varphi(\beta)\right] \\
&+\Phi(\alpha)\,a - \Phi(\beta)\,b + b\Big)^2 \\
=& \sigma^2\left[\frac{a-\mu}{\sigma}\cdot\varphi\left(\frac{a-\mu}{\sigma}\right) - \frac{b-\mu}{\sigma}\cdot\varphi\left(\frac{b-\mu}{\sigma}\right)\right. \\
&\qquad \left.+\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)\right] \\
&+2\mu\sigma\left[\varphi\left(\frac{a-\mu}{\sigma}\right) - \varphi\left(\frac{b-\mu}{\sigma}\right)\right] \\
&+\mu^2\left[\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)\right] \\
&+a^2\Phi\left(\frac{a-\mu}{\sigma}\right) + b^2\left[1-\Phi\left(\frac{b-\mu}{\sigma}\right)\right] \\
&-\left[\mu\left[\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)\right]\right. \\
&\quad +\sigma\left[\varphi\left(\frac{a-\mu}{\sigma}\right) - \varphi\left(\frac{b-\mu}{\sigma}\right)\right] \\
&\quad \left.+\Phi\left(\frac{a-\mu}{\sigma}\right)a + \left[1-\Phi\left(\frac{b-\mu}{\sigma}\right)\right]b\right]^2.
\end{aligned}$$

□

According to (9.3), for the m-phase BS model, the logarithmic return over each phase is

$$\ln\Big(\frac{S_j}{S_{j-1}}\Big) \sim \mathcal{N}\Big((r-\frac{\sigma^2}{2})\frac{T}{m}, \sigma^2\frac{T}{m}\Big), \tag{9.4}$$

where S_j is the stock price at the end of the j-th phase, S_{j-1} equals the stock price at the beginning of the j-th phase and $j \in \mathbb{Z}^+ \cap [1,m]$. Then for the re-distributed normal distribution based on $\mathcal{N}\Big((r-\frac{\sigma^2}{2})\frac{T}{m}, \sigma^2\frac{T}{m}\Big)$, according to Theorem 9.8, the mean is

$$\mu_{rd}\Big((r-\frac{\sigma^2}{2})\frac{T}{m}, \sigma\sqrt{\frac{T}{m}}; a, b\Big) =: \mu_{each} \tag{9.5}$$

and according to Theorem 9.10, the variance is

$$\sigma_{rd}^2\Big((r-\frac{\sigma^2}{2})\frac{T}{m}, \sigma\sqrt{\frac{T}{m}}; a, b\Big) =: \sigma_{each}^2. \tag{9.6}$$

9.4 A recursive Black-Scholes model

For m phases, according to (9.1), the logarithmic return over all phases is

$$\ln\left(\frac{S_m}{S_0}\right) \sim \mathcal{N}\left(m\mu_{each}, m\sigma^2_{each}\right), \tag{9.7}$$

where the initial stock price S_0 is given, S_m is stock price at the end of m-th phase, μ_{each} is the statistical mean and σ^2_{each} is the statistical variance in each phase. In other words, there exist some r_{bs} and σ_{bs}, such that

$$m\mu_{each} = (r_{bs} - \frac{\sigma^2_{bs}}{2})T \tag{9.8}$$

and

$$m\sigma^2_{each} = \sigma^2_{bs}T. \tag{9.9}$$

By solving the above, it is clear that

$$\sigma_{bs} = \sqrt{\frac{m}{T}}\sigma_{each} \tag{9.10}$$

and

$$r_{bs} = \frac{m\mu_{each}}{T} + \frac{m\sigma^2_{each}}{2T}. \tag{9.11}$$

Now we can define an $\mathcal{N}\big(r_{bs}, \sigma^2_{bs}\big)$ using the above results. Using the plain Black-Scholes model (Theorem 4.1), the corresponding European call option is priced as

$$c = S_0\Phi(d_1) - Ke^{-r_{bs}T}\Phi(d_2), \tag{9.12}$$

where

$$d_1 = \frac{\ln(S_0/K) + (r_{bs} + \sigma^2_{bs}/2)T}{\sigma_{bs}\sqrt{T}} \tag{9.13}$$

and

$$d_2 = \frac{\ln(S_0/K) + (r_{bs} - \sigma^2_{bs}/2)T}{\sigma_{bs}\sqrt{T}}. \tag{9.14}$$

We explore a simple case here, where $m = 2$, $S_0 = 1$, $r = 0.03$, $\sigma = 0.3$, $T = 1$, $L_d = 0.9$ and $L_u = 1.1$. According to Definition 9.2,

$$\begin{aligned} a &= \ln\left(\frac{S_0 L_d}{S_0}\right) \\ &= \ln(0.9) \end{aligned} \tag{9.15}$$

and

$$\begin{aligned} b &= \ln\Big(\frac{S_0 L_u}{S_0}\Big) \\ &= \ln(1.1). \end{aligned} \tag{9.16}$$

Figure 9.1 demonstrates the difference

$$d(K) := rd(K) - bs(K) \tag{9.17}$$

and the relative difference in percentage

$$d_r(K) := \frac{rd(K) - bs(K)}{bs(K)} \times 100\%, \tag{9.18}$$

where K is the European call option strike price, $rd(K)$ is the multi-phase model value for K and $bs(K)$ is the recursive Black-Scholes model value for K. Since most of the difference values fall within $[-0.002, 0.002]$, it is clear that these two models produce very close results. The relative difference is reasonably small if $K < 1.1$. For this 2-phase case, it is obvious that $rd(K) = 0$, if $K > S_0 L_u^2$. So $d_r(K)$ rapidly drops to -100%, if $K > 1.1^2$ (shown in Figure 9.1). Furthermore, it is clear that for the m-phase case, $d_r(K)$ should equal -100%, if $K \geqslant S_0(L_u)^m$.

Figure 9.2 demonstrates another case, where $K = 0.5$, $S_0 = 1$, $r = 0.03$, $\sigma = 0.3$, $T = 1$, $L_d = 0.9$ and $L_u = 1.1$. The difference

$$d(m) := rd(m) - bs(m) \tag{9.19}$$

and relative difference

$$d_r(m) := \frac{rd(m) - bs(m)}{bs(m)} \times 100\% \tag{9.20}$$

for different number of phases (m in this chapter) reaffirm our finding that these two models produce very close results. Since there is a closed-form solution for the recursive Black-Scholes model, the computation time for the recursive Black-Scholes model is much less than the computation time for the multi-phase model defined in Chapter 8. So within reasonable range when considering the volatile nature of the options market, the recursive Black-Scholes model can be used as a practical model to replace the multi-phase model.

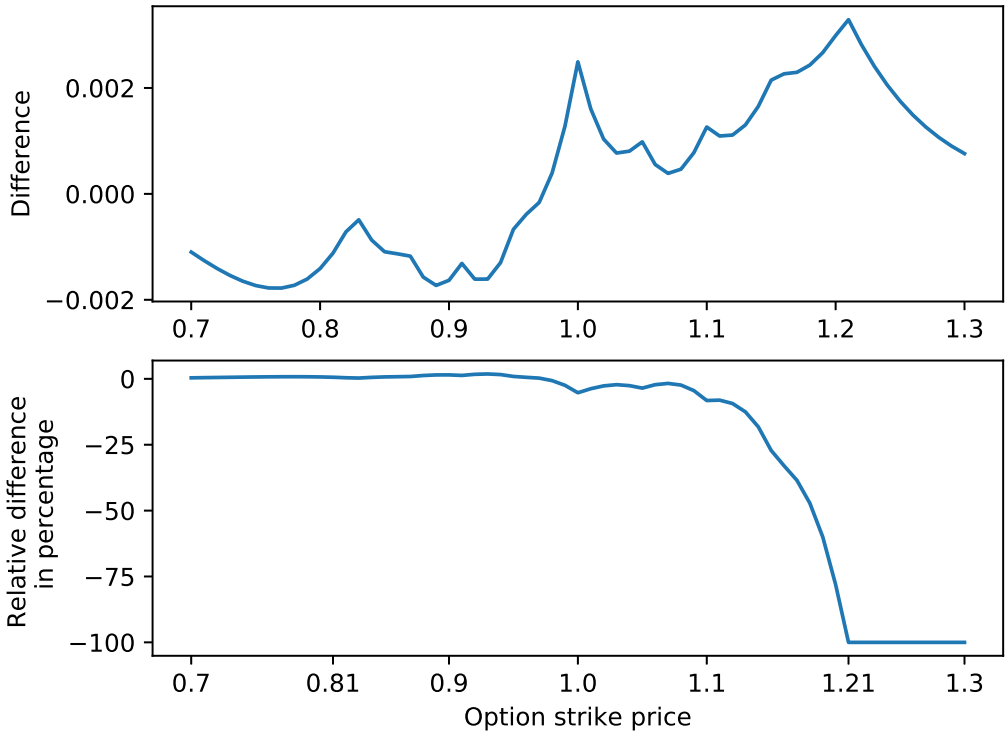


Figure 9.1: Difference between 2-phase re-distributed model and 2-phase recursive Black-Scholes model for different strike price K: $rd(K) - bs(K)$ (top) and $\frac{rd(K) - bs(K)}{bs(K)} \times 100\%$ (bottom).

