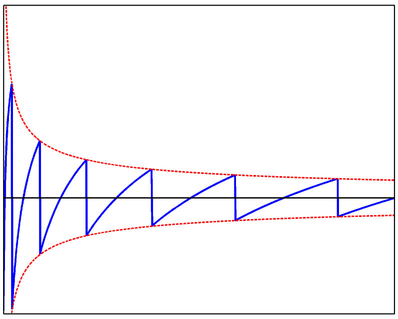
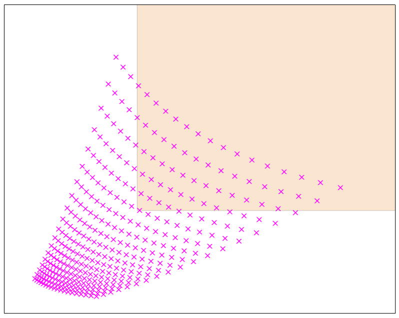
TECHNISCHE UNIVERSITÄT KAISERSLAUTERN

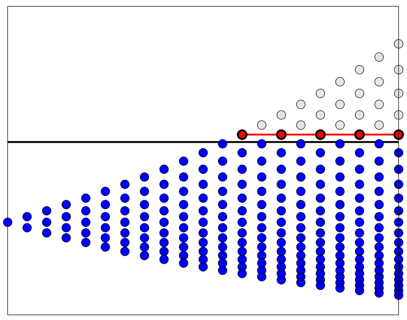
The Binomial Approach to Option Valuation

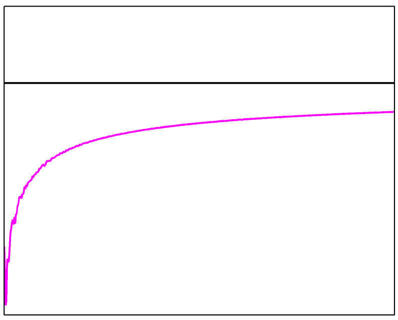
Getting Binomial Trees into Shape

Stefanie Müller









1. Gutachter: Prof. Dr. Ralf Korn
2. Gutachter: Prof. L.C.G. Rogers, Ph.D.

Datum der Disputation: 16. Dezember 2009

Stefanie Müller
Department of Mathematics
University of Kaiserslautern
Erwin-Schrödinger-Straße
67663 Kaiserslautern

stefanie@mathematik.uni-kl.de

The Binomial Approach to Option Valuation

Getting Binomial Trees into Shape

Stefanie Müller

Vom Fachbereich Mathematik der Technischen Universität Kaiserslautern zur Verleihung des akademischen Grades Doktor der Naturwissenschaften (Doctor rerum naturalium, Dr. rer. nat.) genehmigte Dissertation.

1. Gutachter: Prof. Dr. Ralf Korn
2. Gutachter: Prof. L.C.G. Rogers, Ph.D.

Datum der Disputation: 16. Dezember 2009

D386

Acknowledgements

First, I would like to express my deep gratitude to my supervisor, Prof. Dr. Ralf Korn. His support and advice were invaluable to me. I am especially grateful for his close co-operation while working on our joint papers.

I am also very grateful to Prof. Dr. Chris Rogers for accepting to act as a referee for this thesis and for his hospitality during my research visits to the Cambridge Statistical Laboratory. I benefited a lot from his feedback.

I deeply wish to thank Frank Seifried and my father Stefan Müller for their innumerable discussions with me. I very much appreciated their suggestions. Further, many thanks go to Dr. Kalina Natcheva-Acar. In my first year, Kalina kindly shared her valuable experiences in computational finance with me.

I would like to thank my colleagues and friends at the department and at the Fraunhofer ITWM; in particular, Dr. Peter Diesinger, Prof. Dr. Jörn Sass and Nicole Tschauder; Dr. Matthias Fengler and Juniorprof. Dr. Natalie Packham for invitations to the Trading & Derivatives group of Sal. Oppenheim and to the Frankfurt School of Finance & Management; Prof. Dr. Mark Broadie and Dr. Mike Staunton for valuable comments and software code; Prof. Dr. Wolfgang Runggaldier, whose AARMS summer course intensified my interest in research on mathematical finance.

Finally, I gratefully acknowledge the financial support of the Rheinland-Pfalz cluster of excellence DASMOD and the Center for Computational and Mathematical Modelling.

To my parents.

Contents

Chapter 1

Introduction

The discipline of mathematical finance consists of two main subjects: portfolio optimisation and derivative pricing. This thesis deals with an aspect of the latter: The binomial approach to the valuation of financial derivatives. The binomial method is an important technique for numerical option valuation, and the increase in complexity of financial products has further expanded its fields of application. Moreover, since the very beginning of derivative pricing theory, the binomial approach has been of an enormous pedagogical use. In this introductory chapter, we briefly sketch the history of derivative pricing with a particular focus on the binomial approach and its growing fields of application. Afterwards we give an overview of this thesis.

What is derivative pricing? Financial instruments include both primary financial instruments such as stocks, bonds and currencies, and derivative securities, whose value is derived from an underlying. The underlying can be a primary financial instrument, a reference value from the market such as interest rates and indexes, a commodity or - to put it bluntly - in principle anything you could possibly bet on or hedge against. To give an example for an everyday ("non-banking") derivative; from April 1 to June 7, 2008, Deutsche Bahn offered a *Fan BahnCard 25* at EUR 39 / EUR 19 (first / second class) that promised a 25%-discount on train fares until June 31, 2008. In addition, Deutsche Bahn promised that the discount card's duration would be extended by one month for each match that the German soccer team would win during the European Championship in Austria/Switzerland. Clearly, at the purchasing date, the buyer did not know the number of matches that the German team finally won. Thus, this was clearly a bet on the competitiveness of the German soccer team.

In derivative pricing, the "fair price" of a certain derivative contract is determined. Due to the Fundamental Theorem of Asset Pricing, we nowadays know that, if the market is complete and arbitrage-free, there exists a unique equivalent martingale measure under which the price p of a derivative contract on some underlying S with payoff g and maturity T is given by

$$p = E_Q \left(e^{-rT} g(S_t; 0 \le t \le T) \right), \tag{1.1}$$

where r is the risk-free rate. However, it took a lot of time and effort to arrive at this abstract form.

Pricing options with the no-arbitrage principle: The binomial approach and its economic interpretation Early crucial steps to abstract pricing theory were made by R.C. Merton and by F. Black and M.S. Scholes in the early 1970s, who formulated the following no-arbitrage principle:

"If options are correctly priced in the market, it should not be possible to make sure profits by creating portfolios of long and short positions in options and their underlying stocks." (F. Black, M.S. Scholes [BS73])

Based on the no-arbitrage principle, they derived a theoretical valuation formula for European calls and puts on a log-normally distributed stock price. From today's point of view, twelve years after Scholes and Merton received the Nobel prize in economic science (Black had already died at that time), it is almost unbelievable that the Black-Scholes paper had been rejected by two other journals before it was published in the *Journal of Political Economy*. However, in the early 1970s, their ideas were so non-standard and revolutionary that it was hard to believe in them.

In 1975 M. Rubinstein and W.F. Sharpe, himself Nobel Price laureate (1990), had the following discussion on the Black-Scholes model at a conference in Ein Borek, Israel:

"With nothing to do during the breaks (except to take a dip in the sea), ..., we wondered how it was that the then two-year-old Black-Scholes approach to valuing options could recreate a riskless payoff using only the option and its underlying asset. It was then that Sharpe said, I wonder if it's really that there are only two states of the world, but three securities, so that any one of the securities can be replicated by the other two." (M. Rubinstein [Rub92])

This insight was the birth of the binomial approach. Rubinstein and Sharpe realised that by the Central Limit Theorem, the Black-Scholes formula occurs as the limiting form of the corresponding price in a discrete model with successive two-state up-down movements of the underlying asset price. Subsequently, the binomial approach to option pricing theory was presented in Sharpe's textbook *"Investments"* [Sha79] and the model was explained in detail in *"Option pricing: a simplified approach"* [CRR79] by J.C. Cox, S.A. Ross and M. Rubinstein. For many economists, the binomial approach actually justified the continuous-time modelling of Black and Scholes. And even nowadays, it is of an enormous pedagogical use. Here the economic concepts of arbitrage opportunities and market completeness are much easier to understand than in the continuous-time case.

More importantly, the binomial approach became widely used as a numerical pricing tool for American and exotic options when an analytic pricing formula is not available. This is explained below.

American and exotic options: The binomial approach as a numerical pricing tool The option pricing formula (1.1) is only valid for European-type options. European options can only be exercised at one specified date T in the future, the maturity. However, since the early days of trading, numerous option types traded in exchanges belong to the class of American options. They can be exercised at any time between the purchase date and the expiration date. Due to the widespread use of American options, it is important to find appropriate methods to determine their fair price. However, the small conceptual difference between European- and American-style options causes a big difference in pricing because the optimal exercise date $0 \leq \tau_* \leq T$ is not known on the date of purchase. Rather, it depends on the random evolution of the stock price, and it is hence itself random (mathematically, τ_* is a stopping time with respect to the filtration generated by S). For American-style options, the pricing formula (1.1) must be modified to

$$p = E_Q\left(e^{-r\tau_*} g(S_t; 0 \leq t \leq \tau_*)\right). \tag{1.2}$$

However, as τ_* is uncertain, the formula does not readily provide a monetary value for a specific valuation problem. In fact, the American valuation problem continues to engage both researchers in academics and professionals. Ross (1987) writes in the *New Palgrave Dictionary of Economics*:

"This does not mean that there are no important gaps in the (option pricing) theory. Perhaps of most importance, beyond numerical results, ..., very little is known about most American options which expire in finite time .... Despite such gaps, when judged by its ability to explain the empirical data, option pricing theory is the most successful theory not only in finance, but in all of economics."

In contrast to the continuous-time American valuation problem, the American valuation problem can be solved explicitly in the binomial approach. Let us explain: In the Black-Scholes model, the stock price follows a geometric Brownian motion, an infinite variation process. However, in the binomial approach, randomness is modelled on a discrete grid in both time and space. This simplifies the valuation problem considerably because there is only a finite number of possible scenarios. Therefore, we can go step by step backwards in time and decide at each scenario whether it is optimal to exercise or not. Weighting our decisions with respect to the risk-neutral measure leads to the price of the American option in the binomial model. As shown by K. Amin and A. Khanna (1994), the price estimates obtained from the binomial approach converge to the American option price in the Black-Scholes market [AK94]. For a sufficiently large number of periods, the binomial price serves as an estimate for the continuous-time price. As a consequence, the binomial approach became important as a numerical pricing tool for American options.

As explained above, the option price is simply the expected value of a functional of the stock price (compare with the pricing formula (1.1) and (1.2)). In the binomial model, the price is *the expected value of the same function of a simpler process that approximates the original stock price, but that is only driven by discrete random events*. In particular, for numerical option valuation, it is irrelevant whether the sequence of price estimates obtained from the binomial model has the economic interpretation as option prices in the associated discrete markets. Therefore, the transition probabilities in the binomial model need not be risk-neutral; rather, it suffices if the sequence of binomial processes converges weakly to the continuous-time stock price. Early suggestions for this kind of binomial models are made in the paper by Cox, Ross and Rubinstein and also in *"Two-state option pricing"* by R.J. Rendleman and B.J. Bartter. Their paper appeared around the same time as the paper by Cox, Ross and Rubinstein, but has not received the same attention. As a main difference, for the model suggested by Rendleman and Bartter, the probability for moving upwards and downwards is the same [RB79].

For American options, the exercise time can be chosen by the option buyer. Alternatively, one could think of more complex payoff structures. Towards the end of the 1970s and the beginning of the 1980s, standard option trading became better understood and the trading volume exploded. Financial institutions began to search for alternative forms of options - called exotics, special-purpose options or customer-tailored options - meeting the new requirements of the customers [Zha98]. The increase in complexity of the options' structure led to an increasing demand for numerical pricing algorithms, which enhanced the scope of the binomial approach as a pricing tool. As for American options, the price in the binomial model can be determined for any desired structure of the payoff by calculating all possible scenarios and weighting them with respect to an appropriate measure.

Path-dependent options have been of particular interest among these second-generation options; in the late 1990s, they became the most popular options in the OTC market place [Zha98]. Here the payoff depends on the entire path of the underlying asset. Due to Donsker's Theorem, a process version of the Central Limit Theorem, the binomial approach leads to prices for path-dependent options that converge to the option price in the Black-Scholes model. Therefore, the binomial approach can be used a pricing tool for path-dependent options.

Multi-asset options, i.e. options depending on several underlyings, form another important class of exotic derivatives. In the course of increasing cross-market integration and globalisation in financial markets, multi-asset options have become popular to hedge cross-market and global positions [Zha98]. Consequently, since the late 1980s and early 1990s, there have been numerous approaches to adapt the binomial method to the valuation of multi-asset options. Though differing in details, most suggestions are based on a discretisation of the joint evolution of the stock price process [P.P. Boyle (1988); P.P. Boyle, J. Evnine and S. Gibbs (1989), B. Kamrad and P. Ritchken (1991),...]. For multi-asset options however, the binomial approach suffers from the *curse of dimensionality*, i.e the computational effort grows exponentially in the number of underlying assets. Consequently, for high-dimensional valuation problems, the binomial method is currently not practically useful. This is an inherent drawback of the binomial approach as a method based on the discretisation of the underlying assets. However, up to dimension four, let us say, the binomial approach can lead to results that are perfectly competitive and often superior to those obtained by Monte Carlo methods.

Although the binomial approach is, in principle, an efficient method for lower dimensional valuation problems, there are at least two main problems regarding its application: Firstly, binomial methods often exhibit an irregular convergence behaviour of the option prices computed for an increasing number of periods N. Furthermore, traded options often exhibit discontinuities, so that the Berry-Esséen inequality on the sequence of binomial price estimates is in general tight; i.e. conventional tree methods converge no faster than in order $1/\sqrt{N}$. Unfortunately, the fact that the payoff is non-smooth also causes an irregular convergence behaviour that impedes the possibility to achieve a higher of order of convergence via extrapolation methods. The most prominent example of irregular behaviour is the so-called sawtooth-effect. Secondly, in multi-asset markets conventional tree construction methods cannot ensure well-defined transition probabilities for arbitrary correlation structures between the assets. As a major aim of the thesis, we present two approaches to ”get binomial trees into shape”; *the optimal drift model for the valuation of single-asset options* and *the decoupling approach to multi-dimensional option pricing*.

The optimal drift model is presented as a new binomial scheme for single-asset option pricing. It can lead to convergence of order $o(1/N)$ by exploiting the specific structure of the valuation problem under consideration. *The optimal drift model has the potential to outperform even benchmark methods* such as the binomial scheme suggested by D.P.J. Leisen and M. Reimer, which is widely used in practice for American option pricing [LR96]. The decoupling approach is presented as a construction method for multi-dimensional trees. In contrast to the standard approach to multi-dimensional trees, *the trees constructed according to the decoupling approach are well-defined for an arbitrary correlation structure of the underlying assets*. In addition, they yield a more regular convergence behaviour. In fact, the sawtooth effect can even vanish completely, so that extrapolation can be applied. In contrast to the optimal drift model, *the decoupling approach does not assume knowledge of the valuation problem under consideration*. We do not claim that the decoupling approach performs best for any particular type of (exotic) option. However, it shows a strong overall convergence behaviour when applied to arbitrary options. *The decoupling approach is therefore an easy and universal approach to cope with the irregular convergence behaviour in multiple dimensions*. By contrast, *the optimal drift model is based on an advanced construction* *technique that shows superior performance if adapted to a specific valuation problem.*

An overview of the thesis This thesis consists of two parts; binomial pricing in a single-asset Black-Scholes market and its extension to multi-dimensional situations. In Chapter 2, we give a thorough and rigorous overview of the binomial approach to numerical option valuation for a single underlying. We summarise, order and comment on many results from literature. Some of these are well-known, while others are non-standard. In order to complete the picture of the binomial model, we add many pieces of our own work. Furthermore, we introduce the optimal drift model. It is defined by optimising the allocation of probability mass in relation to the strike value, as suggested by Y.S. Tian and by L.-B. Chang and K. Palmer [Tia99], [CP07]. However, while both these models are imposed on the scheme suggested by Cox, Ross and Rubinstein, we *optimise the drift of the underlying binomial model*. As a result, the optimal drift model can admit convergence of order $o(1/N)$.