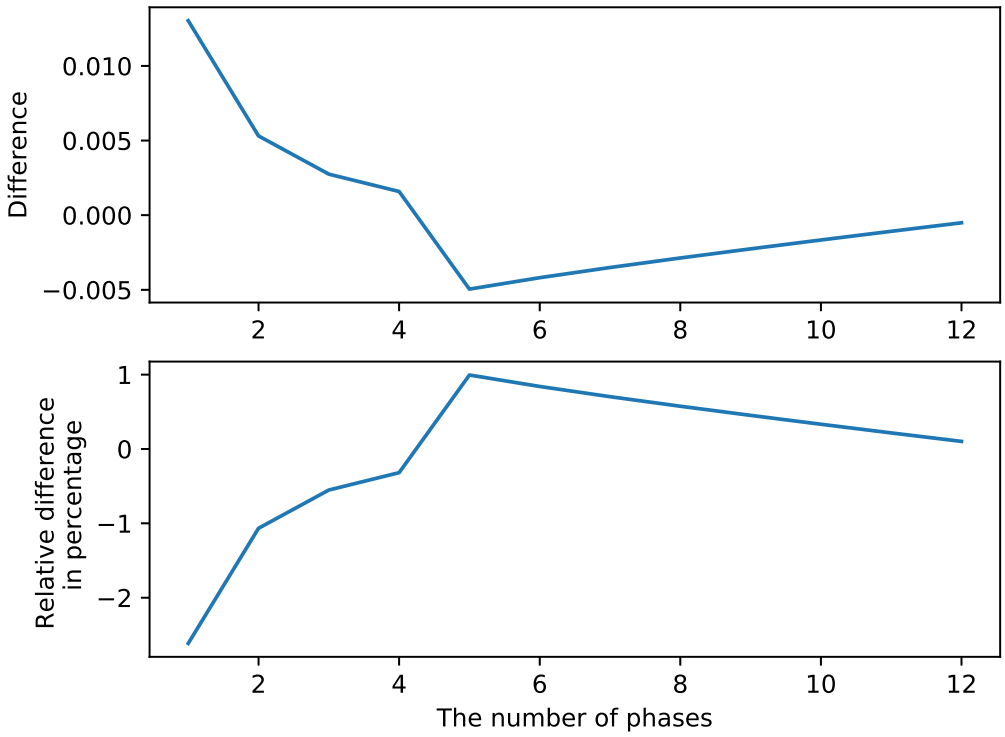


Figure 9.2: Price difference $rd(m) - bs(m)$ (top) and relative price difference $\dfrac{rd(m) - bs(m)}{bs(m)} \times 100\%$ (bottom) between m-phase re-distributed model and recursive Black-Scholes model.

In this part, we introduced several option valuation models. We presented some mathematical preparations in Chapter 5. We explained some properties under a part of the normal distribution and introduced a truncated model in that chapter. Those preparations and properties were used through this part. Following Amemiya's (1973) idea, we introduced a re-weighting model for price-bounded options in Chapter 6. This re-weighting model relied on truncated normal distributions that we explained in Chapter 4. Both the truncated model and the re-weighting model were the stepping stones for our re-distributed model.

Following Arak & Cook's (1997) idea, we introduced the re-distributed model in Chapter 7, using the binomial tree approach. To derive this model, we assumed that the latent variable, underlying economic value, follows a geometric Brownian motion and the observable variable, market price, is censored. The outliers were shifted to the boundaries. We believed that the logarithmic returns of market prices in price-bounded markets follow a re-distributed normal distribution (explained in Chapter 9). Using the raw data we introduced in Part I, setting up portfolios of stocks and options, empirical results suggested that our re-distributed model is better than the Black-Scholes model to evaluate options.

However, the re-distributed model is a 1-phase model for 1-day options only. To evaluate longer-term options, we introduced a multi-phase model in Chapter 8. The multi-phase model relied on recursively defined stock prices and probabilities. No closed-form solution was found. Numerical methods were used to find the option values. Simulation results suggested that the multi-phase model requires extremely long computation time. It took more than 30 minutes to find an option value. It is not practical enough. We believed that this multi-phase model is not suitable for long-term options.

To reduce the computation time, we proposed a re-distributed normal distribution and introduced an m-phase recursive Black-Scholes model in Chapter 9. A normal distribution was used to approximate the re-distributed normal distribution. Hence, a simplified closed-form solution was found. Compared to our re-distributed model, simulation results suggested that the recursive Black-Scholes model is accurate enough. The percentage error was approximately 1%. As there was a closed-form solution, the recursive Black-Scholes model required negligible computation time. It is acceptable. We believed that this recursive Black-Scholes model is suitable for long-term options.

Chapter 10

Conclusion

In this thesis, we analysed Chinese financial market from two perspectives. Although there are 2 parts, they focus on same research objective, to evaluate financial securities on price-bounded markets.

In Part I, we proposed several stock valuation models. We explored the relationship between returns and price-earnings (PE) ratios in Chapter 2. This model suggested that lower PE implies higher future return. However this simple return-PE model can only explain approximately 3% of data. Because of its performance, although it is straightforward, it is not acceptable.

Then we introduced more variables and explored the relationship between market capitalisations and other 16 variables in Chapter 3. The final 16-variable regression model can explain 92.3% of data. It showed a clear relationship among the financial variables. It is usable to evaluate stocks. Although it has much more explanatory power than the return-PE model, this 16-variable model is more complex and requires 16 independent variable data. In real-world applications, analysts prefer simple models which require less datasets. So this 16-variable model is useful as a reference model, but not practical for real-world applications. Furthermore, the mean of beta in this model was divergent. We were not satisfied with this model.

As neither the simple return-PE model nor the 16-variable model produced satisfactory result, we separated our data variables into 5 groups and used principal component analysis techniques to extract the important information. The final 5-variable model can explain more than 57.2% of data. Although this 5-variable model has less explanatory power than the 16-variable model, this explanatory power level is still acceptable. Moreover, the mean of beta was convergent in this model. However, it still required a large set of data. In our analysis, totally 16 datasets were used. It is still not practical enough.

To improve our valuation models and exploit our data, in this thesis, we decided to view the data from another perspective. In Part II, we proposed several option valuation models for price-bounded options. We introduced a re-distributed model in Chapter 7. Using binomial trees and following Hull's (2015) idea, we derived its closed-form solution. In this model, we assumed that there is a pair of boundaries, i.e. a upper boundary and a lower boundary. Obviously, this re-distributed model is a 1-phase model. Although the derivation of the re-distributed model is mathematically rigorous, this model can only be used to evaluate short-term options.

In long-term option valuations, there is a speed-accuracy trade-off between two models. The

multi-phase model is more accurate. And the m-phase recursive Black-Scholes model is faster. Using recursively defined stock prices and probabilities, we introduced a multi-phase model in Chapter 8. As no closed-form solution was derived, it relied on numerical methods and required long computation time. Using plain normal distributions to approximate our re-distributed normal distribution, we introduced an m-phase recursive Black-Scholes model in Chapter 9. As approximation was used, some errors were introduced into the analysis.

If the number of phases went larger, the computation time of the multi-phase model went extremely long and the multi-phase model became unacceptable. So we proposed the m-phase recursive Black-Scholes model as our option valuation model for long-term options, because it required negligible computation time for long-term options and generated approximate 2% error. It is practical and usable.

However, according to the the Moore's law, the expected computation power could double in a few years. Moreover, we believe that the computation time can be further reduced by using some optimised algorithms. Furthermore, as the multi-phase model is mathematically rigorous, it can be used as a useful reference and benchmark to evaluate price-bounded options.

Finally, the main drawback of our stock valuation models is the unrealistic data requirement. Except the return-PE model, all models consumed 16 datasets, before enough explanatory power was gathered. We suspect that there is problem with data quality. A possible further research is to optimise the datasets and reduce the data requirement in the models. We expect that there exist a smaller set of carefully selected data that could also improve the goodness of fit.

Following the idea from Chapter 9, another possible further research is to approximate other distributions, for example beta distributions, using normal distributions. Using beta distributions in option pricing is supported by some literatures. However, it requires complicated techniques to transform beta distributions into binomial trees (see Chapter 4). The results from Chapter 9 suggests that approximation using normal distributions might be an acceptable method.

Appendix A

2-phase model

In this chapter, the approaches for 2-phase models will be discussed. The two phases will be treated as independent objects, and separate binomial trees will be set up for each phase. Since 1-phase models have already been studied in previous chapters, the 1st phase close price S_1, which could be obtained from the 1-phase models, will be used as the initial price for the 2nd-phase binomial tree. Hence, in Section A.1, 1-phase approaching will be summarised. After that, 2-phase models will be derived from 1-phase models in the following sections.

A.1 Summarising of the 1-phase approaching

In this section, we introduce some mathematical preparations. For 1-phase approach, assumed there are i_1 upwards steps, in the phase. Then the 1st phase close price S_1 is

$$S_1(S_0, i_1) = S_0 u^{i_1} d^{n-i_1}. \tag{A.1}$$

The probability P_1 at $S_1(\cdot, i_1)$ is

$$P_1(i_1) = \frac{n!}{(n-i_1)!i_1!} p^{i_1}(1-p)^{n-i_1}. \tag{A.2}$$

According to CRR model, a European call option price is calculated as

$$c := e^{-rT} \sum_{i_1=0}^{n} P_1 \times max(S_1 - K, 0), \tag{A.3}$$

where T is the option life time.

On a regular binomial tree, $i_\cdot \in [0, n]$. However in re-distributed model, there are upper and lower price bound, i.e. $i_1 \in [\gamma, \beta]$. To make a valid probability distribution, we place extra probability at the boundary points.

According to (5.19), the probability above upper bound is

$$P(u^+) := \frac{ln(1/L_u) + (r - \sigma^2/2)T}{\sigma\sqrt{T}}. \tag{A.4}$$

We replace it at the upper bound. And according to (5.31), the probability below lower bound is

$$P(d^-) := 1 - \frac{ln(1/L_d) + (r - \sigma^2/2)T}{\sigma\sqrt{T}}. \tag{A.5}$$

We replace it at the lower bound. Hence

$$\begin{aligned} c =& e^{-rT}\Big[\sum_{\gamma}^{\beta} P \times max(S-K,0) \\ &+ \sum_{i_1=0}^{\gamma} \mathbf{1}_{i_1:K<S(i_1)} P(i_1)(S_0 L_d - K) \\ &+ \sum_{i_1=\beta}^{n} \mathbf{1}_{i_1:K<S(i_1)} P(i_1)(S_0 L_u - K)\Big], \end{aligned} \tag{A.6}$$

where the indicator function $\mathbf{1}_{x:A(x\cdot)}$ is defined as

$$\mathbf{1}_{x:A(x,\cdot)} = \begin{cases} 1, & \text{if event } A(x,\cdot) \text{ occurs} \\ 0, & \text{anything else.} \end{cases}$$

A.2 2 phases: Re-weighting

In this section, a 2-phase re-weighting model will be introduced. Let n denote the number of steps for each phase, T be the time length of one phase, i_1 and i_2 be the number of upward steps in phase 1 and 2.

For the first phase, the close price $S_1(S_0, i_1) = S_0 u^{i_1} d^{n-i_1}$, with probability

$$P_1(i_1) = \frac{n!}{(n-i_1)!i_1!} p^{i_1}(1-p)^{n-i_1}. \tag{A.7}$$

For the second phase, if S_1 is given, the close price $S_2(S_1, i_2) = S_1 u^{i_2} d^{n-i_2}$, with probability

$$P_2(i_2) = \frac{n!}{(n-i_2)!i_2!} p^{i_2}(1-p)^{n-i_2}. \tag{A.8}$$

The above can be combined together to get

$$S_2(S_0, i_1, i_2) = S_0 u^{i_1} d^{n-i_1} u^{i_2} d^{n-i_2}, \tag{A.9}$$

with probability

$$P_2(i_1, i_2) = \frac{n!}{(n-i_1)!i_1!} p^{i_1}(1-p)^{n-i_1} \frac{n!}{(n-i_2)!i_2!} p^{i_2}(1-p)^{n-i_2}. \tag{A.10}$$

Then the 2-phase re-weighting model can be written as

$$\begin{aligned} c = e^{-2rT} \frac{1}{\varrho^2} \sum_{i_1=\gamma}^{\beta} \sum_{i_2=\gamma}^{\beta} & \frac{n!}{(n-i_1)!i_1!} p^{i_1}(1-p)^{n-i_1} \\ & \frac{n!}{(n-i_2)!i_2!} p^{i_2}(1-p)^{n-i_2} \end{aligned} \tag{A.11}$$

$$max(S_0 u^{i_1} d^{n-i_1} u^{i_2} d^{n-i_2} - K, 0).$$

Four possible situations will be discussed for the strike price K:

1. $S_0 L_u^2 \leqslant K$, the stock price $S_2 \leqslant S_0 L_u^2$ is always below strike price K, hence $c = 0$.
2. $S_0 L_u L_d \leqslant K < S_0 L_u^2$.
3. $S_0 L_d^2 < K < S_0 L_d L_u$.
4. $K \leqslant S_0 L_d^2$.

A.2.1 If $K \leqslant S_0 L_d^2$

If $K \leqslant S_0 L_d^2$, since $S_0 L_d^2 \leqslant S_2$,

$$\forall i_1, i_2 \in [\gamma, \beta], \tag{A.12}$$

such that $0 \leqslant S_0 u^{i_1} d^{n-i_1} u^{i_2} d^{n-i_2} - K$, hence

$$\begin{aligned} & max(S_0 u^{i_1} d^{n-i_1} u^{i_2} d^{n-i_2} - K, 0) \\ =& S_0 u^{i_1} d^{n-i_1} u^{i_2} d^{n-i_2} - K. \end{aligned} \tag{A.13}$$

$$\begin{aligned} c = \quad & e^{-2rT} \frac{1}{\varrho^2} \sum_{i_1=\gamma}^{\beta} \sum_{i_2=\gamma}^{\beta} \quad \frac{n!}{(n-i_1)! i_1!} p^{i_1} (1-p)^{n-i_1} \\ & \qquad \frac{n!}{(n-i_2)! i_2!} p^{i_2} (1-p)^{n-i_2} \\ & \qquad (S_0 u^{i_1} d^{n-i_1} u^{i_2} d^{n-i_2} - K) \\ = \ & S_0 e^{-2rT} \frac{1}{\varrho^2} \sum_{i_1=\gamma}^{\beta} \sum_{i_2=\gamma}^{\beta} \quad \frac{n!}{(n-i_1)! i_1!} p^{i_1} (1-p)^{n-i_1} \\ & \qquad \frac{n!}{(n-i_2)! i_2!} p^{i_2} (1-p)^{n-i_2} \\ & \qquad u^{i_1} d^{n-i_1} u^{i_2} d^{n-i_2} \\ & - K e^{-2rT} \frac{1}{\varrho^2} \sum_{i_1=\gamma}^{\beta} \sum_{i_2=\gamma}^{\beta} \frac{n!}{(n-i_1)! i_1!} p^{i_1} (1-p)^{n-i_1} \\ & \qquad \frac{n!}{(n-i_2)! i_2!} p^{i_2} (1-p)^{n-i_2}. \end{aligned} \tag{A.14}$$

Since

$$\sum_{i=m}^{n} \sum_{j=s}^{t} a_i b_j = \sum_{i=m}^{n} a_i \times \sum_{j=s}^{t} b_j, \tag{A.15}$$

we have

$$
\begin{aligned}
c = \; & S_0 e^{-2rT} \frac{1}{\varrho^2} \sum_{i_1=\gamma}^{\beta} \frac{n!}{(n-i_1)!i_1!} p^{i_1}(1-p)^{n-i_1} u^{i_1} d^{n-i_1} \\
& \times \sum_{i_2=\gamma}^{\beta} \frac{n!}{(n-i_2)!i_2!} p^{i_2}(1-p)^{n-i_2} u^{i_2} d^{n-i_2} \\
& - K e^{-2rT} \frac{1}{\varrho^2} \sum_{i_1=\gamma}^{\beta} \frac{n!}{(n-i_1)!i_1!} p^{i_1}(1-p)^{n-i_1} \\
& \times \sum_{i_2=\gamma}^{\beta} \frac{n!}{(n-i_2)!i_2!} p^{i_2}(1-p)^{n-i_2} \\
= \; & S_0 e^{-2rT} \frac{1}{\varrho^2} U_1^2 - K e^{-2rT} \frac{1}{\varrho^2} U_2^2 \\
= \; & \frac{1}{\varrho^2} (S_0 e^{-2rT} U_1^2 - K e^{-2rT} U_2^2) \\
= \; & \frac{1}{\varrho^2} (S_0 e^{-2rT} e^{2rT} (\Phi(d_1) - \Phi(d_3))^2 \\
& - K e^{-2rT} (\Phi(d_2) - \Phi(d_4))^2) \\
= \; & \frac{1}{\varrho^2} \Big(S_0 (\Phi(d_1) - \Phi(d_3))^2 \\
& - K e^{-2rT} (\Phi(d_2) - \Phi(d_4))^2 \Big),
\end{aligned}
$$

where

$$d_1 = \frac{ln(1/L_d) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}, \tag{A.16}$$

$$d_2 = \frac{ln(1/L_d) + (r - \sigma^2/2)T}{\sigma\sqrt{T}}, \tag{A.17}$$

$$d_3 = \frac{ln(1/L_u) + (r + \sigma^2/2)T}{\sigma\sqrt{T}} \tag{A.18}$$

and

$$d_4 = \frac{ln(1/L_u) + (r - \sigma^2/2)T}{\sigma\sqrt{T}}. \tag{A.19}$$

The un-re-weighted result

$$S_0(\Phi(d_1) - \Phi(d_3))^2 - K e^{-2rT} (\Phi(d_2) - \Phi(d_4))^2 \tag{A.20}$$

without ϱ will be used in the distributed model as the 1st part.