In Chapters 3 and 4, we investigate the multi-asset case. Chapter 3 deals with the standard approach to binomial option pricing in a multi-asset Black-Scholes market. *Standard methods are based on an approximation of the joint evolution of the underlying assets*. This will be explained in detail. The standard approach is illustrated with the model suggested by P.P. Boyle, J. Evnine and S. Gibbs which canonically extends the one-dimensional model by Cox, Ross and Rubinstein to a multi-dimensional situation [BEG89]. In addition, we consider a multi-dimensional variant of the model suggested by Rendleman and Bartter. In order to obtain an appropriate approximation to the multi-asset Black-Scholes model under consideration, the correlation structure among the assets has to be matched. Consequently, the number of moment matching conditions grows quadratically in the dimension. As a result, *if we follow the standard approach to multi-dimensional trees, setting up an appropriate binomial model soon gets tedious, and it is sometimes even impossible.* On top of that, conventional multi-dimensional tree construction methods inherit the irregular convergence behaviour observed for the one-dimensional situation.

In Chapter 4, the decoupling approach is introduced as an alternative to binomial pricing of multi-asset options. *The basic idea of the decoupling approach is to transform the multi-dimensional (log-normal) asset price process to a new process with independent components before setting up a discrete model*. The model we suggest contains the 2D example by J. Hull as a special case [Hul06]. Decoupling is an easy approach to overcome the main problems in applications of the standard approach to multi-dimensional trees.

Chapter 4 is essentially based on the paper *"The decoupling approach to binomial pricing of multi-asset options"* published in the *Journal of Computational Finance* [KM09a]. Short extracts from both parts of the thesis are collected in the paper *"Getting multi-dimensional trees into a new shape"*, which has recently appeared in the *Wilmott Journal* [KM09b]. Both papers are joint work with Ralf Korn.

Chapter 2

Binomial Pricing for Single-Asset Options

We consider a one-dimensional Black-Scholes model with stock price dynamics under the risk-neutral measure given by

$$dS(t) = S(t)(r\mathrm{d}t + \sigma \mathrm{d}W_t), \qquad S(0) = s_0 > 0, \tag{2.1}$$

for some volatility parameter $\sigma > 0$. Here r is the risk-free interest rate, and W is a one-dimensional Brownian motion with respect to the risk-neutral measure Q. We fix a time horizon $T > 0$.

2.1 Introduction

This chapter deals with binomial pricing of single-asset options. The underlying stock is assumed to follow the Black-Scholes dynamics defined above. From a theoretical perspective, we therefore investigate two-state Markov chain approximations to a geometric Brownian motion. To apply the binomial approach to numerical option pricing, we want the approximating models to ensure weak convergence to the stock price process in the Black-Scholes setting. Under this condition, the corresponding sequence of binomial price estimates converges to the exact option price for most common payoff structures. However, as discussed in the introductory section, conventional binomial schemes suffer from serious problems in practical applications. The convergence behaviour of the corresponding price estimates is non-monotone and oscillatory. Furthermore, if the payoff exhibits discontinuities, the Berry-Esséen bound on the convergence rate of the pricing error is in general tight. We therefore focus on the construction of binomial approximations that can exploit the structure of the valuation problem under consideration.

Let us briefly outline the contents of this chapter: We first discuss alternative binomial models for the approximation of the stock price process. In particular, we consider the schemes suggested by Cox, Ross and Rubinstein (1979) (for short: CRR) and by Rendleman and Bartter (1979) (for short: RB). All schemes under consideration are defined so as to asymptotically match the first two moments of the logreturns of the stock price. It is well-known that by asymptotic moment matching, the approximating models converge weakly to the stock price process as the step size in the discrete model tends to zero. This is discussed in detail. Afterwards, we investigate the asymptotic behaviour of the discretisation error. We demonstrate that for conventional schemes, the Berry-Esséen bound is tight and the discretisation error converges non-smoothly. This motivates to control the error term, as is done in many advanced models. In particular, we discuss a generalised variant of the advanced models suggested by Tian (1999) and by Chang and Palmer (2007). The latter leads to the new optimal drift model. We will verify that the optimal drift model can admit a superior convergence rate of the discretion error. In Section 2.3, we focus on the application of the binomial approach to numerical option pricing. We see that due to weak convergence, the binomial method can be applied to numerical valuation of most common types of European and American options. The corresponding valuation algorithm is called a *tree procedure* because the possible realisations of the binomial process can be identified with a tree structure. The implementation of binomial option pricing is discussed in detail in Section 2.4. Finally, we analyse the convergence behaviour of binomial trees for the two most common payoff structures; for payoffs that are constant in the terminal value $S(T)$ (i.e. cash-or-nothing options) and for payoffs that are linear in $S(T)$ (i.e. plain vanilla options). We see that amongst the methods under consideration, the shape of the tree constructed by the optimal drift model best exploits the structure of the valuation problem. This results in a superior rate of convergence of the corresponding pricing error.

2.2 Discretisation of the Stock Price and Weak Convergence

In the following, we discuss alternative binomial schemes for the approximation of the stock price process. In the context of numerical option pricing, we want the approximating models to ensure weak convergence to the stock price process S. That is,

Definition 1 (Weak Convergence)**.** *Let M be a metric space and let $P^{(N)}$, $1 \leq N < \infty$, and P be probability measures on $(M, \mathscr{B}(M))$, where we write $\mathscr{B}(M)$ for the Borel σ-field of M; i.e. the smallest σ-field containing all open subsets of M. Then we say that the sequence of probability measures $(\mathbf{P}^{(\mathbf{N})})_{\mathbf{N}}$ **converges weakly to** $\mathbf{P}$, denoted by*

$$P^{(N)} \Rightarrow_w P,$$

if for any bounded, continuous function $f : M \to \mathbb{R}$*, we have*

$$\lim_{N\to\infty} \int_M f(x) P^{(N)}(\mathrm{d}x) = \int_M f(x) P(\mathrm{d}x).$$

Further, let $X^{(N)}$*,* $1 \le N < \infty$*, and* X *be random variables with state space* M *defined on probability spaces* $(\Omega^{(N)}, \mathscr{F}^{(N)}, P^{(N)})$ *and* $(\Omega, \mathscr{F}, P)$*, respectively. Then we say that the sequence of random variables* $(\mathbf{X}^{(\mathbf{N})})_{\mathbf{N}}$ ***converges weakly to*** $\mathbf{X}$*, denoted by*

$$X^{(N)} \Rightarrow_w X,$$

if for any bounded, continuous function $f : M \to \mathbb{R}$*, we have,*

$$\lim_{N\to\infty} E_{P^{(N)}}(f(X^{(N)})) = E_P(f(X))$$

(compare e.g. [Bil68]).

Remark 1. *Of course, weak convergence of random variables is the same as weak convergence of their distributions. In particular, the random variables under consideration need not to be defined on the same probability space, as is used below.*

2.2.1 Binomial Models

Let $N \in \mathbb{N}$ denote the number of periods in the discrete model. A binomial approximation to the stock price allows for two possible scenarios per period, so that the path space is naturally given by

$$\mathscr{E}^{(N)} := \{\omega : \{1, \ldots, N\} \to \{1, -1\}\}$$

endowed with the product σ-field

$$\mathscr{F}^{(N)} := \bigotimes_{k=1}^{N} \mathscr{P}(\{1, -1\}) := \sigma\left(Z_k^{(N)} \,\middle|\, k \in N\right).$$

Here $\mathscr{P}(.)$ denotes the power set and $Z_k^{(N)} : \mathscr{E}^{(N)} \to \{1, -1\}$ is the coordinate mapping $Z_k^{(N)}(\omega) = \omega_k$.

Starting at the initial value of the continuous-time process s_0, we define a binomial process on $(\mathscr{E}^{(N)}, \mathscr{F}^{(N)})$ by

$$S_{k+1}^{(N)} := S_k^{(N)} e^{\alpha(N)\Delta t + \beta\sqrt{\Delta t}Z_{k+1}^{(N)}} \qquad \forall k = 0, \ldots, N-1, \tag{2.2}$$

for some constant $\beta > 0$ and some constant $\alpha(N)$ depending on the number of periods N. Here $\Delta t := T/N$ is the grid size of the discrete-time model. In order to achieve weak convergence to the continuous-time price process, we choose the sequence $(\alpha(N))_N$, the constant $\beta > 0$ and the sequence of probability measures $(P^{(N)})_N$ on $(\mathscr{E}^{(N)}, \mathscr{F}^{(N)})_N$ such that the following conditions are satisfied:

1. For all $N \in \mathbb{N}$, the random variables $Z_k^{(N)}$, $k = 1, \ldots, N$, (for short: RV) are independently and identically distributed (for short: i.i.d.).

2. The first two moments of the one-period logreturns of S are asymptotically matched, i.e. we have that

$$\begin{aligned}\mu(N): &= \frac{1}{\Delta t}\,\mathrm{E}_{P^{(N)}}\left(\ln\left(\frac{S_{k+1}^{(N)}}{S_k^{(N)}}\right) \,\middle|\, S_k^{(N)}\right)\\ &= \alpha(N) + \beta\sqrt{\frac{1}{\Delta t}}\,\mathrm{E}_{P^{(N)}}\left(Z_{k+1}^{(N)}\right)\end{aligned} \tag{2.3}$$

$$\begin{aligned}\sigma^2(N) &:= \frac{1}{\Delta t}\,\mathrm{Var}_{P^{(N)}}\left(\ln\left(\frac{S_{k+1}^{(N)}}{S_k^{(N)}}\right) \,\middle|\, S_k^{(N)}\right)\\ &= \beta^2\,\mathrm{Var}_{P^{(N)}}\left(Z_{k+1}^{(N)}\right)\end{aligned} \tag{2.4}$$

are such that as $N \to \infty$

$$\mu(N) \to r - \frac{1}{2}\sigma^2 \tag{2.5}$$

and

$$\sigma^2(N) \to \sigma^2. \tag{2.6}$$

Under these conditions, it follows from Donsker's Invariance Principle (compare e.g. [Bil68], Theorem 10.1) that the linear interpolation of $S^{(N)}$ (suitably scaled in time) converges weakly to the stock price process. Of course, the first condition is satisfied if and only if $P^{(N)}$ is a product measure of the form

$$P^{(N)} = \bigotimes_{k=1}^{N} P^{(1,N)} \tag{2.7}$$

for some probability measure $P^{(1,N)}$ on $(\{1,-1\},\mathscr{P}(\{1,-1\}))$. Then one-step transitions are independent and they are the same for each period, which we assume from now on. In particular, $S^{(N)}$ is a Markov process.

Introducing the notations $p_u(N) := P^{(1,N)}(\{1\})$ and $p_d(N) := P^{(1,N)}(\{-1\})$, the characteristics of the discrete-time model $\mu(N)$ and $\sigma^2(N)$ write as

$$\mu(N) = \alpha(N) + \beta\sqrt{\tfrac{1}{\Delta t}}\,(2p_u(N)-1) \tag{2.8}$$

and

$$\sigma^2(N) = 4\beta^2 p_u(N)(1-p_u(N)). \tag{2.9}$$

Further, $p_d(N)$ is determined by $p_d(N) = 1-p_u(N)$. Apparently, we can choose among an infinite number of possible discretisation schemes that ensure the moment matching conditions (2.5) and (2.6). Let us remark that the drift parameters $(\alpha(N))_N$ are allowed to be non-constant in N. This will provide some additional flexibility to adapt the binomial scheme to the payoff structure of interest (compare e.g. [Tia99]); more on that to come later. We impose the following condition:

Assumption 1. *The sequence $(\alpha(N))_N$ is assumed to be bounded; i.e. it is assumed to be of order $O(1)$.*

In the following, we investigate some widely used binomial schemes. Firstly, we consider transition probabilities that are given by the risk-neutral measure associated with the discrete market consisting of a stock with dynamics (2.2) and a bond

$$B_{k+1} = B_k e^{r\Delta t}, \qquad k = 0,\ldots,N-1; \qquad B_0 = 1. \tag{2.10}$$

Here r is the interest rate in the continuous-time model. Let us anticipate that under the risk-neutral measure, the moment matching conditions (2.5) and (2.6) are satisfied if and only if $\beta = \sigma$. By contrast, the drift of the discrete-time model is irrelevant. As explained in the introductory chapter, this approach features an economic insight on option pricing in the Black-Scholes model. When we approximate an option on the continuous-time stock price by evaluating a payoff functional along the sequence of binomial models, the resulting price estimates are themselves option prices in the approximating binomial model.

Secondly, we analyse the discretisation schemes suggested by RB and by CRR[1]. Here

[1] The binomial approach suggested by CRR is motivated by its economic insight. However, they also introduce a binomial model with transition probabilities that are only asymptotically identical with the risk-neutral transition probabilities [CRR79]. In literature, the term CRR model appears both for

the binomial model serves as a plain numerical pricing tool; the binomial estimates for an option on the stock price S do not admit an economic interpretation. In particular, they do not coincide with the corresponding option prices in the binomial model. Having clarified this conceptual difference to risk-neutral discretisation schemes, we will use the term ”binomial prices” loosely.

Throughout the thesis, we typically distinguish between the discretisation schemes presented above. Below we additionally investigate advanced schemes that allow for a better performance in numerical option valuation.

Discretisation Schemes with Risk-neutral Transition Probabilities In the following, we consider the discrete-time market consisting of a bond with one-period return $e^{r\Delta t}$ (compare (2.10)) and the stock $S^{(N)}$ with dynamics (2.2). The possible one-period returns of the stock price are denoted by

$$u(N) := e^{\alpha(N)\Delta t+\beta\sqrt{\Delta t}} \qquad \text{and} \qquad d(N) := e^{\alpha(N)\Delta t-\beta\sqrt{\Delta t}}.$$

By our convention $\beta > 0$, $u(N)$ can be interpreted as the one-period return given that ”the economy is in the good state 1”, and $d(N)$ is the realised one-period return if ”the economy is in the bad state -1”. Then there is the following well-known result (compare e.g. [Bjö04], Section 2 and Section 3):

Proposition 1. *We have absence of arbitrage opportunities (for short: AAO) in the discrete market if and only if*

$$d(N) < e^{r\Delta t} < u(N). \tag{2.11}$$

In this case, the discrete market is also complete and the risk-neutral probability measure is given by $Q^{(N)} = \bigotimes_{k=1}^{N} Q^{(1,N)}$, *where*

$$Q^{(1,N)}(1) := \tfrac{e^{r\Delta t}-d(N)}{u(N)-d(N)}. \tag{2.12}$$

Note that (AAO) implies in particular that the measure $Q^{(N)}$ is well-defined.

If we write the condition (2.11) in terms of $\alpha(N)$ and β, we have (AAO) if and only if

$$|r-\alpha(N)|\sqrt{T/N} < \beta.$$

the discretisation scheme with risk-neutral transition probabilities and for the discretisation scheme with transition probabilities defined by $p_u(N) = 1/2 + 1/2(\frac{r-1/2\sigma^2}{\sigma}\sqrt{T/N})$. In the thesis, we use the term CRR model for the latter variant.

By Assumption 1, there is some constant $M \geq 0$ such that $|\alpha(N)| \leq M$ for all $N \in \mathbb{N}$. Thus, we can formulate the following result:

Corollary 1. *Provided that the grid size is sufficiently small, the discrete market is arbitrage-free and complete. It suffices to let*

$$N > \frac{(r+M)^2 T}{\beta^2}.$$

By Corollary 1, we may agree on the assumption that the grid size is always sufficiently small to ensure that risk-neutral transition probabilities are well-defined and unique.

As we will see in the following, when we define transition probabilities in accordance with the risk-neutral measure, weak convergence to the continuous-time stock price S is already ensured if we set $\beta = \sigma$; i.e. it suffices to ensure that the discrete-time stock price is exposed to shocks of appropriate size. Yet this condition is also necessary. By contrast, the drift is irrelevant.

Proposition 2. *Assume that the transition probabilities are determined according to the risk-neutral measure $Q^{(N)}$; i.e. we have $Q^{(N)} = \bigotimes_{k=1}^{N} Q^{(1,N)}$ with $Q^{(1,N)}$ defined in (2.12). Then the moment matching conditions (2.5) and (2.6) are satisfied if and only if $\beta = \sigma$.*
In particular, convergence to the first two moments of the one-period logreturns in the continuous-time model is of order

$$\left|\mu_{Q^{(N)}}(N) - (r - \tfrac{1}{2}\sigma^2)\right| = O(\tfrac{1}{N}) \qquad \text{and} \qquad \left|\frac{\sigma^2}{\sigma^2_{Q^{(N)}}(N)} - 1\right| = O(\tfrac{1}{N}),$$

where $\mu_{Q^{(N)}}(N)$ and $\sigma^2_{Q^{(N)}}(N)$ are computed with respect to $Q^{(N)}$ according to (2.3) and (2.4), respectively.

Proof. By Assumption 1, it follows from a Taylor expansion that

$$\begin{aligned} u(N) \;&=\; 1 + \beta\left(\tfrac{T}{N}\right)^{1/2} + \left(\alpha(N) + \tfrac{1}{2}\beta^2\right)\tfrac{T}{N} + \left(\tfrac{1}{6}\beta^3 + \alpha(N)\beta\right)\left(\tfrac{T}{N}\right)^{3/2} + \\ &\quad \left(\tfrac{1}{2}\alpha^2(N) + \tfrac{1}{2}\beta^2\alpha(N) + \tfrac{1}{24}\beta^4\right)\left(\tfrac{T}{N}\right)^2 + o\left(\tfrac{1}{N^2}\right) \end{aligned} \tag{2.13}$$

$$\begin{aligned} d(N) \;&=\; 1 - \beta\left(\tfrac{T}{N}\right)^{1/2} + \left(\alpha(N) + \tfrac{1}{2}\beta^2\right)\tfrac{T}{N} - \left(\tfrac{1}{6}\beta^3 + \alpha(N)\beta\right)\left(\tfrac{T}{N}\right)^{3/2} + \\ &\quad \left(\tfrac{1}{2}\alpha^2(N) + \tfrac{1}{2}\beta^2\alpha(N) + \tfrac{1}{24}\beta^4\right)\left(\tfrac{T}{N}\right)^2 + o\left(\tfrac{1}{N^2}\right), \end{aligned} \tag{2.14}$$

which implies that $q_u(N) := Q^{(1,N)}(1)$ is of the form

$$q_u(N) = \tfrac{1}{2} + c_1(N)\left(\tfrac{T}{N}\right)^{1/2} + c_3(N)\left(\tfrac{T}{N}\right)^{3/2} + o\left(\tfrac{1}{N^{3/2}}\right), \tag{2.15}$$

where

$$\begin{aligned} c_1(N) &= \tfrac{1}{2\beta}\left(r - \alpha(N) - \tfrac{1}{2}\beta^2\right) \\ c_3(N) &= \tfrac{1}{2\beta}\left(\tfrac{1}{2}(\alpha(N)-r)^2 + \tfrac{1}{6}\beta^2(\alpha(N)-r) + \tfrac{1}{24}\beta^4\right) \end{aligned}$$

(compare also [CP07], p. 97/98). Note that $c_1(N)$ and $c_3(N)$ are of order $O(1)$. We obtain

$$\mu_{Q^{(N)}}(N) \overset{(2.8)}{=} \alpha(N) + \beta\left(\tfrac{N}{T}\right)^{1/2}(2q_u(N)-1) = r - \tfrac{1}{2}\beta^2 + 2\beta c_3(N)\tfrac{T}{N} + o\left(\tfrac{1}{N}\right)$$

and

$$\sigma^2_{Q^{(N)}}(N) \overset{(2.9)}{=} \beta^2 4 q_u(N)(1-q_u(N)) = \beta^2 - 4\beta^2 c_1^2(N)\tfrac{T}{N} + o\left(\tfrac{1}{N}\right).$$

Hence, the moment matching conditions (2.5) and (2.6) are satisfied if and only if $\beta = \sigma$. Moreover, we see from the above equations that the assertion on the order of convergence holds true, which completes the proof. □

Remark 2. *Let us stress that the moment matching conditions are satisfied independently of the particular choice of the sequence* $(\alpha(N))_N$.

Further, equation (2.15) implies the following result on the asymptotic behaviour of the risk-neutral measure:

Corollary 2. *As the number of periods N tends to infinity,*

$$q_u(N) \to \tfrac{1}{2}.$$

The Discretisation Scheme suggested by RB RB suggest to set

$$\alpha(N) = \alpha := r - \tfrac{1}{2}\sigma^2, \qquad \beta = \sigma$$

and

$$p_u(N) = p_d(N) = \tfrac{1}{2}. \tag{2.16}$$

Apparently, for this choice the transition probabilities do not depend on the number of periods N. As required, the model satisfies the moment matching conditions (2.5) and (2.6). In fact, moments are not only matched asymptotically, but for any number of periods N; i.e.

$$\mu(N) = r - \tfrac{1}{2}\sigma^2 \qquad \text{and} \qquad \sigma^2(N) = \sigma^2.$$

The Discretisation Scheme suggested by CRR CRR define an appropriate binomial model via

$$\alpha(N) = \alpha := 0, \qquad \beta = \sigma$$

and

$$p_u(N) = \tfrac{1}{2} + \tfrac{1}{2}\left(\frac{r - \frac{1}{2}\sigma^2}{\sigma}\right)\sqrt{\tfrac{T}{N}}. \tag{2.17}$$

In contrast to the RB model, the transition probabilities vary in the number of periods N. In particular, the corresponding probability measures $P^{(N)}$ is only well-defined if the grid size is sufficiently small, which we assume throughout this thesis. To be precise, we need

$$N \geq \tfrac{1}{\sigma^2}\left(r - \tfrac{1}{2}\sigma^2\right)^2 T. \tag{2.18}$$

Simple calculations show that the moment matching conditions (2.5) and (2.6) are satisfied. In fact, if the model is well-defined for any number of periods N, the first moment of the logreturns is matched exactly. By contrast, the second moment is matched asymptotically only. We have

$$\sigma^2(N) = \sigma^2 - \left(r - \tfrac{1}{2}\sigma^2\right)^2 \tfrac{T}{N}, \tag{2.19}$$

which yields

$$\left|\frac{\sigma^2}{\sigma^2(N)} - 1\right| = O\left(\tfrac{1}{N}\right). \tag{2.20}$$

Note that as the grid size tends to zero, the one-step transition probabilities in the CRR model converge to the corresponding one-step transition probabilities in the RB model. In particular,

$$p_u^{CRR}(N) = p_u^{RB}(N) + O\left(\frac{1}{\sqrt{N}}\right).$$

Further, neither in the CRR model nor in the RB model, the transition probabilities are risk-neutral. However as readily observed from (2.15), there is the following asymptotic relationship between $p_u(N)$ and $q_u(N)$:

Proposition 3. *For the discretisation scheme suggested by RB and by CRR, the one-step transition probability $p_u(N)$, defined in (2.16) and in (2.17), respectively, coincides with the associated risk-neutral probability $q_u(N)$ up to a term of order $1/N^{3/2}$. We have that for the RB model*

$$\begin{aligned} q_u^{RB}(N) &= \tfrac{1}{2} + \tfrac{1}{24}\,\sigma^3 \left(\tfrac{T}{N}\right)^{3/2} + o\left(\tfrac{1}{N^{3/2}}\right) \\ &= p_u^{RB}(N) + \tfrac{1}{24}\,\sigma^3 \left(\tfrac{T}{N}\right)^{3/2} + o\left(\tfrac{1}{N^{3/2}}\right) \end{aligned}$$

and for the CRR model

$$\begin{aligned} q_u^{CRR}(N) &= \tfrac{1}{2} + \tfrac{1}{2\sigma}\left(r - \tfrac{1}{2}\sigma^2\right)\tfrac{T}{N} + \tfrac{1}{4\sigma}\left((r - \tfrac{1}{6}\sigma^2)^2 + \tfrac{1}{18}\,\sigma^4\right)\left(\tfrac{T}{N}\right)^{3/2} + o\left(\tfrac{1}{N^{3/2}}\right) \\ &= p_u^{CRR}(N) + \tfrac{1}{4\sigma}\left((r - \tfrac{1}{6}\sigma^2)^2 + \tfrac{1}{18}\,\sigma^4\right)\left(\tfrac{T}{N}\right)^{3/2} + o\left(\tfrac{1}{N^{3/2}}\right). \end{aligned}$$

In particular, Proposition 3 implies the following result:

Corollary 3. *Both for the models suggested by CRR and by RB, there is some $N_0 \in \mathbb{N}$ such that*

$$q_u(N) \geq p_u(N) \qquad \text{for all } N \geq N_0.$$

In words: For a sufficiently large number of periods N, the chosen probability for an up-movement $p_u(N)$ is smaller than the corresponding risk-neutral probability $q_u(N)$.

Let us stress that while for the discretisation scheme suggested by RB, the drift enters transition states, it enters transition probabilities for the CRR model. As a consequence, the latter implies that the logarithm of the binomial process is symmetric around the starting value. By contrast, the discretisation scheme suggested by RB prefers to have equal weights, which is paid for by a non-symmetric form of the binomial process.

Finally, let us anticipate some aspects of the application of the two models to numerical option pricing. Firstly, we wish to stress that computing binomial price estimates is cheaper for the RB model because every path has exactly the same probability. For the CRR model, we have to multiply each payment with its *specific* probability of occurrence. Secondly, the application of the CRR model suffers from the fact that the requirement (2.18) sets a lower bound on the number of periods N. In particular, if the volatility is small, the transition probabilities are only well-defined for a relative large number of periods. To illustrate this, let us consider the input parameters $r = 0.1$, $T = 1$ and $\sigma = 0.01$. Then the number of periods N is bounded from below by 100. For $\sigma = 0.001$, the required number of periods is already bounded by 10000.

Time-scaling In the following, we map the binomial process $S^{(N)}$ onto a continuous-time process $S^{(c,N)}$ that appropriately approximates the stock price process S on $[0,T]$ so that weak convergence is ensured. By moment matching, it only remains to scale time appropriately; that is, we apply the transformation $k \mapsto k\Delta t$. By linear interpolation, we set

$$S^{(c,N)}(t) := \exp\left(\left(\left[\tfrac{N}{T}t\right] + 1 - \tfrac{N}{T}t \right) \ln\left(S^{(N)}_{\left[\frac{N}{T}t\right]} \right) + \left(\tfrac{N}{T}t - \left[\tfrac{N}{T}t\right] \right) \ln\left(S^{(N)}_{\left[\frac{N}{T}t\right]+1} \right) \right), \quad (2.21)$$

where $[x]$ denotes the greatest integer less or equal to x.

Then for $t = k\Delta t$, $S^{(c,N)}(t)$ has the same distribution as $S^{(N)}_k$ and for $t \in (k\Delta t, (k+1)\Delta t)$, $\ln(S^{(c,N)}(t))$ is obtained by linear interpolation between $\ln(S^{(N)}_k)$ and $\ln(S^{(N)}_{k+1})$. To motivate the time-scaling applied, note that for $s = k\Delta t$ and $t = (k+1)\Delta t$, i.e. $t - s = \Delta t$, the log-increment $\ln(S^{(c,N)}(t)/S^{(c,N)}(s))$ (asymptotically) matches the first two moments of the log-returns of the stock price process over a period of length Δt.

Weak convergence Due to the Central Limit Theorem, the fact that the process $S^{(c,N)}$ is based on the binomial distribution becomes negligible in the limit, so that for all times $t \in [0,T]$, $S^{(c,N)}(t)$ converges in distribution to the time-t value of the stock price S_t. Moreover, for $s = k\Delta t$ and $t = (k+1)\Delta t$, the log-increment $\ln(S^{(c,N)}(t)/S^{(c,N)}(s)) = \alpha(N)\Delta t + \beta Z^{(N)}_{k+1}\sqrt{\Delta t}$ is independent of $\sigma(S^{(c,N)}(u); 0 \le u \le s) = \sigma(Z^{(N)}_1, \ldots, Z^{(N)}_k)$. Hence, together with the moment matching conditions, we anticipate that the following result holds:

Proposition 4. *The sequence of approximating processes $(S^{(c,N)})_N$ converges weakly to the geometric Brownian motion S; for short we write*

$$S^{(c,N)} \Rightarrow_w S.$$

Proposition 4 is a key result in the theory of numerical option pricing. It is a simple consequence of Donsker's Invariance Principle that provides a process version of the Central Limit Theorem. This is discussed in the following. Yet the assertion is not at all non-trivial because all bits of hard work, in particular proving tightness, are hidden in the invariance principle.

Since the RVs $Z^{(N)}_k$, $k = 1, \ldots, N$, are i.i.d. for fixed N only, we need a special variant of Donsker's Theorem that considers triangular schemes:

Theorem 1 (The Invariance Principle for Triangular Schemes)**.** *Let $\xi_{N1},\ldots,\xi_{Nk_N}$ be i.i.d. with mean* 0 *and variance* $0 < \sigma_{N1}^2 < \infty$*; put* $S_{Ni} = \sum_{l=1}^{i} \xi_{Nl}$*,* $s_{Ni}^2 = i\sigma_{N1}^2$*, and* $s_N^2 = s_{Nk_N}^2$*. Let* $X^{(N)}$ *be the random function that is linear on each interval* $[s_{N,i-1}^2/s_N^2, s_{Ni}^2/s_N^2]$ *and has values* $X^{(N)}(s_{Ni}^2/s_N^2) = S_{Ni}/s_N$ *at the grid points. Then,* $X^{(N)}$ *converges weakly to a Brownian motion (compare e.g. [Bil68], problem 1, p. 77).*

Remark 3. *Note in particular that the above variant of the invariance principle is applicable to the binomial model suggested by CRR, for which the one-step transition probabilities depend on the number of periods N.*

In our application, $S^{(c,N)}$ is a continuous function of the embedded process $X^{(N)}$ defined above. But weak convergence is preserved under continuous mappings. In addition, it is known by Slutsky's Theorem that an asymptotic matching of moments suffices:

Theorem 2 (Continuous Mapping Principle)**.** *Let M and M' be metric spaces. Let X and $X^{(N)}$,* $1 \le N < \infty$*, be M-valued RVs and let $h : M \to M'$ be continuous. Then, if $X^{(N)} \Rightarrow_w X$, it holds that $h \circ X^{(N)} \Rightarrow_w h \circ X$ (compare e.g. [Bil68], Theorem 5.1).*

Theorem 3 (Slutsky's Theorem)**.** *Let (M,d) be a metric space. Let $(X_1^{(N)}, X_2^{(N)})_N$ be a sequence of $(M \times M)$-valued RVs defined on a probability space $(\Omega^{(N)}, \mathscr{F}^{(N)}, P^{(N)})$. Suppose that $X_1^{(N)} \Rightarrow_w X_1$ for some M-valued RV X_1. If for all* $\varepsilon > 0$,

$$P^{(N)}\left(d(X_1^{(N)}, X_2^{(N)}) > \varepsilon\right) \to 0,$$

then $X_2^{(N)} \Rightarrow_w X_1$ (compare e.g. [EK86], Corollary 3.3.3.).

Proof of Proposition 4. In order to re-write the dynamics of the process $Y^{(c,N)}$ in terms of normalised RVs, we define

$$Y_k^{(N)} = \frac{1}{\sqrt{\mathrm{Var}_{P^{(N)}}\left(Z_k^{(N)}\right)}} \left(Z_k^{(N)} - \mathrm{E}_{P^{(N)}}(Z_k^{(N)})\right), \tag{2.22}$$

so that $\mathrm{E}_{P^{(N)}}(Y_k^{(N)}) = 0$ and $\mathrm{Var}_{P^{(N)}}(Y_k^{(N)}) = 1$. Then

$$\begin{aligned} S_{k+1}^{(N)} &= S_k^{(N)} \exp\left(\alpha(N)\Delta t + \beta\sqrt{\Delta t}\,\mathrm{E}_{P^{(N)}}(Z_{k+1}^{(N)}) + \beta\sqrt{\mathrm{Var}_{P^{(N)}}\left(Z_{k+1}^{(N)}\right)}\sqrt{\Delta t}\,Y_{k+1}^{(N)}\right) \\ &= S_k^{(N)} \exp\left(\mu(N)\Delta t + |\sigma(N)|\sqrt{\Delta t}\,Y_{k+1}^{(N)}\right), \end{aligned}$$

which implies that the process $S^{(c,N)}$ writes as

$$S^{(c,N)}(t) = s_0 e^{\mu(N)t+|\sigma(N)|\sqrt{T}\,Y^{(c,N)}\left(\frac{t}{T}\right)}, \qquad t \in [0,T], \tag{2.23}$$

with

$$Y^{(c,N)}(s) = \tfrac{1}{\sqrt{N}} \left(\sum_{k=1}^{[Ns]} Y_k^{(N)} + (Ns - [Ns])\, Y_{[Ns]+1}^{(N)} \right), \qquad s \in [0,1]. \tag{2.24}$$

Let us define

$$X_1^{(c,N)}(t) := \left(r - \tfrac{1}{2}\sigma^2\right)t + \sigma\sqrt{T}\,Y^{(c,N)}\left(\tfrac{t}{T}\right)$$

and

$$X_2^{(c,N)}(t) := \ln\left(S_t^{(c,N)}\right) = \mu(N)t + |\sigma(N)|\sqrt{T}\,Y^{(c,N)}\left(\tfrac{t}{T}\right).$$