A.2.2 If $S_0 L_u L_d \leqslant K < S_0 L_u^2$

If $S_0 L_u L_d \leqslant K < S_0 L_u^2$. For the second phase, if $S_1 = S_0 u^{2i_1 - n}$ is given, for i_2, such that $K < S_1 u^{i_2} d^{n-i_2}$, we have $i_2 > \eta := \frac{1}{2}(n - \frac{ln(\frac{S_0 u^{2i_1-n}}{K})}{\sigma\sqrt{T/n}})$.

Let S_1 such that $K = S_1 L_u$. Since $u = e^{\sigma\sqrt{T/n}}$,

$$\begin{aligned} & \frac{K}{L_u} = S_1 = S_0 u^{i_1} d^{n-i_1} = S_0 u^{2i_1 - n} \\ \iff & \frac{K}{L_u S_0} = u^{2i-n} \\ \iff & ln(\frac{K}{L_u S_0}) = (2i_1 - n) ln(u) \\ \iff & \frac{ln(K) - ln(L_u S_0)}{ln(u)} = 2i_1 - n \\ \iff & \frac{ln(K) - ln(L_u S_0)}{ln(u)} + n = 2i \\ \iff & \frac{1}{2}(\frac{ln(K) - ln(L_u S_0)}{\sigma\sqrt{T/n}} + n) = i_1. \end{aligned} \tag{A.21}$$

It is clear that

Lemma A.1: $K = S_1 L_u \iff \frac{1}{2}(\frac{ln(K) - ln(L_u S_0)}{\sigma\sqrt{T/n}} + n) = i_1$, *where* $u = e^{\sigma\sqrt{T/n}}$.

Set $\zeta := \frac{1}{2}(\frac{ln(\frac{K}{L_u S_0})}{\sigma\sqrt{T/n}} + n)$. Let $S_1 := S_0 u^{i_1} d^{n-i_1}$. Then

$$\begin{aligned} \forall i_1 < \zeta, \text{ such that } & S_1 L_u < K \\ & \iff max(S_1 L_u - K, 0) = 0. \end{aligned} \tag{A.22}$$

$$\begin{aligned} c = & e^{-2rT} \frac{1}{\varrho^2} \lim_{n \longrightarrow \infty} \sum_{i_1 = \zeta}^{\beta} \sum_{i_2 = \eta}^{\beta} \frac{n!}{(n-i_1)! i_1!} p^{i_1} (1-p)^{n-i_1} \\ & \frac{n!}{(n-i_2)! i_2!} p^{i_2} (1-p)^{n-i_2} \\ & (S_0 u^{i_1} d^{n-i_1} u^{i_2} d^{n-i_2} - K) \\ = & e^{-2rT} \frac{1}{\varrho^2} S_0 \lim_{n \longrightarrow \infty} \sum_{i_1 = \zeta}^{\beta} \frac{n!}{(n-i_1)! i_1!} p^{i_1} (1-p)^{n-i_1} u^{i_1} d^{n-i_1} \\ & \lim_{n \longrightarrow \infty} \sum_{i_2 = \eta}^{\beta} \frac{n!}{(n-i_2)! i_2!} p^{i_2} (1-p)^{n-i_2} u^{i_2} d^{n-i_2} \\ & - e^{-2rT} \frac{1}{\varrho^2} K \lim_{n \longrightarrow \infty} \sum_{i_1 = \zeta}^{\beta} \frac{n!}{(n-i_1)! i_1!} p^{i_1} (1-p)^{n-i_1} \\ & \lim_{n \longrightarrow \infty} \sum_{i_2 = \eta}^{\beta} \frac{n!}{(n-i_2)! i_2!} p^{i_2} (1-p)^{n-i_2} \\ = & e^{-rT} \frac{1}{\varrho^2} S_0 \lim_{n \longrightarrow \infty} \sum_{i_1 = \zeta}^{\beta} \frac{n!}{(n-i_1)! i_1!} p^{i_1} (1-p)^{n-i_1} u^{i_1} d^{n-i_1} \\ & (\Phi(d_1') - \Phi(d_3)) \end{aligned} \tag{A.23}$$

$$
\begin{aligned}
& - e^{-2rT}\frac{1}{\varrho^2}K \lim_{n\longrightarrow\infty}\sum_{i_1=\zeta}^{\beta}\frac{n!}{(n-i_1)!i_1!}p^{i_1}(1-p)^{n-i_1} \\
& \qquad\qquad (\Phi(d_2')-\Phi(d_4)) \\
= \quad & e^{-rT}\frac{1}{\varrho^2}S_0 \qquad (\Phi(d_1)-\Phi(d_3))(-\Phi(d_3)) \\
& + e^{-rT}\frac{1}{\varrho^2}S_0 \lim_{n\longrightarrow\infty}\sum_{i_1=\zeta}^{\beta}\frac{n!}{(n-i_1)!i_1!}p^{i_1}(1-p)^{n-i_1}u^{i_1}d^{n-i_1}\Phi(d_1') \\
& - e^{-2rT}\frac{1}{\varrho^2}K \qquad (\Phi(d_2)-\Phi(d_4))(-\Phi(d_4)) \\
& - e^{-2rT}\frac{1}{\varrho^2}K \lim_{n\longrightarrow\infty}\sum_{i_1=\zeta}^{\beta}\frac{n!}{(n-i_1)!i_1!}p^{i_1}(1-p)^{n-i_1}\Phi(d_2'),
\end{aligned}
$$

where

$$d_1 = \frac{ln(S_0L_u/K)+(r+\sigma^2/2)T}{\sigma\sqrt{T}}, \tag{A.24}$$

$$d_1' = \frac{ln(S_0u^{2i_1-n}/K)+(r+\sigma^2/2)T}{\sigma\sqrt{T}}, \tag{A.25}$$

$$d_2 = \frac{ln(S_0L_u/K)+(r-\sigma^2/2)T}{\sigma\sqrt{T}}, \tag{A.26}$$

$$d_2' = \frac{ln(S_0u^{2i_1-n}/K)+(r-\sigma^2/2)T}{\sigma\sqrt{T}}, \tag{A.27}$$

$$d_3 = \frac{ln(1/L_u)+(r+\sigma^2/2)T}{\sigma\sqrt{T}} \tag{A.28}$$

and

$$d_4 = \frac{ln(1/L_u)+(r-\sigma^2/2)T}{\sigma\sqrt{T}}. \tag{A.29}$$

A.2.3 If $S_0L_d^2 < K < S_0L_dL_u$

If $S_0L_d^2 < K < S_0L_dL_u$, let S_1 such that $S_1L_d < K$. According to Lemma A.1, $i_1 < \frac{1}{2}(\frac{ln(\frac{K}{L_dS_0})}{\sigma\sqrt{T/n}} + n)$.

Now we set $\theta := \frac{1}{2}(\frac{ln(\frac{K}{L_dS_0})}{\sigma\sqrt{T/n}} + n)$. The binomial tree is above K, if $i_1 > \theta$. The binomial tree is intersected with K, if $i_1 \leqslant \theta$. Hence

$$
\begin{aligned}
c = \; & e^{-2rT}\frac{1}{\varrho^2}\sum_{i_1=\gamma}^{\theta}\sum_{i_2=\eta}^{\beta} \quad \frac{n!}{(n-i_1)!i_1!}p^{i_1}(1-p)^{n-i_1} \\
& \qquad\qquad \frac{n!}{(n-i_2)!i_2!}p^{i_2}(1-p)^{n-i_2}
\end{aligned} \tag{A.30}
$$

$$
\begin{aligned}
& & & (S_0 u^{i_1} d^{n-i_1} u^{i_2} d^{n-i_2} - K) \\
& + e^{-2rT} \frac{1}{\varrho^2} \sum_{i_1=\theta}^{\beta} \sum_{i_2=\gamma}^{\beta} & & \frac{n!}{(n-i_1)!i_1!} p^{i_1} (1-p)^{n-i_1} \\
& & & \frac{n!}{(n-i_2)!i_2!} p^{i_2} (1-p)^{n-i_2} \\
& & & (S_0 u^{i_1} d^{n-i_1} u^{i_2} d^{n-i_2} - K) \\
= & \quad e^{-rT} \frac{1}{\varrho^2} S_0 & & \Big[(\Phi(d_1) - \Phi(d_3'))(-\Phi(d_3)) \\
& & & + \lim_{n \longrightarrow \infty} \sum_{i_1=\gamma}^{\theta} \frac{n!}{(n-i_1)!i_1!} p^{i_1} (1-p)^{n-i_1} u^{i_1} d^{n-i_1} \Phi(d_1') \Big] \\
& - e^{-2rT} \frac{1}{\varrho^2} K & & \Big[(\Phi(d_2) - \Phi(d_4'))(-\Phi(d_4)) \\
& & & + \lim_{n \longrightarrow \infty} \sum_{i_1=\gamma}^{\theta} \frac{n!}{(n-i_1)!i_1!} p^{i_1} (1-p)^{n-i_1} \Phi(d_2') \Big] \\
& + \frac{1}{\varrho^2} & & \Big(S_0 (\Phi(d_3') - \Phi(d_3))(\Phi(d_1) - \Phi(d_3)) \\
& & & - K e^{-2rT} (\Phi(d_4') - \Phi(d_4))(\Phi(d_2) - \Phi(d_4)) \Big),
\end{aligned}
$$

where

$$d_1 = \frac{ln(1/L_d) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}, \tag{A.31}$$

$$d_1' = \frac{ln(S_0 u^{2i_1 - n}/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}, \tag{A.32}$$

$$d_2 = \frac{ln(1/L_d) + (r - \sigma^2/2)T}{\sigma\sqrt{T}}, \tag{A.33}$$

$$d_2' = \frac{ln(S_0 u^{2i_1 - n}/K) + (r - \sigma^2/2)T}{\sigma\sqrt{T}}, \tag{A.34}$$

$$d_3 = \frac{ln(1/L_u) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}, \tag{A.35}$$

$$d_3' = \frac{ln(S_0 L_d/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}, \tag{A.36}$$

$$d_4 = \frac{ln(1/L_u) + (r - \sigma^2/2)T}{\sigma\sqrt{T}} \tag{A.37}$$

and

$$d_4' = \frac{ln(S_0 L_d/K) + (r - \sigma^2/2)T}{\sigma\sqrt{T}}. \tag{A.38}$$

In this section, a re-weighting approach is discussed. The un-re-weighted results will also be used in the distributed model in the next section.

A.3 2 phases: Re-distributed model

For 2-phase distributed approaching, we set up separate binomial tree for each phase.

On the first phase, it is a simple binomial tree between the price boundaries. So the close price

$$S_1(S_0, i_1) = S_0 u^{i_1} d^{n-i_1}, \tag{A.39}$$

with probability

$$P_1(i_1) = \frac{n!}{(n-i_1)!i_1!} p^{i_1}(1-p)^{n-i_1}. \tag{A.40}$$

At the price boundaries, the extra probability is distributed, since the price cannot beyond the boundaries. So the boundary points themselves have the probability beyond the boundaries. Hence at the lower bound, $S_1 = S_0 L_d$, the extra probability is

$$P(d^-) := 1 - \Phi(\frac{ln(1/L_d) + (r - \sigma^2/2)T}{\sigma\sqrt{T}}). \tag{A.41}$$

At the upper bound, $S_1 = S_0 L_u$, the extra probability is

$$P(u^+) := \Phi(\frac{ln(1/L_u) + (r - \sigma^2/2)T}{\sigma\sqrt{T}}). \tag{A.42}$$

For the second phase, if S_1 is given, the close price $S_2(S_1, i_2) = S_1 u^{i_2} d^{n-i_2}$, with probability $P_2(i_2) = \frac{n!}{(n-i_2)!i_2!} p^{i_2}(1-p)^{n-i_2}$. We can calculate the S_1 value and its probability, using the 1-phase model. So we combine above together to get

$$S_2(S_0, i_1, i_2) = S_0 u^{i_1} d^{n-i_1} u^{i_2} d^{n-i_2}, \tag{A.43}$$

with probability

$$P_2(i_1, i_2) = \frac{n!}{(n-i_2)!i_2!} p^{i_2}(1-p)^{n-i_2} P_1(i_1). \tag{A.44}$$

Lemma A.2: *At the lower bound, $S_2 = S_1 L_d$, the extra probability is*

$$P(d^-) := (1 - \Phi(\frac{ln(1/L_d) + (r - \sigma^2/2)T}{\sigma\sqrt{T}}))P_1(i_1). \tag{A.45}$$

Lemma A.3: *At the upper bound, $S_2 = S_1 L_u$, the extra probability is*

$$P(u^+) := \Phi(\frac{ln(1/L_u) + (r - \sigma^2/2)T}{\sigma\sqrt{T}})P_1(i_1). \tag{A.46}$$

We use the above model to find S_2 value and its probability P_2 value. Hence, a European call option price can be calculated as

$$c = e^{-rT} \sum_{i_1=0}^{\infty} \sum_{i_2=0}^{\infty} max(S_2 - K, 0) \times P_2. \tag{A.47}$$

To simplify the calculation, we divide the model into four separated parts, which are

- c_1, both S_1 and S_2 are between the boundaries, which is the un-re-distributed 2-phase re-distributed model.
- c_2, only S_2 is between the boundaries, and S_1 is beyond a boundary, which means the 2nd phase is a un-re-distributed 1-phase re-distributed model.
- c_3, only S_1 is between the boundaries, which means the 1st phase is a un-re-distributed 1-phase re-distributed model.
- c_4, both S_1 and S_2 are beyond the boundaries.

Therefore, we have

$$
\begin{aligned}
c :=& c_1 + c_2 + c_3 + c_4 \qquad (A.48)\\
=& e^{-2rT}\Big[\sum_{i_1=\gamma}^{\beta}\sum_{i_2=\gamma}^{\beta} \frac{n!}{i_1!(n-i_1)!}\frac{n!}{i_2!(n-i_2)!}p^{i_1+i_2}(1-p)^{n-i_1+n-i_2}\\
& \qquad max(S_0u^{i_1+i_2}d^{n-i_1+n-i_2}-K,0)\\
& + \sum_{i_2=\gamma}^{\beta} \frac{n!}{i_2!(n-i_2)!}p^{i_2}(1-p)^{n-i_2}\\
& \qquad ((1-\Phi(\frac{ln(1/L_d)+(r-\sigma^2/2)T}{\sigma\sqrt{T}}))\\
& \qquad max(S_0L_du^{i_2}d^{n-i_2}-K,0)\\
& \qquad +\Phi(\frac{ln(1/L_u)+(r-\sigma^2/2)T}{\sigma\sqrt{T}})\\
& \qquad max(S_0L_uu^{i_2}d^{n-i_2}-K,0))\\
& + \sum_{i_1=\gamma}^{\beta} \frac{n!}{i_1!(n-i_1)!}p^{i_1}(1-p)^{n-i_1}\\
& \qquad ((1-\Phi(\frac{ln(1/L_d)+(r-\sigma^2/2)T}{\sigma\sqrt{T}}))\\
& \qquad max(S_0u^{i_1}d^{n-i_1}L_d-K,0)\\
& \qquad +\Phi(\frac{ln(1/L_u)+(r-\sigma^2/2)T}{\sigma\sqrt{T}})\\
& \qquad max(S_0u^{i_1}d^{n-i_1}L_u-K,0))\\
& + \sum_{i_1=0}^{\gamma}\sum_{i_2=0}^{\gamma} \mathbf{1}_{i_1,i_2:K<S_0u^{i_1}d^{n-i_1}u^{i_2}d^{n-i_2}}\\
& \qquad \frac{n!}{i_1!(n-i_1)!}\frac{n!}{i_2!(n-i_2)!}p^{i_1+i_2}(1-p)^{n-i_1+n-i_2}\\
& \qquad max(S_0L_dL_d-K,0)\\
& + \sum_{i_1=\beta}^{n}\sum_{i_2=0}^{\gamma} \mathbf{1}_{i_1,i_2:K<S_0u^{i_1}d^{n-i_1}u^{i_2}d^{n-i_2}}\\
& \qquad \frac{n!}{i_1!(n-i_1)!}\frac{n!}{i_2!(n-i_2)!}p^{i_1+i_2}(1-p)^{n-i_1+n-i_2}\\
& \qquad max(S_0L_uL_d-K,0)\\
& + \sum_{i_1=0}^{\gamma}\sum_{i_2=\beta}^{n} \mathbf{1}_{i_1,i_2:K<S_0u^{i_1}d^{n-i_1}u^{i_2}d^{n-i_2}}
\end{aligned}
$$

$$\frac{n!}{i_1!(n-i_1)!}\frac{n!}{i_2!(n-i_2)!}p^{i_1+i_2}(1-p)^{n-i_1+n-i_2}$$
$$max(S_0L_dL_u-K,0)$$
$$+\sum_{i_1=\beta}^{n}\sum_{i_2=\beta}^{n}\quad \mathbf{1}_{i_1,i_2:K<S_0u^{i_1}d^{n-i_1}u^{i_2}d^{n-i_2}}$$
$$\frac{n!}{i_1!(n-i_1)!}\frac{n!}{i_2!(n-i_2)!}p^{i_1+i_2}(1-p)^{n-i_1+n-i_2}$$
$$max(S_0L_uL_u-K,0)\Big].$$