Then

$$\left\|X_1^{(c,N)} - X_2^{(c,N)}\right\|_\infty = \sup_{t\in[0,T]} \left|(r - \tfrac{1}{2}\sigma^2 - \mu(N))t + (\sigma - |\sigma(N)|)\sqrt{T}\,Y^{(c,N)}\left(\tfrac{t}{T}\right)\right|$$

measures ”the impact of an asymptotic matching of moments” in terms of the sup-norm, i.e. $||f||_\infty = \sup_{t\in[0,T]}|f(t)|$. Let $\varepsilon > 0$. Apparently,

$$\begin{aligned} P^{(N)}\left(\left\|X_1^{(c,N)} - X_2^{(c,N)}\right\|_\infty > \varepsilon\right) &\le P^{(N)}\left(\left|\mu(N) - (r - \tfrac{1}{2}\sigma^2)\right| T > \tfrac{\varepsilon}{2}\right) + \\ &\quad P^{(N)}\left(\left\|(\sigma - |\sigma(N)|)\sqrt{T}\,Y^{(c,N)}\left(\tfrac{t}{T}\right)\right\|_\infty > \tfrac{\varepsilon}{2}\right). \end{aligned} \tag{2.25}$$

For a sufficiently large number of periods N, the first term on the right-hand side of equation (2.25) is zero by asymptotic moment matching. It remains to investigate the second term. Clearly,

$$\begin{aligned} P^{(N)}\left(\left\|(\sigma - |\sigma(N)|)\sqrt{T}\,Y^{(c,N)}\left(\tfrac{t}{T}\right)\right\|_\infty > \tfrac{\varepsilon}{2}\right) &= \\ P^{(N)}\left(|\sigma - |\sigma(N)||\sqrt{T}\max_{k=1,\dots,N}\left|M_k^{(N)}\right| > \tfrac{\varepsilon}{2}\right), \end{aligned}$$

where the discrete process $M^{(N)}$ is defined by

$$M_k^{(N)} := \frac{1}{\sqrt{N}} \sum_{i=1}^{N} Y_i^{(N)}, \quad 1 \le k \le N, \qquad M_0^{(N)} := 0. \tag{2.26}$$

As $M^{(N)}$ is a discrete martingale, it follows from Doob's martingale inequality (compare e.g. [Dur05], Theorem 4.4.2) that

$$\begin{aligned} P^{(N)}\left(|\sigma - |\sigma(N)|| \sqrt{T} \max_{k=1,\dots,N} \left| M_k^{(N)} \right| > \frac{\varepsilon}{2} \right) &\le \frac{4}{\varepsilon^2} (\sigma - |\sigma(N)|)^2 T E_{P^{(N)}} \left(\left| M_N^{(N)} \right|^2 \right) \\ &= \frac{4}{\varepsilon^2} (\sigma - |\sigma(N)|)^2 T. \end{aligned} \tag{2.27}$$

By the moment matching conditions, the right-hand side of inequality (2.27) tends to zero as N tends to infinity, which implies that

$$P^{(N)}\left(\left\| X_1^{(c,N)} - X_2^{(c,N)} \right\|_\infty > \varepsilon \right) \to 0.$$

Consequently, according to Slutsky's Theorem it suffices to consider the process $X_1^{(c,N)}$ instead of $X_2^{(c,N)}$.

Further, the invariance principle for triangular schemes applies to the sequence $(Y^{(c,N)})_N$; i.e.

$$\{Y^{(c,N)}(s)\}_{\{s \in [0,1]\}} \Rightarrow_w \{B(s)\}_{\{s \in [0,1]\}}, \tag{2.28}$$

where $\{B_s, \tilde{\mathscr{F}}_s; 0 \le s \le 1\}$ is a Brownian motion. Define the time-change $t : [0,1] \to [0,T]$ by

$$t : s \to Ts$$

and the filtration $\mathscr{F}_t := \tilde{\mathscr{F}}_{t/T}$ indexed according to the "new time". Then by the time-scaling property of Brownian paths (compare [KS98], Lemma 2.9.4), $\{W_t, \mathscr{F}_t; 0 \le t \le T\}$ with

$$W_t := \sqrt{T} B(t/T) \qquad 0 \le t \le T$$

is again a Brownian motion. Consequently, according to the continuous mapping principle, (2.28) implies that

$$s_0 \exp\left(X_1^{(c,N)} \right) \Rightarrow_w S,$$

which proves the assertion. □

2.2.2 Distributional Fit

In the following, we investigate how well the process $S^{(c,N)}$ fits the stock price process S for a fixed time $t \in [0,T]$; i.e. we are interested in the distance

$$d^{(N)}(t,x) := \left| P^{(N)}\left(S_t^{(c,N)} \leq x\right) - Q\left(S_t \leq x\right) \right|.$$

The Minimal Convergence Rate

The Berry-Esséen inequality suggests that the distance $d^{(N)}(t,x)$ converges to zero in order $1/\sqrt{N}$ uniformly in $x \in \mathbb{R}$. Yet in our application, the moments are only asymptotically matched. We see in the following that if the moments themselves converge in order $1/\sqrt{N}$, the Berry-Esséen bound is maintained.

Theorem 4 (Berry-Esséen inequality)**.** *Let $X_1,\ldots,X_N$ be independent RVs such that $EX_j = 0$, $E|X_j|^3 < \infty$ $(j = 1,\ldots,N)$. We write*

$$\sigma_j^2 = EX_j^2, \qquad B_N = \sum_{j=1}^{N} \sigma_j^2, \qquad F^{(N)}(x) = P\left(B_N^{-1/2} \sum_{j=1}^{N} X_j < x\right)$$

and

$$L^{(N)} = B_N^{-3/2} \sum_{j=1}^{N} E|X_j|^3.$$

Then,

$$\sup_x \left| F^{(N)}(x) - \Phi(x) \right| \leq A L^{(N)},$$

where $\Phi(x)$ denotes the standard normal distribution function and A is some positive constant (compare e.g. [Pet75], Theorem 5.3).

Proposition 5. *Suppose that $S^{(N)}$ is the binomial process (2.2) with*

$$|\mu(N) - \mu| = O\left(\frac{1}{\sqrt{N}}\right) \qquad \text{and} \qquad \left|\frac{\nu^2}{\sigma^2(N)} - 1\right| = O\left(\frac{1}{\sqrt{N}}\right) \tag{2.29}$$

for some constants $\mu, \nu \in \mathbb{R}$, $\nu > 0$. *Then for all times* $t \in [0,T]$,

$$\sup_x \left| P^{(N)}\left(S^{(c,N)}(t) < x\right) - \Phi\left(\frac{\ln\left(\frac{x}{s_0}\right) - \mu t}{\nu\sqrt{t}}\right)\right| = O\left(\frac{1}{\sqrt{N}}\right). \tag{2.30}$$

Remark 4. *For* $\mu = r - 1/2\sigma^2$ *and* $\nu = \sigma$, *the condition (2.29) is a sharpening of the moment matching conditions (2.5) and (2.6). It additionally requires a minimal order of convergence.*

Proof of Proposition 5. Note that

$$\Phi\left(\frac{\ln\left(\frac{x}{s_0}\right) - \mu(N)t}{|\sigma(N)|\sqrt{t}}\right) = \Phi\left(\frac{\ln\left(\frac{x}{s_0}\right) - \mu t}{\nu\sqrt{t}}\,\varepsilon_1^{(N)} + \varepsilon_2^{(N)}\right), \tag{2.31}$$

where $\varepsilon_i^{(N)}$ $(i = 1,2)$ are the correction terms

$$\varepsilon_1^{(N)} = \frac{\nu}{|\sigma(N)|} \qquad \text{and} \qquad \varepsilon_2^{(N)} = \frac{(\mu - \mu(N))t}{|\sigma(N)|\sqrt{t}}$$

that appear because the constants μ and ν are only asymptotically matched. By the triangular inequality,

$$\begin{aligned}\sup_x &\left| P^{(N)}\left(S^{(c,N)}(t) < x\right) - \Phi\left(\frac{\ln\left(\frac{x}{s_0}\right) - \mu t}{\nu\sqrt{t}}\right)\right| \leq \\ &\sup_x \left| P^{(N)}\left(S^{(c,N)}(t) < x\right) - \Phi\left(\frac{\ln\left(\frac{x}{s_0}\right) - \mu(N)t}{|\sigma(N)|\sqrt{t}}\right)\right| + \\ &\sup_x \left| \Phi\left(\frac{\ln\left(\frac{x}{s_0}\right) - \mu(N)t}{|\sigma(N)|\sqrt{t}}\right) - \Phi\left(\frac{\ln\left(\frac{x}{s_0}\right) - \mu t}{\nu\sqrt{t}}\right)\right|.\end{aligned}$$

Using definition of the correction terms $\varepsilon_1^{(N)}$ and $\varepsilon_2^{(N)}$, we further observe that

$$\begin{aligned}\sup_x &\left| \Phi\left(\frac{\ln\left(\frac{x}{s_0}\right) - \mu(N)t}{|\sigma(N)|\sqrt{t}}\right) - \Phi\left(\frac{\ln\left(\frac{x}{s_0}\right) - \mu t}{\nu\sqrt{t}}\right)\right| = \\ &\sup_x \left| \Phi\left(\frac{\ln\left(\frac{x}{s_0}\right) - \mu t}{\nu\sqrt{t}}\,\varepsilon_1^{(N)} + \varepsilon_2^{(N)}\right) - \Phi\left(\frac{\ln\left(\frac{x}{s_0}\right) - \mu t}{\nu\sqrt{t}}\right)\right| \leq \\ &\sup_x \left| \Phi\left(x + \varepsilon_2^{(N)}\right) - \Phi(x)\right| + \sup_x \left| \Phi\left(x\,\varepsilon_1^{(N)}\right) - \Phi(x)\right|\end{aligned}$$

Thus, in total,

$$\sup_x \left| P^{(N)}\left(S^{(c,N)}(t) < x\right) - \Phi\left(\frac{\ln\left(\frac{x}{s_0}\right) - \mu t}{v\sqrt{t}}\right)\right| \leq$$
$$\sup_x \left| P^{(N)}\left(S^{(c,N)}(t) < x\right) - \Phi\left(\frac{\ln\left(\frac{x}{s_0}\right) - \mu(N)t}{|\sigma(N)|\sqrt{t}}\right)\right| +$$
$$\sup_x \left|\Phi\left(x + \varepsilon_2^{(N)}\right) - \Phi(x)\right| + \sup_x \left|\Phi\left(x\varepsilon_1^{(N)}\right) - \Phi(x)\right| \quad (2.32)$$

In the following, we investigate each term on the right-hand side of the inequality above. The first term can be written as

$$\sup_x \left| P^{(N)}\left(S^{(c,N)}(t) < x\right) - \Phi\left(\frac{\ln\left(\frac{x}{s_0}\right) - \mu(N)t}{|\sigma(N)|\sqrt{t}}\right)\right| =$$
$$\sup_x \left| P^{(N)}\left(S^{(c,N)}(t) < s_0 e^{\mu(N)t + |\sigma(N)|\sqrt{t}x}\right) - \Phi(x)\right| =$$
$$\sup_x \left| P^{(N)}\left(\frac{1}{\sqrt{tN/T}}\left(\sum_{j=1}^{[tN/T]} Y_j^{(N)} + \left(t\frac{N}{T} - \left[t\frac{N}{T}\right]\right) Y_{[tN/T]+1}^{(N)}\right) < x\right) - \Phi(x)\right|,$$

where $Y_k^{(N)}$ are the normalised RVs defined in (2.22). Hence, we observe that

$$\sup_x \left| P^{(N)}\left(S^{(c,N)}(t) < x\right) - \Phi\left(\frac{\ln\left(\frac{x}{s_0}\right) - \mu(N)t}{|\sigma(N)|\sqrt{t}}\right)\right| \leq$$
$$\max\left\{ \sup_x \left| P^{(N)}\left(\frac{1}{\sqrt{tN/T}}\left(\sum_{j=1}^{[tN/T]} Y_j^{(N)} + \sqrt{t\frac{N}{T} - \left[t\frac{N}{T}\right]}\left|Y_{[tN/T]+1}^{(N)}\right|\right) < x\right) - \Phi(x)\right|,\right.$$
$$\left.\sup_x \left| P^{(N)}\left(\frac{1}{\sqrt{\left[t\frac{N}{T}\right]+1}}\left(\sum_{j=1}^{[tN/T]} Y_j^{(N)} - \left|Y_{[tN/T]+1}^{(N)}\right|\right) < x\right) - \Phi(x)\right|\right\}$$

Note from applying the Berry-Esséen inequality to both arguments on the right-hand side of the inequality above that

$$\sup_x \left| P^{(N)}\left(S^{(c,N)}(t) < x\right) - \Phi\left(\frac{\ln\left(\frac{x}{s_0}\right) - \mu(N)t}{|\sigma(N)|\sqrt{t}}\right)\right| = O\left(\frac{1}{\sqrt{N}}\right). \quad (2.33)$$

It remains to consider the last two terms on the right-hand side of inequality (2.32) that are due to the presence of the correction terms $\varepsilon_1^{(N)}$ and $\varepsilon_2^{(N)}$. By assumption (2.29), we have

$$\left(\varepsilon_1^{(N)}\right)^2 = 1 + h(N),$$

where for sufficiently large N,

$$|h(N)| \leq \tfrac{c}{N}$$

for some constant $c > 0$. Consequently, we obtain by the Binomial Series Theorem that

$$\left|\varepsilon_1^{(N)} - 1\right| = O\left(\tfrac{1}{\sqrt{N}}\right). \tag{2.34}$$

Further,

$$\left|\varepsilon_2^{(N)}\right| = \left|\tfrac{(\mu-\mu(N))t}{\nu\sqrt{T}}\varepsilon_1^{(N)}\right| \leq \left|\tfrac{(\mu-\mu(N))t}{\nu\sqrt{T}}\right| \left|\varepsilon_1^{(N)} - 1\right| + \left|\tfrac{(\mu-\mu(N))t}{\nu\sqrt{T}}\right| = O\left(\tfrac{1}{\sqrt{N}}\right). \tag{2.35}$$

By the Mean-Value Theorem, there is some $\Theta_{x,1}^{(N)} \in [0,1]$ such that

$$\begin{aligned}\sup_x \left|\Phi\left(x\varepsilon_1^{(N)}\right) - \Phi(x)\right| &\leq \sup_x \left|\Phi'\left(x + \Theta_{x,1}^{(N)}\left(\varepsilon_1^{(N)} - 1\right)x\right)x\right| \left|\varepsilon_1^{(N)} - 1\right| \\ &\leq \sup_x |\Phi'(x)\,x| \sup_x \left|\tfrac{x}{x+\Theta_{x,1}^{(N)}\left(\varepsilon_1^{(N)}-1\right)x}\right| \left|\varepsilon_1^{(N)} - 1\right|.\end{aligned}$$

As $\Phi'(x)x$ is bounded, it follows from (2.34) that

$$\sup_x \left|\Phi\left(x\varepsilon_1^{(N)}\right) - \Phi(x)\right| = O\left(\tfrac{1}{\sqrt{N}}\right). \tag{2.36}$$

Similarly, there is $\Theta_{x,2}^{(N)} \in [0,1]$ such that

$$\sup_x \left|\Phi\left(x + \varepsilon_2^{(N)}\right) - \Phi(x)\right| \leq \sup_x \left|\Phi'\left(x + \Theta_{x,2}^{(N)}\,\varepsilon_2^{(N)}\right)\right| \left|\varepsilon_2^{(N)}\right|.$$

Hence, by boundedness of $\Phi'(x)$, (2.35) implies that

$$\sup_x \left|\Phi\left(x + \varepsilon_2^{(N)}\right) - \Phi(x)\right| = O\left(\tfrac{1}{\sqrt{N}}\right). \tag{2.37}$$

Finally, the assertion follows by combining the results (2.33), (2.37) and (2.36). □

Let $\mu = r - 1/2\sigma^2$ and $\nu = \sigma$. As discussed previously, for the RB model the moment matching conditions (2.5) and (2.6) are satisfied exactly, i.e. they hold for any number of periods N. By contrast, for the CRR model the volatility is asymptotically matched only. Here $\sigma^2/\sigma^2(N)$ converges to one in order $1/N$. Further according to Proposition 2, if we use risk-neutral transition probabilities, the corresponding characteristics $\mu(N)$ and $\sigma^2(N)$ of the discrete-time model converge in order $1/N$ if and only if $\beta = \sigma$. Hence by Proposition 5, for the models under consideration, convergence to the limiting moments is sufficiently fast to maintain the minimal convergence rate $1/\sqrt{N}$ suggested by the Berry-Esséen inequality:

Corollary 4. *Let $S^{(N)}$ be the process suggested by CRR, the process suggested by RB or any binomial process (2.2) with $\beta = \sigma$ and risk-neutral transition probabilities. Then for any time $t \in [0,T]$,*

$$\sup_x \left| P^{(N)}\left(S^{(c,N)}(t) < x\right) - Q\left(S(t) < x\right)\right| = O\left(\frac{1}{\sqrt{N}}\right) \tag{2.38}$$

Remark 5. *The Berry-Esséen inequality sets a lower bound on the convergence rate of the discretisation error in the approximation to the stock price. It depends on the specific distribution of the discrete-time model whether this minimal convergence rate is attained. Of course, the convergence rate of the discretisation error can also be faster.*

For the RB model, it is easy to observe that the Berry-Esséen inequality is tight:

Proposition 6. *Let $S_N^{(N)}$ be the terminal value of the RB model. Then the distributional fit at the median of the continuous-time model $x = s_0 e^{(r-1/2\sigma^2)T}$ is of the following order:*

$$\left| P^{(N)}\left(S_N^{(N)} < s_0 e^{\left(r-1/2\sigma^2\right)T}\right) - \frac{1}{2}\right| = 0 \qquad \text{for } N \text{ odd},$$

and

$$\lim_{\substack{N\to\infty \\ N \text{ even}}} \frac{\left| P^{(N)}\left(S_N^{(N)} < s_0 e^{\left(r-1/2\sigma^2\right)T}\right) - \frac{1}{2}\right|}{\frac{1}{\sqrt{2\pi}}} \sqrt{N} = 1.$$

Proof. The RB model is tilted in such a way that

$$P^{(N)}\left(S_N^{(N)} < s_0 e^{\left(r-1/2\sigma^2\right)T}\right) = P\left(X < \tfrac{N}{2}\right),$$

where X is a $\mathrm{Bin}(N, 1/2)$-RV on some probability space $(\Omega, \mathscr{A}, P)$. Further, it follows from symmetry that

$$\begin{aligned} P\left(X < \tfrac{N}{2}\right) &= \tfrac{1}{2}\left(1 - P\left(X = \tfrac{N}{2}\right)\right) \\ &= \begin{cases} \tfrac{1}{2}\left(1 - \binom{N}{N/2}\tfrac{1}{2^N}\right) & N \text{ even} \\ \tfrac{1}{2} & N \text{ odd} \end{cases} \end{aligned}$$

Finally, by Stirling's formula (compare e.g. [AS72], Formula 6.1.38), we have

$$\binom{N}{N/2}\frac{1}{2^N} = \frac{2}{\sqrt{2\pi}}\frac{1}{\sqrt{N}} + o\left(\frac{1}{\sqrt{N}}\right),$$

which yields the assertion. □

We see that the RB model leads to a terminal value $S_N^{(N)}$, for which the Berry-Esséen bound is tight in the sense that

$$\limsup_{N\to\infty}\left\{\sup_x \left| P^{(N)}\left(S_N^{(N)} < x\right) - \Phi\left(\frac{\ln\left(\frac{x}{s_0}\right) - (r - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}\right)\right| \sqrt{N}\right\} \geq$$
$$\limsup_{N\to\infty}\left\{\left| P^{(N)}\left(S_N^{(N)} < s_0 e^{(r-1/2\sigma^2)T}\right) - \frac{1}{2}\right| \sqrt{N}\right\} = \frac{1}{\sqrt{2\pi}} > 0.$$

In the general case, it is more involved to decide whether the Berry-Esséen inequality is tight. In the next paragraph, we investigate the asymptoptic behaviour of the discretisation error for a fairly general class of binomial processes. We will essentially follow Chang and Palmer (2007). However, while they restrict to risk-neutral transition probabilities, we take a more general approach, so that our results also apply to the models suggested by CRR and by RB. In particular, we will find that the Berry-Esséen inequality is tight for the models under consideration.

The Asymptotic Behaviour of the Discretisation Error

For simplicity, we limit the following analysis of the discretisation error to the terminal distribution of the stock price. For the Black-Scholes model, the distribution at maturity is given by

$$Q(S_T \geq x) = \Phi(d_2),$$

where

$$d_2 := d_2(x) := \frac{\ln(s_0/x)+(r-1/2\sigma^2)T}{\sigma\sqrt{T}}.$$

Let us assume that the drift in the discrete-time model is constant in N; i.e. $\alpha(N) \equiv \alpha$. As we see in the following, even in this case there is no asymptotic expansion of $P^{(N)}\left(S_N^{(N)} \geq x\right)$ around $Q(S_T \geq x) = \Phi(d_2)$ in the conventional sense. That is, a function $f(\varepsilon)$ has an asymptotic expansion in powers of ε up to order k with constant coefficients $(c_i)_{i=0,\ldots,k}$, if for any $m = 0,\ldots,k$,

$$\lim_{\varepsilon \to 0^+} \varepsilon^{-m}\left(f(\varepsilon) - \sum_{i=0}^{m} c_i \varepsilon^i\right) = 0.$$

However, as suggested by Diener and Diener [DD04], the asymptotic behaviour of the discretisation error can be described with an appropriate ”extended asymptotic calculus”. Let us explain: Apparently, the distribution of the binomial process at maturity $P^{(N)}\left(S_N^{(N)} \geq x\right)$ writes as

$$P^{(N)}\left(S_N^{(N)} \geq x\right) = \sum_{i=l(N)}^{N} \binom{N}{i} p_u(N)^i (1-p_u(N))^{N-i} =: F_x(N),$$

where $l(N)$ is the smallest integer l such that $s_0 u(N)^l d(N)^{N-l} \geq x$. If $\alpha(N)$ is constant in N, it follows from a Taylor expansion that $u(N)$ and $d(N)$ admit an asymptotic expansion in powers of $1/\sqrt{N}$ up to an arbitrary order k in the conventional sense. Consequently, we might be tempted to suppose that if $p_u(N)$ admits an asymptotic expansion in powers of $1/\sqrt{N}$ in the conventional sense (which is in particular valid for the models under consideration), the distribution $F_x(N)$ will do so, too. However, there are problems arising from $l(N)$: Let us introduce

$$a(N) := \frac{\ln(x/s_0) - N\ln d(N)}{\ln u(N) - \ln d(N)} = \frac{1}{2}N + \frac{\ln(x/s_0) - \alpha T}{2\beta\sqrt{T}}\sqrt{N}, \tag{2.39}$$

which is the solution to $s_0 u(N)^a d(N)^{N-a} = x$. Then,

$$l(N) = [a(N)] + 1 = a(N) + 1 - \{a(N)\} = a(N) + \{-a(N)\}, \tag{2.40}$$

where $\{.\}$ denotes the fractional part. Note that while $a(N)$ is a polynomial in $\sqrt{N}$ (compare (2.39)), this is not the case for the integer $l(N)$ because the latter involves the fractional part $\{-a(N)\}$. In particular, $\{-a(N)\}$ has no limit as N tends to infinity; but it is known to be bounded between 0 and 1. Therefore, Diener and Diener introduce the following extended asymptotic calculus (compare [DD04], Definition 2.1):

Definition 2. *Let $(f_i)_{i=0,\ldots,k}$ be* ***bounded functions*** *of $\varepsilon > 0$; we shall say that a function $f(\varepsilon)$ has an asymptotic expansion in powers of ε up to order k with coefficients $(f_i)_{i=0,\ldots,k}$ if for any $m = 0,\ldots,k$,*

$$\lim_{\varepsilon\to 0^+} \varepsilon^{-m}\left(f(\varepsilon) - \sum_{i=0}^{m} f_i(\varepsilon)\varepsilon^i\right) = 0$$

The term $f_i(\varepsilon)\varepsilon^i$ is called the term of order i of the expansion.

Remark 6. *Obviously, there is no uniqueness for the expansion with bounded coefficients of a given function. Moreover, it is clear that, if the sequence $(f_i)_{i=0,\ldots,k}$ is a sequence of constant functions in ε, the function f has an asymptotic expansion up to order k (in the conventional sense).*

We now formulate the key result of this section which describes the asymptotics of the discretisation error in the approximation to the terminal stock price S_T for a fairly general class of binomial processes. As motivated above, the extended asymptotic calculus introduced by Diener and Diener is suitable for this purpose.

Proposition 7. *Let $S^{(N)}$ be the process (2.2) with $\beta = \sigma$ and $\alpha(N) \equiv \alpha$ constant in N. Assume further that the one-step transition probability $p_u(N)$ admits an asymptotic expansion up to order $k = 3$ in powers of $1/\sqrt{N}$ in the conventional sense with constant $c_1 = 1/(2\sigma)(r-\alpha-1/2\sigma^2)$ and $c_2 = 0$, i.e.*

$$p_u(N) = \tfrac{1}{2} + \tfrac{1}{2\sigma}\left(r-\alpha-\tfrac{1}{2}\sigma^2\right)\left(\tfrac{T}{N}\right)^{1/2} + c_3\left(\tfrac{T}{N}\right)^{3/2} + o\left(\tfrac{1}{N^{3/2}}\right) \tag{2.41}$$

for some constant c_3. Then, $P^{(N)}\left(S_N^{(N)} \geq x\right)$ admits an ***asymptotic expansion with bounded coefficients*** *around $Q(S_T \geq x) = \Phi(d_2)$ up to order $k = 2$ in powers of $1/\sqrt{N}$. It can be written as*

$$\begin{aligned} P^{(N)}\left(S_N^{(N)} \geq x\right) = \Phi(d_2) + \tfrac{e^{-\frac{1}{2}d_2^2}}{\sqrt{2\pi}} b(N)\tfrac{1}{\sqrt{T}}\left(\tfrac{T}{N}\right)^{1/2} + \\ \tfrac{e^{-\frac{1}{2}d_2^2}}{\sqrt{2\pi}}\left(g + \left(\tfrac{r-\alpha-1/2\sigma^2}{3\sigma\sqrt{T}} - \tfrac{d_2}{12T}\right)(1-d_2^2) - \tfrac{d_2}{2}b^2(N)\tfrac{1}{T}\right)\tfrac{T}{N} + o\left(\tfrac{1}{N}\right), \end{aligned} \tag{2.42}$$

where

$$g := \tfrac{1}{2\sigma^2}(r-\alpha-\tfrac{1}{2}\sigma^2)^2 d_2 + 2c_3\sqrt{T}$$

with

$$b(N) := 1 - 2\{-a(N)\}. \tag{2.43}$$

Remark 7. *Proposition 7 is a generalisation of the main result in Chang and Palmer ([CP07], p. 93/94). They consider the case $\beta = \sigma$, as we do above. However, their analysis is limited to risk-neutral transition probabilities. Yet in fact, the result extends to arbitrary transition probabilities $p_u(N)$ of the form (2.41). As a consequence, the assumption of Proposition 7 covers the RB model and the CRR model, which are not considered by Chang and Palmer.*

In the following, we give a proof to Proposition 7. It is similar to that of Chang and Palmer in large parts. They suggest to apply the following extended version of results by J.V. Uspensky (compare [Usp37], Chapter 7) on the approximation of the normal distribution to the binomial distribution ([CP07], Lemma 1):

Lemma 1. *Provided that $p_u(N) \to 1/2$ as $N \to \infty$, and $0 \le l(N) \le N+1$ for N sufficiently large,*

$$\sum_{k=l(N)}^{N} \binom{N}{k} p_u^k(N)(1-p_u(N))^{N-k} = \frac{1}{\sqrt{2\pi}} \int_{\xi_1(N)}^{\xi_2(N)} e^{-\frac{1}{2}u^2} \mathrm{d}u +$$
$$\frac{1-2p_u(N)}{6\sqrt{2\pi N p_u(N)(1-p_u(N))}} \left((1-\xi_2^2(N)) e^{-\frac{1}{2}\xi_2^2(N)} - (1-\xi_1^2(N)) e^{-\frac{1}{2}\xi_1^2(N)} \right) +$$
$$\frac{1}{12N\sqrt{2\pi}} \left(\xi_2(N) e^{-\frac{1}{2}\xi_2^2(N)} (\xi_2^2(N)-1) - \xi_1(N) e^{-\frac{1}{2}\xi_1^2(N)} (\xi_1^2(N)-1) \right) + o\left(\frac{1}{N}\right),$$

where

$$\xi_1(N) = \frac{l(N) - Np_u(N) - 1/2}{\sqrt{Np_u(N)(1-p_u(N))}} \quad \text{and} \quad \xi_2(N) = \frac{N - Np_u(N) + 1/2}{\sqrt{Np_u(N)(1-p_u(N))}}.$$

Proof of Proposition 7. We assume that N is sufficiently large to ensure that $0 < p_u(N) < 1$ and $0 \le l(N) \le N+1$. Note from (2.39) that for $\beta = \sigma$

$$a(N) = \frac{1}{2}N + \frac{\ln(x/s_0) - \alpha T}{2\sigma\sqrt{T}} \sqrt{N},$$

and hence

$$\begin{aligned} -2a(N) + N + 2Nc_1 \left(\frac{T}{N}\right)^{1/2} &= -\frac{\ln(x/s_0) - \alpha T}{\sigma\sqrt{T}} \sqrt{N} + 2N \left(\frac{1}{2\sigma}(r - \alpha - \frac{1}{2}\sigma^2)\right) \left(\frac{T}{N}\right)^{1/2} \\ &= \sqrt{N} d_2. \end{aligned}$$

It follows from the asymptotic expansion (2.41) of $p_u(N)$ that

$$\begin{aligned} -2l(N) + 2Np_u(N) + 1 &= -2a(N) + 2Np_u(N) + b(N) \\ &= \sqrt{N} d_2 + b(N) + 2c_3 T \left(\frac{T}{N}\right)^{1/2} + o\left(\frac{1}{N^{1/2}}\right). \quad (2.44) \end{aligned}$$

Further, we have

$$p_u(N)(1-p_u(N)) = \tfrac{1}{4} - c_1^2 \tfrac{T}{N} + O\left(\tfrac{1}{N^{3/2}}\right). \tag{2.45}$$

Hence, it follows from the the Binomial Series Theorem that

$$\frac{1}{2\sqrt{p_u(N)(1-p_u(N))}} = \frac{1}{\sqrt{1-4c_1^2\frac{T}{N}+O(\frac{1}{N^{3/2}})}} = 1 + 2c_1^2\frac{T}{N} + O\left(\frac{1}{N^{3/2}}\right) \tag{2.46}$$

Combining the results (2.44) and (2.46), we get

$$\begin{aligned} -\xi_1(N) &= \frac{1}{2\sqrt{Np_u(N)(1-p_u(N))}}\left(-2l(N)+2Np_u(N)+1\right) = \\ &d_2 + b(N)\frac{1}{\sqrt{T}}\left(\tfrac{T}{N}\right)^{1/2} + 2\left(c_1^2 d_2 + c_3\sqrt{T}\right)\tfrac{T}{N} + o\left(\tfrac{1}{N}\right) = \\ &d_2 + b(N)\frac{1}{\sqrt{T}}\left(\tfrac{T}{N}\right)^{1/2} + g\tfrac{T}{N} + o\left(\tfrac{1}{N}\right). \end{aligned} \tag{2.47}$$

We now analyse the terms in Lemma 1 one by one. Here

$$\int_{\xi_1(N)}^{\xi_2(N)} e^{-\frac{1}{2}u^2}\mathrm{d}u = \int_{\xi_1(N)}^{\infty} e^{-\frac{1}{2}u^2}\mathrm{d}u - \int_{\xi_2(N)}^{\infty} e^{-\frac{1}{2}u^2}\mathrm{d}u := I_1(N) - I_2(N).$$

Note that

$$I_1(N) = \sqrt{2\pi}\Phi(d_2) + h(-\xi_1(N)), \tag{2.48}$$

where $h(x) = \int_{d_2}^{x} e^{-u^2/2}\mathrm{d}u$. Next we apply a third-order Taylor expansion of $h(-\xi_1(N))$ about the point d_2: Since h''' is bounded, the expansion (2.47) of $\xi_1(N)$ implies that

$$h(-\xi_1(N)) = e^{-\frac{1}{2}d_2^2} b(N)\frac{1}{\sqrt{T}}\left(\tfrac{T}{N}\right)^{1/2} + e^{-\frac{1}{2}d_2^2}\left(g - \tfrac{d_2}{2}b^2(N)\tfrac{1}{T}\right)\tfrac{T}{N} + o\left(\tfrac{1}{N}\right)$$

(a detailed Taylor expansion argument can be found in [CP07]). By (2.48), we then obtain that

$$I_1(N) = \sqrt{2\pi}\Phi(d_2) + e^{-\frac{1}{2}d_2^2} b(N)\frac{1}{\sqrt{T}}\left(\tfrac{T}{N}\right)^{1/2} + e^{-\frac{1}{2}d_2^2}\left(g - \tfrac{d_2}{2}b^2(N)\tfrac{1}{T}\right)\tfrac{T}{N} + o\left(\tfrac{1}{N}\right). \tag{2.49}$$