The upper bound

$$\beta := \frac{1}{2}(\frac{ln(L_u)}{\sigma\sqrt{T/n}}+n), \tag{A.49}$$

where $L_u = 1.1$. The lower bound

$$\gamma := \frac{1}{2}(\frac{ln(L_d)}{\sigma\sqrt{T/n}}+n), \tag{A.50}$$

where $L_d = 0.9$. And

$$p = \frac{e^{\frac{rT}{n}}-e^{-\sigma\sqrt{\frac{T}{n}}}}{e^{\sigma\sqrt{\frac{T}{n}}}-e^{-\sigma\sqrt{\frac{T}{n}}}}. \tag{A.51}$$

A.3.1 Simplifying (A.48): the 1st, 2nd, 3rd part

For the 1st, 2nd, 3rd part of (A.48), each model is an un-re-weighted re-weighting model. The first part, which is between price bounds for both phases, is an un-re-weighted 2-phase re-weighting model without ϱ. The result has been shown on (A.48). For the second and third part, these are un-re-weighted 1-phase re-weighting models with two fixed pre-known probability values at upper and lower bound. The $P(u^+)$ and $P(d^-)$ values at upper bound and lower bound have been shown at the beginning of this section on Lemma A.2 and A.3.

Hence, we combine the re-weighting models and summarise them as follows.

If $K \leqslant S_0L_d^2$,

$$c_1 = S_0(\Phi(d_1)-\Phi(d_3))^2 - Ke^{-2rT}(\Phi(d_2)-\Phi(d_4))^2, \tag{A.52}$$

where

$$d_1 = \frac{ln(1/L_d)+(r+\sigma^2/2)T}{\sigma\sqrt{T}}, \tag{A.53}$$

$$d_2 = \frac{ln(1/L_d)+(r-\sigma^2/2)T}{\sigma\sqrt{T}}, \tag{A.54}$$

$$d_3 = \frac{ln(1/L_u)+(r+\sigma^2/2)T}{\sigma\sqrt{T}} \tag{A.55}$$

and

$$d_4 = \frac{ln(1/L_u) + (r - \sigma^2/2)T}{\sigma\sqrt{T}}. \tag{A.56}$$

$$\begin{aligned} c_2 = \; & (S_0 L_d \Phi(d_1) - Ke^{-rT}\Phi(d_2) - S_0 L_d \Phi(d_3) + Ke^{-rT}\Phi(d_4)) \\ & (1 - \Phi(\frac{ln(1/L_d) + (r - \sigma^2/2)T}{\sigma\sqrt{T}})) \\ & + (S_0 L_u \Phi(d_1) - Ke^{-rT}\Phi(d_2) - S_0 L_u \Phi(d_3) + Ke^{-rT}\Phi(d_4)) \\ & (\Phi(\frac{ln(1/L_u) + (r - \sigma^2/2)T}{\sigma\sqrt{T}})). \end{aligned} \tag{A.57}$$

$$\begin{aligned} c_3 = \; & (S_0 L_d \Phi(d_1) - Ke^{-rT}\Phi(d_2) - S_0 L_d \Phi(d_3) + Ke^{-rT}\Phi(d_4)) \\ & (1 - \Phi(\frac{ln(1/L_d) + (r - \sigma^2/2)T}{\sigma\sqrt{T}})) \\ & + (S_0 L_u \Phi(d_1) - Ke^{-rT}\Phi(d_2) - S_0 L_u \Phi(d_3) + Ke^{-rT}\Phi(d_4)) \\ & (\Phi(\frac{ln(1/L_u) + (r - \sigma^2/2)T}{\sigma\sqrt{T}})), \end{aligned} \tag{A.58}$$

where

$$d_1 = \frac{ln(1/L_d) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}, \tag{A.59}$$

$$d_2 = \frac{ln(1/L_d) + (r - \sigma^2/2)T}{\sigma\sqrt{T}}, \tag{A.60}$$

$$d_3 = \frac{ln(1/L_u) + (r + \sigma^2/2)T}{\sigma\sqrt{T}} \tag{A.61}$$

and

$$d_4 = \frac{ln(1/L_u) + (r - \sigma^2/2)T}{\sigma\sqrt{T}}. \tag{A.62}$$

If $S_0 L_u L_d \leqslant K < S_0 L_u^2$,

$$\begin{aligned} c_1 = \; & e^{-rT} S_0 \qquad (\Phi(d_1) - \Phi(d_3))(-\Phi(d_3)) \\ & + e^{-rT} S_0 \lim_{n \longrightarrow \infty} \sum_{i_1 = \zeta}^{\beta} \frac{n!}{(n - i_1)! i_1!} p^{i_1} (1 - p)^{n - i_1} u^{i_1} d^{n - i_1} \Phi(d_1') \\ & - e^{-2rT} K \qquad (\Phi(d_2) - \Phi(d_4))(-\Phi(d_4)) \\ & - e^{-2rT} K \lim_{n \longrightarrow \infty} \sum_{i_1 = \zeta}^{\beta} \frac{n!}{(n - i_1)! i_1!} p^{i_1} (1 - p)^{n - i_1} \Phi(d_2'), \end{aligned} \tag{A.63}$$

where

$$d_1 = \frac{ln(S_0 L_u/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}, \tag{A.64}$$

$$d_1' = \frac{ln(S_0 u^{2i_1 - n}/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}, \tag{A.65}$$

$$d_2 = \frac{ln(S_0 L_u/K) + (r - \sigma^2/2)T}{\sigma\sqrt{T}}, \tag{A.66}$$

$$d_2' = \frac{ln(S_0 u^{2i_1 - n}/K) + (r - \sigma^2/2)T}{\sigma\sqrt{T}}, \tag{A.67}$$

$$d_3 = \frac{ln(1/L_u) + (r + \sigma^2/2)T}{\sigma\sqrt{T}} \tag{A.68}$$

and

$$d_4 = \frac{ln(1/L_u) + (r - \sigma^2/2)T}{\sigma\sqrt{T}}. \tag{A.69}$$

$$\begin{aligned} c_2 =& (S_0 L_u \Phi(d_1) - K e^{-rT}\Phi(d_2) - S_0 L_u \Phi(d_3) + K e^{-rT}\Phi(d_4)) \\ & (\Phi(\frac{ln(1/L_u) + (r - \sigma^2/2)T}{\sigma\sqrt{T}})). \end{aligned} \tag{A.70}$$

$$\begin{aligned} c_3 =& (S_0 L_u \Phi(d_1) - K e^{-rT}\Phi(d_2) - S_0 L_u \Phi(d_3) + K e^{-rT}\Phi(d_4)) \\ & (\Phi(\frac{ln(1/L_u) + (r - \sigma^2/2)T}{\sigma\sqrt{T}})), \end{aligned} \tag{A.71}$$

where

$$d_1 = \frac{ln(S_0 L_u/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}, \tag{A.72}$$

$$d_2 = \frac{ln(S_0 L_u/K) + (r - \sigma^2/2)T}{\sigma\sqrt{T}}, \tag{A.73}$$

$$d_3 = \frac{ln(1/L_u) + (r + \sigma^2/2)T}{\sigma\sqrt{T}} \tag{A.74}$$

and

$$d_4 = \frac{ln(1/L_u) + (r - \sigma^2/2)T}{\sigma\sqrt{T}}. \tag{A.75}$$

If $S_0 L_d^2 < K < S_0 L_d L_u$,

$$
\begin{aligned}
c_1 = \; & e^{-rT} S_0 \quad \Big[(\Phi(d_1) - \Phi(d_3'))(-\Phi(d_3)) \\
& \qquad + \lim_{n \longrightarrow \infty} \sum_{i_1=\gamma}^{\theta} \frac{n!}{(n-i_1)! i_1!} p^{i_1} (1-p)^{n-i_1} u^{i_1} d^{n-i_1} \Phi(d_1') \Big] \\
& - e^{-2rT} K \Big[(\Phi(d_2) - \Phi(d_4'))(-\Phi(d_4)) \\
& \qquad + \lim_{n \longrightarrow \infty} \sum_{i_1=\gamma}^{\theta} \frac{n!}{(n-i_1)! i_1!} p^{i_1} (1-p)^{n-i_1} \Phi(d_2') \Big] \\
& + \quad \Big(S_0 (\Phi(d_3') - \Phi(d_3))(\Phi(d_1) - \Phi(d_3)) \\
& \qquad - K e^{-2rT} (\Phi(d_4') - \Phi(d_4))(\Phi(d_2) - \Phi(d_4)) \Big),
\end{aligned}
\tag{A.76}
$$