Note next that since $p_u(N) \to 1/2$ as $N \to \infty$, it follows that $\xi_2(N)/\sqrt{N} \to 1$. Consequently, the integral $I_2(N)$ does not contribute to the terms of order $1/N$; i.e.

$$I_2(N) = o\left(\tfrac{1}{N}\right) \tag{2.50}$$

(compare [CP07] for details). Regarding the second term in Lemma 1, note that by (2.45),

$$\begin{aligned}\frac{1-2p_u(N)}{\sqrt{Np_u(N)(1-p_u(N))}} &= \left(\frac{2}{\sqrt{N}}+O\left(\frac{1}{N^{3/2}}\right)\right)\left(-2c_1\left(\frac{T}{N}\right)^{1/2}+O\left(\frac{1}{N^{3/2}}\right)\right)\\ &= -4c_1\frac{1}{\sqrt{T}}\frac{T}{N}+O\left(\frac{1}{N^{3/2}}\right).\end{aligned}$$

Since $-\xi_1(N)\to d_2$ and $\xi_2(N)\to\infty$ as $N\to\infty$, we now obtain that

$$\frac{1-2p_u(N)}{6\sqrt{2\pi Np_u(N)(1-p_u(N))}}\left((1-\xi_2^2(N))e^{-\frac{1}{2}\xi_2^2(N)}-(1-\xi_1^2(N))e^{-\frac{1}{2}\xi_1^2(N)}\right)= \frac{2c_1}{3\sqrt{2\pi}\sqrt{T}}(1-d_2^2)e^{-\frac{1}{2}d_2^2}\frac{T}{N}+o\left(\frac{1}{N}\right), \quad (2.51)$$

Further, we observe that the third term in Lemma 1 admits the form

$$\frac{1}{12N\sqrt{2\pi}}\left(\xi_2(N)e^{-\frac{1}{2}\xi_2^2(N)}(\xi_2^2(N)-1)-\xi_1(N)e^{-\frac{1}{2}\xi_1^2(N)}(\xi_1^2(N)-1)\right)= \frac{d_2e^{-1/2d_2^2}(d_2^2-1)}{12\sqrt{2\pi}T}\frac{T}{N}+o\left(\frac{1}{N}\right). \quad (2.52)$$

Finally, combining the results (2.49) to (2.52) yields the assertion. □

Next we wish to interpret the previous result on the asymptotic behaviour of the discretisation error. Apparently, the factor $b(N)=1-2\{-a(N)\}$ enters both the coefficient of the term of order $1/\sqrt{N}$ and that of the term of order $1/N$. Since $b(N)$ is non-constant in N, the discretisation error converges non-smoothly, although $u(N)$, $d(N)$ and $p_u(N)$ admit an asymptotic expansion in the conventional sense.

However, as $b(N)$ is bounded by 1 and -1, we obtain the following bounds on the oscillations of the leading error term:

Corollary 5. *The leading term of the discretisation error is bounded by*

$$-\frac{e^{-\frac{1}{2}d_2^2(x)}}{\sqrt{2\pi}}\frac{1}{\sqrt{T}}\leq\frac{e^{-\frac{1}{2}d_2^2(x)}}{\sqrt{2\pi}}b(N)\frac{1}{\sqrt{T}}\leq-\frac{e^{-\frac{1}{2}d_2^2(x)}}{\sqrt{2\pi}}\frac{1}{\sqrt{T}}.$$

Clearly, the coefficient of the leading error term can also be bounded uniformly in $x\in\mathbb{R}$:

Corollary 6. *We have*

$$0\leq\left|\frac{e^{-\frac{1}{2}d_2^2(x)}}{\sqrt{2\pi}}b(N)\frac{1}{\sqrt{T}}\right|\leq\left|\frac{1}{\sqrt{2\pi}\sqrt{T}}\right| \qquad \textit{for all } N\in\mathbb{N} \textit{ and for all } x\in\mathbb{R}.$$

Note that $1/\sqrt{2\pi}$ is the maximum value of $1/(\sqrt{2\pi})e^{-1/2d_2^2(x)}$, which is attained at $x = s_0 e^{(r-1/2\sigma^2)T}$, the median of the continuous-time model. This matches the intuitive idea that by error accumulation, the discretisation error in the approximation to the distribution function should be largest at the median.

Let us now investigate the discretisation error for the binomial models under consideration. First, we consider binomial schemes with risk-neutral transition probabilities. Then by (2.15) the assumption $\beta = \sigma$ implies that the corresponding transition probability $q_u(N)$ is of the form (2.41) with

$$c_3 = \frac{1}{2\sigma}\left(\frac{1}{2}(\alpha-r)^2 + \frac{1}{6}\sigma^2(\alpha-r) + \frac{1}{24}\sigma^4\right).$$

Consequently, the asymptotics of the discretisation error can be determined from Proposition 7:

Corollary 7. *Let $S^{(N)}$ be the binomial process (2.2) with $\beta = \sigma$ and $\alpha(N) \equiv \alpha$ constant in N. If the transition probabilities are defined according to the risk-neutral measure, we have*

$$Q^{(N)}\left(S_N^{(N)} \geq x\right) = \Phi(d_2) + \frac{e^{-\frac{1}{2}d_2^2}}{\sqrt{2\pi}\sqrt{T}} b(N)\left(\frac{T}{N}\right)^{1/2} +$$
$$\frac{e^{-\frac{1}{2}d_2^2}}{\sqrt{2\pi}}\left(\frac{1}{2\sigma^2}d_1(r-\alpha)^2 + \frac{2-d_1d_2-d_1^2}{6\sigma\sqrt{T}}(r-\alpha) + \frac{d_1^3+d_1d_2^2+2d_2-4d_1}{24T} - \frac{d_2}{2T}b^2(N)\right)\frac{T}{N} + o\left(\frac{1}{N}\right),$$

where $d_1 := d_1(x) := d_2(x) + \sigma\sqrt{T}$ (compare [CP07]).

Let us recall that according to Proposition 2, for risk-neutral schemes, the assumption $\beta = \sigma$ is necessary and sufficient to ensure that the moments are asymptotically matched. Consequently, the above result is valid for the schemes that are of relevance for numerical option pricing.

Next we analyse the discretisation error for the models suggested by RB and by CRR. Clearly, both models satisfy the assumption of Proposition 7. In particular, the transition probabilities are of the form (2.41). Hence, we can formulate the following results:

Corollary 8. *Let $S^{(N)}$ be the binomial process suggested by RB. Then,*

$$P^{(N)}\left(S_N^{(N)} \geq x\right) = \Phi(d_2) + \frac{e^{-\frac{1}{2}d_2^2}}{\sqrt{2\pi}\sqrt{T}} b(N)\left(\frac{T}{N}\right)^{1/2} + \frac{e^{-\frac{1}{2}d_2^2}}{\sqrt{2\pi}}\left(\frac{d_2^3-d_2}{12T} - \frac{d_2}{2T}b^2(N)\right)\frac{T}{N} + o\left(\frac{1}{N}\right),$$

where in this case

$$b(N) = 1 - 2\{-\tfrac{1}{2}N + \tfrac{1}{2}d_2\sqrt{N}\}.$$

Corollary 9. *Let $S^{(N)}$ be the binomial process suggested by CRR. Then,*

$$P^{(N)}\left(S_N^{(N)} \geq x\right) = \Phi(d_2) + \frac{e^{-\frac{1}{2}d_2^2}}{\sqrt{2\pi}\sqrt{T}}\, b(N)\left(\tfrac{T}{N}\right)^{1/2} + \frac{e^{-\frac{1}{2}d_2^2}}{\sqrt{2\pi}}\left(\tfrac{1}{2\sigma^2}d_2(r-\tfrac{1}{2}\sigma^2)^2 + \tfrac{1-d_2^2}{3\sigma\sqrt{T}}(r-\tfrac{1}{2}\sigma^2) + \tfrac{d_2^3-d_2}{12T} - \tfrac{d_2}{2T}b^2(N)\right)\tfrac{T}{N} + o\left(\tfrac{1}{N}\right),$$

where in this case

$$b(N) = 1 - 2\{-\tfrac{1}{2}N + \tfrac{\ln(s_0/x)}{2\sigma\sqrt{T}}\sqrt{N}\}.$$

As we anticipate from the results observed for risk-neutral transition probabilities, there is a relationship between moment matching and the fact that the transition probability $p_u(N)$ is of the form (2.41). In fact, we have the following:

Proposition 8. *Let $S^{(N)}$ be the binomial process (2.2) with $\beta = \sigma$ and $\alpha(N) \equiv \alpha$ constant in N.*

1. *If $p_u(N)$ is of the form (2.41), the corresponding first two moments of the one-period logreturns converge in order*

$$\left|\mu_{P^{(N)}}(N) - (r - \tfrac{1}{2}\sigma^2)\right| = O\left(\tfrac{1}{N}\right) \qquad \text{and} \qquad \left|\frac{\sigma^2}{\sigma^2_{P^{(N)}}(N)} - 1\right| = O\left(\tfrac{1}{N}\right).$$

2. *Assume that the moment matching condition (2.5) on the expectation of the logreturns is satisfied with*

$$\left|\mu_{P^{(N)}}(N) - (r - 1/2\sigma^2)\right| = O\left(\tfrac{1}{N}\right),$$

   then $p_u(N)$ is of the form (2.41).

Proof. The first part of the assertion follows directly by computing the corresponding characteristics of the discrete-time model $\mu_{P^{(N)}}(N)$ and $\sigma^2_{P^{(N)}}(N)$. For the second part of the assertion note from (2.8) that if $|\mu_{P^{(N)}}(N) - (r - 1/2\sigma^2)| = O(\frac{1}{N})$, we have

$$\left|\alpha + \sigma\sqrt{\tfrac{N}{T}}\,(2p_u(N) - 1) - (r - \tfrac{1}{2}\sigma^2)\right| = O\left(\tfrac{1}{N}\right).$$

It then follows by re-arranging terms that

$$\left|p_u(N) - \left(\tfrac{1}{2} + \tfrac{1}{2\sigma}\left(r - \alpha - \tfrac{1}{2}\sigma^2\right)\sqrt{\tfrac{T}{N}}\right)\right| = O\left(\tfrac{1}{N^{3/2}}\right),$$

which proves the assertion. □

We see from the result above that if the assumption of Proposition 7 is satisfied, the Berry-Esséen inequality is applicable (compare Proposition 5), which yields

$$\sup_x \left| P^{(N)}\left(S_N^{(N)} \geq x\right) - Q\left(S(T) \geq x\right)\right| = O\left(\tfrac{1}{\sqrt{N}}\right);$$

i.e. the discretisation error converges to zero in order $1/\sqrt{N}$. Due to Proposition 7, the Berry-Esséen bound is now known to be tight: Clearly, for all $x \in \mathbb{R}$ and all $\alpha, \sigma \in \mathbb{R}$, there is some subsequence $(N_k)_{k \in \mathbb{N}} \subseteq \mathbb{N}$ for which $b(N_k) \neq 0$. Hence, according to Proposition 7, we have

$$\limsup_{N\to\infty} \left| P^{(N)}\left(S_N^{(N)} \geq x\right) - \Phi(d_2(x))\right| \sqrt{N} > 0, \qquad \text{for all } x \in \mathbb{R},$$

which implies the following result:

Corollary 10. *With the assumption of Proposition 7, we have*

$$\limsup_{N\to\infty} \left\{ \sup_x \left| P^{(N)}\left(S_N^{(N)} \geq x\right) - \Phi(d_2(x))\right| \sqrt{N} \right\} > 0.$$

We wish to stress that Proposition 7 readily extends to the case where $\alpha(N)$ is non-constant, but bounded in N as required by Assumption 1. This will allow to investigate the order of convergence for advanced binomial schemes.

Proposition 9. *Let $S^{(N)}$ be the process (2.2) with $\beta = \sigma$. Assume further that the transition probability $p_u(N)$ admits an asymptotic expansion with bounded coefficients up to order $k = 3$ in powers of $1/\sqrt{N}$ with $c_1(N) = 1/(2\sigma)(r - \alpha(N) - 1/2\sigma^2)$ and $c_2(N) \equiv 0$, i.e.*

$$p_u(N) = \tfrac{1}{2} + \tfrac{1}{2\sigma}\left(r - \alpha(N) - \tfrac{1}{2}\sigma^2\right)\left(\tfrac{T}{N}\right)^{1/2} + c_3(N)\left(\tfrac{T}{N}\right)^{3/2} + o\left(\tfrac{1}{N^{3/2}}\right) \tag{2.53}$$

for some bounded sequence $(c_3(N))_N$. Then

$$P^{(N)}\left(S_N^{(N)} \geq x\right) = \Phi(d_2) + \frac{e^{-\frac{1}{2}d_2^2}}{\sqrt{2\pi}\sqrt{T}} b(N)\left(\tfrac{T}{N}\right)^{1/2} + \frac{e^{-\frac{1}{2}d_2^2}}{\sqrt{2\pi}}\left(g(N) + \left(\frac{r-\alpha(N)-1/2\sigma^2}{3\sigma\sqrt{T}} - \frac{d_2}{12T}\right)(1-d_2^2) - \frac{d_2}{2}b^2(N)\frac{1}{T}\right)\frac{T}{N} + o\left(\tfrac{1}{N}\right),$$

where

$$g(N) := \frac{1}{2\sigma^2}(r - \alpha(N) - \tfrac{1}{2}\sigma^2)^2 d_2 + 2c_3(N)\sqrt{T}.$$

For the special case that the transition probabilities are given by the risk-neutral measure (compare also [CP07]), we have

$$Q^{(N)}\left(S_N^{(N)} \geq x\right) = \Phi(d_2) + \frac{e^{-\frac{1}{2}d_2^2}}{\sqrt{2\pi}\sqrt{T}} b(N) \left(\tfrac{T}{N}\right)^{1/2} + \frac{e^{-\frac{1}{2}d_2^2}}{\sqrt{2\pi}} \left(\tfrac{1}{2\sigma^2} d_1 (r - \alpha(N))^2\right) \tfrac{T}{N} +$$
$$\frac{e^{-\frac{1}{2}d_2^2}}{\sqrt{2\pi}} \left(\tfrac{2 - d_1 d_2 - d_1^2}{6\sigma\sqrt{T}}(r - \alpha(N)) + \tfrac{d_1^3 + d_1 d_2^2 + 2d_2 - 4d_1}{24T} - \tfrac{d_2}{2T} b^2(N)\right) \tfrac{T}{N} + o\left(\tfrac{1}{N}\right).$$

In the next section, we investigate advanced models for which the asymptotic behaviour of the discretisation error is improved. In particular, we present the optimal drift model which can lead to convergence of order $o(1/N)$; i.e. in this case, the discretisation error converges *strictly faster than in order* $1/N$. To the best of our knowledge, the optimal drift model is new.

2.2.3 Advanced Binomial Models

As discussed above, for conventional tree methods (i.e. $\alpha(N) \equiv \alpha$ constant), the error in the approximation to the distribution of the terminal stock price converges non-smoothly; i.e. $P^{(N)}(S_N^{(N)} \geq x)$ in general only admits an asymptotic expansion with bounded, but non-constant coefficients. The oscillations of the coefficients are described by the quantity $b(N)$ defined in (2.43). This suggests that the asymptotic behaviour of the discretisation error can be improved by controlling $b(N)$. Firstly, controlling $b(N)$ can lead to smooth convergence of the leading error term. Secondly, it can even help to improve the order of convergence.

There is a vast number of articles on the control of the leading error term, amongst which are Leisen and Reimer (1995), Leisen (1998), Tian (1999) and Chang and Palmer (2007). Leisen and Reimer use an odd number of steps with the tree centred around the strike [LR96]. Leisen uses an even number of steps with the central node placed exactly at the strike [Lei98]. These methods can require moving the centre of the tree a long distance. By contrast, for the model suggested by Tian and by Chang and Palmer, the nodes in the tree are moved only a small distance so that the strike falls onto a neighbouring node or onto the geometric average of the two neighbouring nodes, respectively. A different approach to improve the convergence behaviour of the discretisation error can be found in Rogers and Stapleton (1998) [RS98].

As shown by Chang and Palmer, the discretisation error in the approximation suggested by Tian converges smoothly[2] in order $1/\sqrt{N}$, while their approach even achieves convergence of order $1/N$. We will explain the basic idea behind these models. Afterwards, we introduce the optimal drift model that further improves the asymptotic behaviour of the discretisation error.