where

$$
d_1 = \frac{ln(1/L_d) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}, \tag{A.77}
$$

$$
d_1' = \frac{ln(S_0 u^{2i_1 - n}/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}, \tag{A.78}
$$

$$
d_2 = \frac{ln(1/L_d) + (r - \sigma^2/2)T}{\sigma\sqrt{T}}, \tag{A.79}
$$

$$
d_2' = \frac{ln(S_0 u^{2i_1 - n}/K) + (r - \sigma^2/2)T}{\sigma\sqrt{T}}, \tag{A.80}
$$

$$
d_3 = \frac{ln(1/L_u) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}, \tag{A.81}
$$

$$
d_3' = \frac{ln(S_0 L_d/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}, \tag{A.82}
$$

$$
d_4 = \frac{ln(1/L_u) + (r - \sigma^2/2)T}{\sigma\sqrt{T}} \tag{A.83}
$$

and

$$
d_4' = \frac{ln(S_0 L_d/K) + (r - \sigma^2/2)T}{\sigma\sqrt{T}}. \tag{A.84}
$$

$$
\begin{aligned}
c_2 = \; & (S_0 L_d \Phi(d_1) - K e^{-rT} \Phi(d_2) - S_0 L_d \Phi(d_3) + K e^{-rT} \Phi(d_4)) \\
& (1 - \Phi(\frac{ln(1/L_d) + (r - \sigma^2/2)T}{\sigma\sqrt{T}})) \\
& + (S_0 L_u \Phi(d_1') - K e^{-rT} \Phi(d_2') - S_0 L_u \Phi(d_3) + K e^{-rT} \Phi(d_4)) \\
& (\Phi(\frac{ln(1/L_u) + (r - \sigma^2/2)T}{\sigma\sqrt{T}})).
\end{aligned}
\tag{A.85}
$$

$$
\begin{aligned}
c_3 = \; & (S_0 L_d \Phi(d_1) - K e^{-rT} \Phi(d_2) - S_0 L_d \Phi(d_3) + K e^{-rT} \Phi(d_4)) \\
& (1 - \Phi(\frac{ln(1/L_d) + (r - \sigma^2/2)T}{\sigma\sqrt{T}})) \\
& + (S_0 L_u \Phi(d_1') - K e^{-rT} \Phi(d_2') - S_0 L_u \Phi(d_3) + K e^{-rT} \Phi(d_4)) \\
& (\Phi(\frac{ln(1/L_u) + (r - \sigma^2/2)T}{\sigma\sqrt{T}})),
\end{aligned} \tag{A.86}
$$

where

$$
d_1 = \frac{ln(S_0 L_d/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}, \tag{A.87}
$$

$$
d_2 = \frac{ln(S_0 L_d/K) + (r - \sigma^2/2)T}{\sigma\sqrt{T}}, \tag{A.88}
$$

$$
d_3 = \frac{ln(1/L_u) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}, \tag{A.89}
$$

$$
d_4 = \frac{ln(1/L_u) + (r - \sigma^2/2)T}{\sigma\sqrt{T}}, \tag{A.90}
$$

$$
d_1' = \frac{ln(1/L_d) + (r + \sigma^2/2)T}{\sigma\sqrt{T}} \tag{A.91}
$$

and

$$
d_2' = \frac{ln(1/L_d) + (r - \sigma^2/2)T}{\sigma\sqrt{T}}. \tag{A.92}
$$

A.3.2 Simplifying (A.48): the 4th part

For the 4th part of (A.48), there is no simplified form. Hence, we have to use the original calculation shown on (A.48).

$$
\begin{aligned}
c_4 = \; & \sum_{i_1=0}^{\gamma} \sum_{i_2=0}^{\gamma} \mathbf{1}_{i_1,i_2:K<S_0 u^{i_1} d^{n-i_1} u^{i_2} d^{n-i_2}} \\
& \frac{n!}{i_1!(n-i_1)!} \frac{n!}{i_2!(n-i_2)!} p^{i_1+i_2} (1-p)^{n-i_1+n-i_2} \\
& max(S_0 L_d L_d - K, 0) \\
& + \sum_{i_1=\beta}^{n} \sum_{i_2=0}^{\gamma} \mathbf{1}_{i_1,i_2:K<S_0 u^{i_1} d^{n-i_1} u^{i_2} d^{n-i_2}} \\
& \frac{n!}{i_1!(n-i_1)!} \frac{n!}{i_2!(n-i_2)!} p^{i_1+i_2} (1-p)^{n-i_1+n-i_2} \\
& max(S_0 L_u L_d - K, 0).
\end{aligned}
$$

$$+\sum_{i_1=0}^{\gamma}\sum_{i_2=\beta}^{n}\mathbf{1}_{i_1,i_2:K<S_0u^{i_1}d^{n-i_1}u^{i_2}d^{n-i_2}}$$
$$\frac{n!}{i_1!(n-i_1)!}\frac{n!}{i_2!(n-i_2)!}p^{i_1+i_2}(1-p)^{n-i_1+n-i_2}$$
$$max(S_0L_dL_u-K,0)$$
$$+\sum_{i_1=\beta}^{n}\sum_{i_2=\beta}^{n}\mathbf{1}_{i_1,i_2:K<S_0u^{i_1}d^{n-i_1}u^{i_2}d^{n-i_2}}$$
$$\frac{n!}{i_1!(n-i_1)!}\frac{n!}{i_2!(n-i_2)!}p^{i_1+i_2}(1-p)^{n-i_1+n-i_2}$$
$$max(S_0L_uL_u-K,0).$$

Bibliography

Albuquerque, R. (2012), 'Skewness in stock returns: Reconciling the evidence on firm versus aggregate returns', *The Review of Financial Studies* **25**(5), 1630–1673. May.

Amemiya, T. (1973), 'Regression analysis when the dependent variable is truncated normal', *Econometrica: Journal of the Econometric Society* **41**(6), 997–1016. November.

Arak, M. & Cook, R. E. (1997), 'Do daily price limits act as magnets? the case of treasury bond futures', *Journal of financial services research* **12**(1), 5–20. August, Springer.

Azhar, A. K. M., Osman, S. I. W. & Parinduri, R. A. (2009), On capital market ratios and stock valuation: A geometric approach, Technical report, 22nd Australasian Finance and Banking Conference 2009. November.

Barrett, J. P. (1974), 'The coefficient of determination -some limitations', *The American Statistician* **28**(1), 19–20. Taylor & Francis, American Statistical Association, February.

Basu, S. (1977), 'Investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis', *The journal of finance* **32**(3), 663–682. June.

Black, F. & Scholes, M. (1973), 'The pricing of options and corporate liabilities', *Journal of Political Economy* **81**(3), 637–654. University of Chicago Press, May-June.

Brown, C. A. & Robinson, D. M. (2002), 'Skewness and kurtosis implied by option prices: A correction', *The journal of financial research* **25**(2), 279–282. Summer.

Burden, R. L. & Faires, J. D. (2011), *Numerical Analysis*, 9th edn, Brooks Cole, Cengage Learning, Boston.

Campbell, J. Y. & Shiller, R. J. (1986), Cointegration and tests of present value models, Journal of Political Economy 95(5), National Bureau of Economic Research. 1062-1088.

Campbell, J. Y. & Shiller, R. J. (1988*a*), 'The dividend-price ratio and expectations of future dividends and discount factors', *Review of financial studies* **1**(3), 195–228. Autumn,Soc Financial Studies, Oxford University Press.

Campbell, J. Y. & Shiller, R. J. (1988*b*), 'Stock prices, earnings, and expected dividends', *The Journal of Finance* **43**(3), 661–676. December, Wiley Online Library.

Campbell, J. Y. & Shiller, R. J. (1991), 'Yield spreads and interest rate movements: A bird's eye view', *The Review of Economic Studies* **58**(3), 495–514. The Econometrics of Financial Markets, special issue, Oxford University Press.

Campbell, J. Y. & Shiller, R. J. (1998), 'Valuation ratios and the long-run stock market outlook: ratios are extraordinarily bearish', *The Journal of Portfolio Management* **24**(2), 11–26. Institutional Investor Journals.

Cho, D. D., Russell, J., Tiao, G. C. & Tsay, R. (2003), 'The magnet effect of price limits: evidence from high-frequency data on taiwan stock exchange', *Journal of Empirical Finance* **10**(1-2), 133–168. Elsevier.

Corrado, C. J. & Su, T. (1996*a*), 'Skewness and kurtosis in s&p 500 index returns implied by option prices', *The journal of financial research* **19**(2), 175–192. Summer.