Let us first follow Chang and Palmer and interpret the quantity $b(N)$ we wish to control in order to "get binomial schemes into shape". Assume that l is the integer value for which

$$s_N^{(N)}(l-1) := s_0 u^{l-1}(N) d^{N-l+1}(N) < x \le s_0 u^l(N) d^{N-l}(N) =: s_N^{(N)}(l),$$

then it follows from (2.39) that

$$\frac{\ln(s_N^{(N)}(l)/x)}{\ln(s_N^{(N)}(l)/s_N^{(N)}(l-1))} = -a(N) + l, \tag{2.54}$$

which implies by (2.40) that

$$\{-a(N)\} = \frac{\ln(s_N^{(N)}(l)/x)}{\ln(s_N^{(N)}(l)/s_N^{(N)}(l-1))}.$$

Consequently, the quantity $\{-a(N)\}$ admits the following interpretation: It measures *the position of x on the log-scale in relation to the two adjacent terminal values of the binomial process* (compare [CP07]). In particular, $\{-a(N)\}$ is strictly decreasing on $(s_N^{(N)}(l-1), s_N^{(N)}(l)]$ with

$$\{-a(N)\} = \begin{cases} 0 & \text{for } x = s_N^{(N)}(l) \\ \frac{1}{2} & \text{for } x = \sqrt{s_N^{(N)}(l) s_N^{(N)}(l-1)}, \end{cases} \tag{2.55}$$

and $\{-a(N)\}$ converges to 1 as x tends to $s_N^{(N)}(l-1)$. Consequently, $b(N) = 1 - 2\{-a(N)\}$ is strictly increasing on $(s_N^{(N)}(l-1), s_N^{(N)}(l)]$ with

$$b(N) = \begin{cases} 1 & \text{for } x = s_N^{(N)}(l) \\ 0 & \text{for } x = \sqrt{s_N^{(N)}(l) s_N^{(N)}(l-1)}, \end{cases} \tag{2.56}$$

and $b(N)$ converges to -1 as x tends to $s_N^{(N)}(l-1)$.

[2]In the original article by Tian, the author illustrates smooth convergence with numerical examples, while a mathematical proof is given by Chang and Palmer.

According to Proposition 2, if we choose risk-neutral transition probabilities, the asymptotic moment matching conditions are automatically satisfied for $\beta = \sigma$. By contrast, the sequence of drift parameters $(\alpha(N))_N$ can be chosen freely. In principle, the advanced models by Tian and by Chang and Palmer, and also the optimal drift model we suggest, exploit the flexibility of the drift parameter to modify the allocation of probability mass, so that $b(N)$ is controlled. Note that the preferred drift depends on the number of periods N. Hence, practical relevance of these advanced models relies on the fact that weak convergence to the stock price can also be ensured for the case that $(\alpha(N))_N$ is non-constant in N.

The Tian Model

Let $x \in \mathbb{R}$ be arbitrary. For binomial option valuation, the point x will be the strike value.

The basic idea behind the Tian model is that for any number of periods N, the terminal distribution of the corresponding binomial model admits a realisation that is placed exactly at the point x. To be precise: We start with the binomial process $S_\alpha^{(N)}$ of the form (2.2) with $\beta = \sigma$ and $\alpha(N) \equiv \alpha$ constant in N. For each $N \in \mathbb{N}$, there is some integer $l_\alpha(N)$ for which $x \in (s_N^{(N)}(l_\alpha(N)-1), s_N^{(N)}(l_\alpha(N))]$. In general, $x \neq s_N^{(N)}(l_\alpha(N))$. Then the corresponding equation $s_0 u(N)^a d(N)^{N-a} = x$ is solved by some value $a_\alpha(N)$ with $l_\alpha(N) - 1 < a_\alpha(N) < l_\alpha(N)$ (compare (2.39)). Given the sequence $(l_\alpha(N))_N$, we define a sequence $(\tilde{\alpha}(N))_N$ with

$$\tilde{\alpha}(N) := \frac{\ln(x/s_0) - (2l_\alpha(N) - N)\sigma\sqrt{T/N}}{T}. \tag{2.57}$$

This will become the sequence of drift parameters for the Tian model[3]; i.e. for any number of periods N, the corresponding Tian model $S_{\tilde{\alpha}(N)}^{(N)}$ is defined as the process (2.2) with $\beta = \sigma$ and drift $\tilde{\alpha}(N)$. Note in particular that the new sequence of drift parameters is non-constant in N. The process $S_{\tilde{\alpha}(N)}^{(N)}$ is defined such that the corresponding equation $s_0 u(N)^a d(N)^{N-a} = x$ is solved by

$$a_{\tilde{\alpha}(N)}(N) := \tfrac{1}{2}N + \frac{\ln(x/s_0) - \tilde{\alpha}(N)T}{2\sigma\sqrt{T}}\sqrt{N} = l_\alpha(N),$$

where the last equality follows from definition of $\tilde{\alpha}(N)$ (compare (2.57)). Consequently, in contrast to the quantity $a_\alpha(N)$ obtained for the original model, the corresponding quantity $a_{\tilde{\alpha}(N)}(N)$ obtained for the superimposed Tian model is integer-valued. Hence,

[3] In the original article by Tian, the author chooses either $l_\alpha(N)$ or $l_\alpha(N) - 1$ depending on which is closer to $a_\alpha(N)$ [Tia99].

as we can observe from (2.54),

$$S^{(N)}_{\tilde{\alpha}(N)}(l_\alpha(N)) = x \qquad \text{for all } N. \tag{2.58}$$

In words: *For any number of periods N, the terminal distribution of the corresponding Tian model allocates probability mass to the point x.* As a consequence, the corresponding quantity $b(N)$ does not depend on N; rather, $b(N) = 1$ for all N.

It remains to show that the new sequence of drift parameters $(a_\alpha(N))_N$ is bounded in N. Then the moment matching conditions will be satisfied for the risk-neutral transition probabilities, so that the Tian model will ensure weak convergence to the stock price process by Proposition 4. In essence, the assumption on boundedness is valid due to the fact that the mass points are only moved a small distance compared to the original model: This can be observed by writing the new drift $\tilde{\alpha}(N)$ in terms of the original drift α. By (2.39), we get

$$\tilde{\alpha}(N) = \tfrac{2\sigma}{\sqrt{T}\sqrt{N}}\,(a_\alpha(N) - l_\alpha(N)) + \alpha, \tag{2.59}$$

which implies that

$$\tfrac{-2\sigma}{\sqrt{T}\sqrt{N}} + \alpha \leq \tilde{\alpha}(N) < \alpha.$$

We observe that the new drift $\tilde{\alpha}(N)$ in the Tian model differs from the original drift by

$$\tilde{\alpha}(N) = \alpha + o(1). \tag{2.60}$$

In particular, the new drift satisfies Assumption 1, i.e. $\tilde{\alpha}(N) = O(1)$, so that we can formulate the following result:

Proposition 10. *The sequence of processes $\left(S^{(c,N)}_{\tilde{\alpha}(N)}\right)_N$ defined from the Tian model by linear interpolation and an appropriate time-scaling (compare (2.21)) converges weakly to the stock price process S.*

Compared to conventional binomial methods with constant drift α, the Tian model shows an improved convergence behaviour of the discretisation error in the approximation of the terminal stock price, which is due to the fact the corresponding quantity $b(N)$ is constant in N. By Proposition 9, we obtain the following result on the asymptotic behaviour of the discretisation error in the Tian model:

Proposition 11. *[The Tian Model] Let $x \in \mathbb{R}$. Let $S^{(N)}_\alpha$ be the binomial process (2.2) with $\beta = \sigma$ and $\alpha(N) \equiv \alpha$ constant in N. If we superimpose the Tian model $S^{(N)}_{\tilde{\alpha}(N)}$ (associated with the given point x), the new transition states and the associated risk-neutral measure* *(denoted by $Q^{(N)}_{\tilde{\alpha}(N)}$) are such that $Q^{(N)}_{\tilde{\alpha}(N)}(S^{(N)}_{\tilde{\alpha}(N)}(N) \geq x)$ admits an asymptotic expansion (in the conventional sense) in powers of $1/\sqrt{N}$ up to order $k = 2$. We have*

$$Q^{(N)}_{\tilde{\alpha}(N)}\left(S^{(N)}_{\tilde{\alpha}(N)}(N) \geq x\right) = \Phi(d_2) + \frac{e^{-\frac{1}{2}d_2^2}}{\sqrt{2\pi}\sqrt{T}}\left(\frac{T}{N}\right)^{1/2} +$$
$$\frac{e^{-\frac{1}{2}d_2^2}}{\sqrt{2\pi}}\left(\frac{1}{2\sigma^2}d_1(r-\alpha)^2 + \frac{2-d_1d_2-d_1^2}{6\sigma\sqrt{T}}(r-\alpha) + \frac{d_1^3+d_1d_2^2+2d_2-4d_1}{24T} - \frac{d_2}{2T}\right)\frac{T}{N} + o\left(\frac{1}{N}\right).$$

Proof. According to Proposition 9, we have

$$Q^{(N)}_{\tilde{\alpha}(N)}\left(S^{(N)}_{\tilde{\alpha}(N)}(N) \geq x\right) = \Phi(d_2) + \frac{e^{-\frac{1}{2}d_2^2}}{\sqrt{2\pi}\sqrt{T}}\left(\frac{T}{N}\right)^{1/2} +$$
$$\frac{e^{-\frac{1}{2}d_2^2}}{\sqrt{2\pi}}\left(\frac{1}{2\sigma^2}d_1(r-\tilde{\alpha}(N))^2 + \frac{2-d_1d_2-d_1^2}{6\sigma\sqrt{T}}(r-\tilde{\alpha}(N)) + \frac{d_1^3+d_1d_2^2+2d_2-4d_1}{24T} - \frac{d_2}{2T}\right)\frac{T}{N} + o\left(\frac{1}{N}\right),$$

which yields the assertion by (2.60). □

Hence for the Tian model, the discretisation error in the approximation to the distribution of the terminal stock price converges smoothly in order $1/\sqrt{N}$, where here and in the rest of Chapter 2, the term ”smooth” is used if the coefficient of the leading error term is constant and if oscillations of higher order terms are negligible. Thus compared to conventional methods, the discretisation error converges smoothly, but the order of convergence is not improved. The Berry-Esséen bound remains tight in the sense that

$$\limsup_{N\to\infty}\left\{\sup_{x\in\mathbb{R}}\left|Q^{(N)}_{\tilde{\alpha}(N)}\left(S^{(N)}_{\tilde{\alpha}(N)}(N) \geq x\right) - \Phi(d_2(x))\right|\sqrt{N}\right\} > 0.$$

Remark 8. *In this thesis, the definition of the Tian model (and also of the Chang-Palmer model discussed below) is more general than in the original papers by Tian and by Chang and Palmer. These authors limit their analysis to advanced models that are superimposed on the CRR model; i.e. they consider only the case $\alpha = 0$. By contrast, we allow the drift of the original model to be an arbitray constant. As we see below, this increases flexibility to further improve the convergence rate of the discretisation error.*

Next we investigate the model suggested by Chang and Palmer (for short: CP model). In contrast to the Tian model, the discretisation error in the CP model admits a higher order of convergence.

The Chang-Palmer Model

Let $x \in \mathbb{R}$. As observed by Chang and Palmer, the analysis of the quantity $b(N)$ suggests that the original drift α should be replaced by some sequence of drift parameters $(\underline{\alpha}(N))_N$ for which x coincides with the geometric average of $s_N^{(N)}(l(N))$ and $s_N^{(N)}(l(N)-1)$ (compare (2.56)). As a consequence, the discretisation error will exhibit a higher order of convergence. This is the basic idea behind the CP model ("the centered binomial model"). In more detail: Let $S_\alpha^{(N)}$ be the binomial process (2.2) with $\beta = \sigma$ and $\alpha(N) \equiv \alpha$ constant in N. As before, $l_\alpha(N)$ denotes the corresponding integers for which $x \in (s_N^{(N)}(l_\alpha(N)-1), s_N^{(N)}(l_\alpha(N))]$. Then we determine the sequence of new drift parameters $(\underline{\alpha}(N))_N$ by

$$\underline{\alpha}(N) = \frac{\ln(x/s_0) - (2l_\alpha(N) - N - 1)\sigma\sqrt{T/N}}{T}. \tag{2.61}$$

For any number of periods N, the superimposed CP model $S_{\underline{\alpha}(N)}^{(N)}$ is defined as the process (2.2) with $\beta = \sigma$ and drift $\underline{\alpha}(N)$. By (2.61), the CP model is defined such that the equation $s_0 u(N)^a d(N)^{N-a} = x$ is solved by

$$a_{\underline{\alpha}(N)}(N) := \tfrac{1}{2}N + \frac{\ln(x/s_0) - \underline{\alpha}(N)T}{2\sigma\sqrt{T}}\sqrt{N} = l_\alpha(N) - \tfrac{1}{2}.$$

Hence, it follows from (2.54) that

$$\left(S_{\tilde{\alpha}(N)}^{(N)}(l_\alpha(N)) S_{\tilde{\alpha}(N)}^{(N)}(l_\alpha(N)-1)\right)^{1/2} = x \qquad \text{for all } N. \tag{2.62}$$

In words: *For any number of periods N, the terminal distribution of the corresponding CP model is such that the point x is at the geometric average of two neighbouring mass points.* As a result, the corresponding quantity $b(N)$ is equal to zero for all N, which will improve the order of convergence of the discretisation error in the approximation to $\Phi(d_2(x))$.

Next we show that $\underline{\alpha}(N) = \alpha + o(1)$. This is again a direct consequence of the fact that the probability mass is only moved a small distance. Similar to the above results, it can be seen that

$$\underline{\alpha}(N) = \frac{2\sigma}{\sqrt{T}\sqrt{N}}\left(a_\alpha(N) - l_\alpha(N) + \tfrac{1}{2}\right) + \alpha, \tag{2.63}$$

where $a_\alpha(N)$ is again the solution to the equation $s_0 u(N)^a d(N)^{N-a} = x$ in the original model. We then get from (2.63) that

$$\alpha - \frac{\sigma}{\sqrt{T}\sqrt{N}} \leq \underline{\alpha}(N) < \alpha + \frac{\sigma}{\sqrt{T}\sqrt{N}}.$$

Consequently, we obtain the following results. Firstly, we know that the CP model ensures weak convergences to stock price process:

Proposition 12. *The sequence of processes* $\left(S^{(c,N)}_{\underline{\alpha}(N)}\right)_N$ *defined from the CP model by linear interpolation and an appropriate time-scaling converges weakly to the stock price process S.*

Secondly, according to Proposition 9, the asymptotic behaviour of the discretisation error is not only superior to that of conventional methods, but also to that of the Tian model:

Proposition 13. *[The Chang-Palmer Model ("The Centered Binomial Model")] Let* $x \in \mathbb{R}$. *Let* $S^{(N)}_{\alpha}$ *be the binomial process (2.2) with* $\beta = \sigma$ *and* $\alpha(N) \equiv \alpha$ *constant in* N. *If we superimpose the CP model* $S^{(N)}_{\underline{\alpha}(N)}$ *(associated with the given point x), the new transition states and the associated risk-neutral measure (denoted by* $Q^{(N)}_{\underline{\alpha}(N)}$*) are such that* $Q^{(N)}_{\underline{\alpha}(N)}\left(S^{(N)}_{\underline{\alpha}(N)}(N) \geq x\right)$ *admits an asymptotic expansion (in the conventional sense) in powers of* $1/N$ *up to order* $k = 1$. *We have*

$$Q^{(N)}_{\underline{\alpha}(N)}\left(S^{(N)}_{\underline{\alpha}(N)}(N) \geq x\right) = \Phi(d_2) + \frac{e^{-\frac{1}{2}d_2^2}}{\sqrt{2\pi}}\left(\frac{1}{2\sigma^2}d_1(r-\alpha)^2 + \frac{2-d_1d_2-d_1^2}{6\sigma\sqrt{T}}(r-\alpha) + \frac{d_1^3+d_1d_2^2+2d_2-4d_1}{24T}\right)\frac{T}{N} + o\left(\frac{1}{N}\right).$$

According to Proposition 13, the CP model leads to a discretisation error with a higher order of convergence: Compared to the methods considered before, the rate of convergence is improved from $1/\sqrt{N}$ to $1/N$. We have

$$\limsup_{N\to\infty}\left\{\sup_{x\in\mathbb{R}}\left|Q^{(N)}_{\underline{\alpha}(N)}\left(S^{(N)}_{\underline{\alpha}(N)}(N) \geq x\right) - \Phi(d_2(x))\right|\sqrt{N}\right\} = 0$$

and

$$\limsup_{N\to\infty}\left\{\sup_{x\in\mathbb{R}}\left|Q^{(N)}_{\underline{\alpha}(N)}\left(S^{(N)}_{\underline{\alpha}(N)}(N) \geq x\right) - \Phi(d_2(x))\right|N\right\} > 0,$$

which shows in particular, that *if the binomial process is defined according to the CP model, the Berry-Esséen bound ceases to be tight*. In addition, the leading term of the discretisation error converges monotonically; yet it is not clear whether the convergence behaviour of the discretisation error is affected by oscillations of higher order. We give a numerical example for binomial valuation of cash-or-nothing options below. For the example, the oscillations of higher order are not negligible.