Corrado, C. J. & Su, T. (1996*b*), 'S&p 500 index option tests of jarrow and rudd's approximate option valuation formula', *Journal of Futures Markets* **16**(6), 611–629. September.

Cox, J. C. & Ross, S. A. (1975), The pricing of options for jump processes, Technical Report 2-75, Wharton School Rodney L. White Center for Financial Research. Working paper.

Cox, J. C., Ross, S. A. & Rubinstein, M. E. (1979), 'Option pricing: A simplified approach', *Journal of financial Economics* **7**(3), 229–263. Elsevier, North-Holland, July.

Doob, J. L. (1955), 'Martingales and one-dimensional diffusion', *Transactions of the American Mathematical Society* **78**(1), 168–208. January to April.

Fama, E. F. (1965), 'The behaviour of stock market prices', *Journal of Business* **38**(1), 34–105. January.

Fama, E. F., Fisher, L., Jensen, M. C. & Roll, R. (1969), 'The adjustment of stock prices to new information', *International Economic Review* **10**(1), 1–21. February, Wiley.

Fama, E. F. & French, K. R. (1993), 'Common risk factors in the returns on stocks and bonds', *Journal of Financial Economics* **33**(1), 3–56. February.

Freedman, D., Pisani, R. & Purves, R. (2007), *Statistics*, 4th edn, W. W. Norton & Company, New York.

French, K. R., Schwert, G. W. & Stambaugh, R. F. (1987), 'Expected stock returns and volatility', *Journal of financial Economics* **19**(1), 3–29. December.

Friedmann, R. & Sanddorf-Köhle, W. G. (2007), 'A conditional distribution model for limited stock index returns', *Journal of Economic Dynamics and Control* **31**(3), 721–741. March, Elsevier.

Goedhart, M., Koller, T. & Wessels, D. (2005), 'The right role for multiples in valuation', *McKinsey on Finance* pp. 7–11. Spring.

Gordon, M. J. & Shapiro, E. (1956), 'Capital equipment analysis: the required rate of profit', *Management science* **3**(1), 102–110. October.

Gottwald, R. (2012), 'The use of the p/e ratio to stock valuation', *GRANT journal* **01**(02), 21–24. December.

Graham, B. & Dodd, D. L. (2009), *Security Analysis: Principles and Technique*, Finance, Investing, sixth edn, Whittlesey House, McGraw-Hill Book Co., New York City. Updated with Commentary.

Gui, M. J. (2006), Shanghai stock exchange market quality report, Research Report 9, Shanghai Stock Exchange Innovation Laboratory, Shang Hai. June.

Hansen, L. P. (1982), 'Large sample properties of generalized method of moments estimators', *Econometrica* **50**(4), 1029–1054. July, Econometric Society.

Hong, H. & Stein, J. C. (2003), 'Differences of opinion, short-sales constraints, and market crashes', *The Review of Financial Studies* **16**(2), 487–525. Summer.

Hooper, D., Coughlan, J. & Mullen, M. (2008), 'Structural equation modelling: guidelines for determining model fit', *Electronic Journal of Business Research Methods* **6**(1), 53–60. ARROW at Dublin Institute of Technology, January.

Hu, L. & Bentler, P. M. (1999), 'Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives', *Structural Equation Modeling: A Multidisciplinary Journal* **6**(1), 1–55.

Hull, J. C. (2015), *Options, futures, and other derivatives*, 9th edn, Pearson Education, Upper Saddle River.

Istanbul (2015), 'Markets', https://www.borsaistanbul.com/en/products-and-markets/markets. Borsa Istanbul.

Ivanovski, Z., Stojanovski, T. & Narasanov, Z. (2015), 'Volatility and kurtosis of daily stock returns at mse', *UTMS Journal of Economics* **6**(2), 209–221. August.

Jarrow, R. & Rudd, A. (1982), 'Approximate option valuation for arbitrary stochastic processes', *Journal of financial Economics* **10**(3), 347–369. Elsevier.

Jensen, M. C., Black, F. & Scholes, M. S. (1972), 'The capital asset pricing model: Some empirical tests', *Studies in the Theory of Capital Markets* .

Johnson, N. L., Kotz, S. & Balakrishnan, N. (1994), *Continuous Univariate Distributions*, Vol. 1 of *Wiley Series in Probability and Statistics*, 2nd edn, publisher, New York.

Jolliffe, I. T. (2002), *Principal Component Analysis: With 28 Illustrations*, Vol. Statistical Theory and Methods of *Springer Series in Statistics*, 2nd edn, Springer Science & Business Media, New York.

Joshi, M. S. (2009), 'The convergence of binomial trees for pricing the american put', *The journal of risk* **11**(4), 87–108. June.

Ki, H., Choi, B., Chang, K.-H. & Lee, M. (2005), 'Option pricing under extended normal distribution', *Journal of Futures Markets: Futures, Options, and Other Derivative Products* **25**(9), 845–871. July, Wiley.

Kodres, L. E. (1993), 'Tests of unbiasedness in the foreign exchange futures markets: An examination of price limits and conditional heteroscedasticity', *Journal of Business* **66**(3), 463–490. July.

Leduc, G. (2013), 'A european option general first-order error formula', *The ANZIAM Journal* **54**(4), 248–272. Cambridge University Press.

Lin, S.-H., Huang, H.-H. & Li, S.-H. (2015), 'Option pricing under truncated gram-charlier expansion', *The North American Journal of Economics and Finance* **32**, 77–97. April, Elsevier.

Malkiel, B. G. & Fama, E. F. (1970), 'Efficient capital markets: A review of theory and empirical work', *The Journal of Finance* **25**(2), 383–417. May, Blackwell Publishing Ltd.

Mendenhall, W. & Sincich, T. T. (2012), *A Second Course in Statistics: Regression Analysis*, Upper-Level Statistics, 7th edn, Prentice Hall, London. January.

Merton, R. C. (1973), 'Theory of rational option pricing', *The Bell Journal of economics and management science* **4**(1), 141–183. Spring.

Miranda, L. C. & Riera, R. (2001), 'Truncated lévy walks and an emerging market economic index', *Physica A: Statistical Mechanics and its Applications* **297**(3-4), 509–520. 15 August, Elsevier.

Molenberghs, G., Fitzmaurice, G., Kenward, M. G., Tsiatis, A. & Verbeke, G., eds (2014), *Handbook of Missing Data Methodology*, Chapman & Hall/CRC Handbooks of Modern Statistical Methods, Taylor & Francis. CRC Press.

Nissim, D. & Penman, S. H. (2001), 'Ratio analysis and equity valuation: From research to practice', *Review of accounting studies* **6**(1), 109–154. Springer.

Osborne, M. F. M. (1959), 'Brownian motion in the stock market', *Operations Research* **7**(2), 145–173. March-April.

Ozer, D. J. (1985), 'Correlation and the coefficient of determination', *Quantitative Methods in Psychology* **97**(2), 307–315. Psychological Bulletin, American Psychological Association.

Rubinstein, M. E. (1998), 'Edgeworth binomial trees', *The Journal of Derivatives* **5**(3), 20–27. Institutional Investor Journals, Spring.

Shiller, R. J. (1979), 'The volatility of long-term interest rates and expectations models of the term structure', *The Journal of Political Economy* **87**(6), 1190–1219. December, The University of Chicago Press.

Shiller, R. J. (1981), 'Do stock prices move too much to be justified by subsequent change in dividends?', *The American Economic Review* **71**(3), 421–436. June, American Economic Association.

Shiller, R. J. (2005), *Irrational exuberance*, 2nd edn, Princeton University Press, New Jersey.

Skorokhod, A. V. (1961), 'Stochastic equations for diffusion processes in a bounded region', *Theory of Probability & Its Applications* **6**(3), 264–274. Society for Industrial and Applied Mathematics, Translated by N. Greenleaf.

SSE (2018), 'Trading rules of shanghai stock exchange', http://english.sse.com.cn/start/rules/sse/trading/. Shanghai Stock Exchange.

Taylor, R. (1990), 'Interpretation of the correlation coefficient: A basic review', *Journal of Diagnostic Medical Sonography* **6**(1), 35–39. Sage Journals, January.

TSE (2019), 'Operating rules of the taiwan stock exchange corporation', https://twse-regulation.twse.com.tw/. Taiwan Stock Exchange.

Walpole, R. E., Myers, R. H., Myers, S. L. & Ye, K. (2012), *Probability & statistics for engineers & scientists*, Engineering—Statistical methods, 9th edn, Prentice Hall, Boston. Pearson Education, Inc.

Williams, J. B. (1938), *The Theory of Investment Value*, Harvard University Press.

Zhu, S.-P. & He, X.-J. (2018), 'A modified black–scholes pricing formula for european options with bounded underlying prices', *Computers & Mathematics with Applications* **75**(5), 1635–1647. March, Elsevier.