Remark 9. *As for the Tian model, the definition of the Chang-Palmer model given above is more general than in the original paper. While the authors fix $\alpha = 0$, we allow the drift of the embedded binomial model to be an arbitrary constant. Though straightforward, the generalisation we suggest is the key result to introduce the optimal drift model. Here we optimise the drift of the original process α to further improve the rate of convergence of the discretisation error.*

The Optimal Drift Model

Before we introduce the optimal drift model, let us stress that the generalisation of the CP model we suggested above has the following impact on the asymptotic behaviour of the discretisation error: In the original paper by Chang and Palmer, the coefficient of the leading error term is constant; i.e. it only depends on the input parameters and on the given point $x \in \mathbb{R}$. By contrast, for the variant of the CP model we introduced, the coefficient of the leading error term is a quadratic function of the drift of the embedded binomial process $S_\alpha^{(N)}$; that is,

$$f_2(\alpha) = \frac{e^{-\frac{1}{2}d_2^2}}{\sqrt{2\pi}} \left(\frac{1}{2\sigma^2} d_1 (r-\alpha)^2 + \frac{2 - d_1 d_2 - d_1^2}{6\sigma\sqrt{T}} (r-\alpha) + \frac{d_1^3 + d_1 d_2^2 + 2d_2 - 4d_1}{24T} \right).$$

This suggests that we can optimise the free parameter α so that $f_2(\alpha)$ is minimal in absolute values. In particular, if $f_2(\alpha)$ intercepts the α-axis, i.e.

$$D := -d_1^4 + \sigma\sqrt{T} d_1^3 - d_1^2 \left(1 + \sigma^2 T\right) + 5\sigma\sqrt{T} d_1 + 2 \geq 0,$$

we choose α_0 such that

$$f_2(\alpha_0) = 0. \tag{2.64}$$

In this case, the leading term in the discretisation error cancels out, so that the rate of convergence of the error is further improved compared to the CP model. The discretisation error exhibits the rate $o(1/N)$; i.e. we have

$$\limsup_{N\to\infty} \left| Q_{\underline{\alpha}(N)}^{(N)} \left(S_{\underline{\alpha}(N)}^{(N)}(N) \geq x \right) - \Phi(d_2(x)) \right| N = 0, \qquad \text{for } x \in \mathbb{R} \text{ with } D(x) \geq 0.$$

If $f_2(\alpha)$ does not intercept the α-axis (i.e. $D < 0$), we choose the parameter α_0 for which $f_2(\alpha_0)$ is the vertex of the parabola, i.e. the coefficient of the leading error term is set to

$$f_2(\alpha_0) = \frac{e^{-\frac{1}{2}d_2^2}}{\sqrt{2\pi}} \frac{1}{36 T d_1} \left(d_1^4 + d_1^2 d_2^2 + 5 d_1 d_2 - 4 d_1^2 - 2 - d_1^3 d_2 \right).$$

In this case, the discretisation error continues to converge in order $1/N$, but the coefficient of its leading error term is always smaller than that obtained by the CP model. In essence, for the original paper by Chang and Palmer, the choice $\alpha = 0$ results in some uncontrolled value on the parabola $f_2(\alpha)$.

Proposition 14 (The Optimal Drift Model)**.** *Let $x \in \mathbb{R}$. Assume that α_0 is such that $f_2(\alpha_0)$ is minimal in absolute values and let $S^{(N)}_{\alpha_0}$ be the binomial process (2.2) with $\beta = \sigma$ and $\alpha(N) \equiv \alpha_0$. The OD model $S^{(N)}_{\underline{\alpha}_0(N)}$ is an alternative to the original CP model that is superimposed on $S^{(N)}_{\alpha_0}$. Consequently, if $f_2(\alpha_0) = 0$, the new transition states and the associated risk-neutral measure (denoted by $Q^{(N)}_{\underline{\alpha}_0(N)}$) are such that*

$$Q^{(N)}_{\underline{\alpha}_0(N)}\left(S^{(N)}_{\underline{\alpha}_0(N)}(N) \geq x\right) = \Phi(d_2) + o(\tfrac{1}{N}).$$

If $(\alpha_0, f_2(\alpha_0))$ is the vertex of the parabola, the new transition states and the associated risk-neutral measure are such that $Q_{\underline{\alpha}_0}(N)^{(N)}\left(S_{\underline{\alpha}_0}(N)^{(N)}(N) \geq x\right)$ admits an asymptotic expansion (in the conventional sense) in powers of $1/N$ up to order $k = 1$; we have

$$Q^{(N)}_{\underline{\alpha}_0(N)}\left(S^{(N)}_{\underline{\alpha}_0(N)}(N) \geq x\right) = \Phi(d_2) + \frac{e^{-\frac{1}{2}d_2^2}}{\sqrt{2\pi}} \frac{1}{36Td_1}\left(d_1^4 + d_1^2 d_2^2 + 5d_1 d_2 - 4d_1^2 - 2 - d_1^3 d_2\right)\frac{1}{N} + o(\tfrac{1}{N}).$$

Since the OD model is a variant of the CP model, we know from Proposition 12 that it ensures weak convergence to the stock price:

Proposition 15. *The sequence of processes $\left(S^{(c,N)}_{\underline{\alpha}_0(N)}\right)_N$ defined from the OD model by linear interpolation and an appropriate time-scaling converges weakly to the stock price process S.*

To summarise, the OD model can further improve the order of convergence of the discretisation error to $o(1/N)$. In any case, the coefficient of the error term of order $1/N$ is smaller than that obtained by the original CP model ($\alpha = 0$).

Remark 10. *Essentially, the optimal drift model admits the rate $o(1/N)$ if $d_1(x)$ is sufficiently small in absolute values. As we see below this covers most cases of practical relevance in numerical option pricing.*

In Section 2.5, we investigate the convergence behaviour of binomial prices for common types of options. We will consider the schemes by CRR and RB and also the advanced schemes discussed above. In particular, we will analyse the impacts of the above results. We will see that for the options under consideration, the OD model achieves a convergence behaviour superior to that obtained by the methods from the literature.

Before, we wish to justify the application of the binomial approach to numerical option valuation. As discussed previously, the binomial processes under consideration ensure that the corresponding sequence of processes $(S^{(c,N)})_N$ (obtained by linear interpolation and time-scaling) converges weakly to the stock price process S. As we see in the next section, the above property provides the theoretical basis for binomial option valuation.

2.3 Convergence of Binomial Option Prices

This section deals with the application of binomial models to numerical option valuation. Assume that the corresponding sequence of approximating processes $(S^{(c,N)})_N$ converges weakly to the stock price process S, which is in particular satisfied for the conventional and the advanced binomial models we considered previously. To apply the binomial model to option valuation, we evaluate the payoff functional g along the sequence of approximating processes $S^{(c,N)}$. By definition of weak convergence, the resulting sequence of binomial prices converges to the exact price provided the payoff functional is bounded and continuous.

We see in the following that the assumption of weak convergence to the stock price process leads to much stronger consequences than the above result on bounded and continuous payoff functions. In particular, according to Skorohod's Theorem, weak convergence can be identified with almost sure convergence on an appropriate probability space:

Theorem 5 (Skohorod)**.** *Let* $X^{(N)}$, $1 \leq N < \infty$, *and* X *be random random variables that take values in a separable metric space* M *such that* $X^{(N)} \Rightarrow_w X$. *Then there exists a probability space* $(\Omega, \mathscr{F}, P)$ *together with some random elements* $Y^{(N)}$, $1 \leq N < \infty$ *and* Y, *such that* $Y^{(N)}$ *and* Y *have the same distribution as* $X^{(N)}$ *and* X, *respectively, (for short:* $Y^{(N)} \sim X^{(N)}$, $Y \sim X$*) and the sequence* $(Y^{(N)})_N$ *converges almost surely to* Y, *for short we write*

$$Y^{(N)} \to Y \quad a.s.$$

(compare e.g. [Kal01], Theorem 4.30).

Concerning practical applications, we show that *the sequence of binomial option prices converges to the exact price for most common European and American types of options*. This justifies the application of the binomial approach to numerical option valuation.

2.3.1 European-Type Options

In this section, we consider the binomial approach to the pricing of European-type options. In the next section, we will briefly discuss the main differences to the American case.

Let $g : C[0,T] \to [0,\infty)$ be the payoff function that describes the future payment at maturity T. The exact option price in the Black-Scholes model is obtained as the discounted expected value of the future payment with respect to the risk-neutral measure Q; that is, the price is given by $E_Q(e^{-rT}g(S))$. Similarly, we obtain the corresponding "binomial price" $E_{P^{(N)}}(e^{-rT}g(S^{(c,N)}))$ if we evaluate the payoff function along the process $S^{(c,N)}$. To avoid misunderstandings, let us stress again that we continue to use the term "binomial price" in a loose sense.

In the following, we investigate the relationship between the exact option price $E_Q(e^{-rT}g(S))$ and the sequence of binomial option prices $(E_{P^{(N)}}(e^{-rT}g(S^{(c,N)})))_N$. In particular, we discuss convergence conditions. But before we demonstrate that for continuous payoff functions, the limes inferior of the binomial prices is an upper bound for the option price.

An Upper Bound on the Option Price

As we anticipate from Fatou's Lemma, the option price is bounded from above by the limes inferior of the corresponding binomial prices if the given payoff functional is continuous almost everywhere. To be precise on the above argument, let us first recall the following technical fact:

Lemma 2. *Let M and M' be metric spaces, and let $\mathscr{B}(M)$ and $\mathscr{B}(M')$ denote the corresponding Borel σ-fields. Then for any function $f : M \to M'$, the set of discontinuities of f*

$$D_f := \{x \in M \mid f \ \underline{\text{not}} \text{ continuous at } x\} \subseteq M$$

is $\mathscr{B}(M)$-measurable (compare e.g. [Bil68], p. 225).

Let g be the payoff function. By the above lemma, the set of discontinuities D_g is $\mathscr{B}(C[0,T])$-measurable and we can hence determine the probability of the set D_g with respect to the law of the stock price process S denoted by $Q \circ S^{-1}(D_g)$.

We can now formulate the following result on the limes inferior of the sequence of binomial option prices:

Proposition 16. *Let $S^{(N)}$ be any binomial process (2.2) that (asymptotically) satisfies the moment matching conditions (2.5) and (2.6), so that $S^{(c,N)} \Rightarrow_w S$. Assume that the* *given payoff functional g is continuous almost everywhere; i.e.* $Q \circ S^{-1}(D_g) = 0$. *Then we have*

$$E_Q\left(e^{-rT} g(S)\right) \leq \liminf_{N\to\infty} E_{P^{(N)}}\left(e^{-rT} g(S^{(c,N)})\right).$$

Proof. According to Skorohod's Theorem, there is a probability space $(\Omega, \mathscr{F}, P)$ together with random variables $Y^{(N)}$, $1 \leq N < \infty$ and Y such that $Y^{(N)} \sim S^{(c,N)}$, $Y \sim S$ and $Y^{(N)} \to Y$ a.s. Consequently, as g is continuous almost everywhere, we have that $g(Y^{(N)}) \to g(Y)$ a.s. Thus, as $g \geq 0$ Fatou's Lemma yields

$$E_P\left(e^{-rT} g(Y)\right) \leq \liminf_{N\to\infty} E_P\left(e^{-rT} g\left(Y^{(N)}\right)\right),$$

which proves the assertion. □

We see that if the payoff functional is continuous almost everywhere and the sequence of corresponding binomial prices converges to some limit, *the exact option price is never below this limit.* We next discuss conditions that ensure convergence to the exact option price.

Convergence Conditions

By definition of weak convergence, the sequence of binomial prices converges to the exact price if the payoff functional is bounded and continuous. Yet with respect to practical applications this result is clearly unsatisfactory. Many traded options have an unbounded payoff function, the plain vanilla call being an obvious example. Further, for many traded options the payoff functional is discontinuous in the stock price process. A prominent example are barrier options for which the right to exercise either appears or disappears on certain regions of the path space of S.

As we show now, the continuity assumption can actually be weakened to the assumption that the set of discontinuities of the payoff function has zero probability with respect to the law of S. This is a direct consequence of the following variant of the continuous mapping principle:

Lemma 3 (Continuous Mapping Principle II)**.** *Let M and M' be metric spaces and let X, $(X^{(N)})_N$ be M-valued RVs defined on probability spaces $(\Omega^{(N)}, \mathscr{F}^{(N)}, P^{(N)})$ and $(\Omega, \mathscr{F}, P)$, respectively. Further, let $h : M \to M'$ be Borel measurable with $P \circ X^{-1}(N_h) = 0$, where $P \circ X^{-1}$ is the distribution of X and N_h is the set of discontinuities of h. Then, if $X^{(N)} \Rightarrow_w X$, it holds that $h(X^{(N)}) \Rightarrow_w h(X)$ (compare e.g. [Bil68], Theorem 5.1).*

Furthermore, the boundedness assumption on the payoff can be weakened to uniform integrability (for short: UI) of the sequence of RVs $(g(S^{(c,N)}))_N$; that is,

$$\lim_{C\to\infty}\left(\sup_{N\in\mathbb{N}} E_{P^{(N)}}\left(\left|g\left(S^{(c,N)}\right)\right| 1_{\{|g(S^{(c,N)})|>C\}}\right)\right) = 0.$$

This a consequence of the following well-known result:

Theorem 6. *Let M be a metric space. Let $(X^{(N)})_N$ and X be M-valued RVs defined on probability spaces $(\Omega^{(N)}, \mathscr{F}^{(N)}, P^{(N)})$ and $(\Omega, \mathscr{F}, P)$, respectively. Assume that $X^{(N)} \Rightarrow_w X$. Then, if the sequence $(X^{(N)})_N$ is UI,*

$$\lim_{N\to\infty} E_{P^{(N)}}(X^{(N)}) = E_P(X).$$

(compare e.g. [Bil68], Theorem 5.4.)

Combining the new assumptions, we obtain the following result on convergence of binomial option prices to the exact price:

Proposition 17. *Let $S^{(N)}$ be any binomial process (2.2) which (asymptotically) satisfies the moment matching conditions (2.5) and (2.6), so that $S^{(c,N)} \Rightarrow_w S$. We assume that*

- *the payoff functional g is continuous almost everywhere, and*
- *the sequence $(g(S^{(c,N)}))_N$ is UI.*

Then the corresponding sequence of binomial option prices converges to the exact price; i.e.

$$E_{P^{(N)}}\left(e^{-rT} g\left(S^{(c,N)}\right)\right) \to E_Q\left(e^{-rT} g(S)\right) \qquad \text{as } N\to\infty.$$

Proof. According to the above variant of the continuous mapping principle, it follows from weak convergence of $S^{(c,N)}$ to the stock price S that

$$g(S^{(c,N)}) \Rightarrow_w g(S).$$

Theorem 6 then yields the assertion. □

Remark 11. *If the payoff function satisfies the assumption of Proposition 17, the binomial price $E_{P^{(N)}}(e^{-rT} g(S^{(c,N)}))$ provides an estimate of the exact price if the corresponding number of periods N is sufficiently large. This justifies the application of the binomial method as a numerical pricing technique.*

In certain situations, it may turn out to be difficult to establish uniform integrability. However, the following criterion by de la Vallée-Poussin is often useful in this context:

Lemma 4 (De la Vallée-Poussin criterion)**.** *Let $(X^{(N)})_N$ be a sequence of integrable RVs each defined on a probability space $(\Omega^{(N)}, \mathscr{F}^{(N)}, P^{(N)})$. Assume that $f : [0,\infty) \to [0,\infty)$ is an increasing function which is such that*

$$\lim_{t\to\infty} \frac{f(t)}{t} = \infty$$

and $E_{P^{(N)}}(f(|X^{(N)}|))$ is uniformly bounded, i.e.

$$\sup_{N\in\mathbb{N}} E_{P^{(N)}}(f(|X^{(N)}|)) < \infty.$$

Then the sequence $(X^{(N)})_N$ is UI (compare e.g. [Ŝ88], Lemma 6.3).

In the following we discuss the assumption of Proposition 17. We will see that the requirements are satisfied for most common types of options. First, we consider two special cases; namely, barrier options and plain vanilla options. Afterwards, we show that Proposition 17 actually applies to every payoff functional that is polynomially bounded and continuous almost everywhere.

Barrier options In this section, we show that the binomial approach can be applied to the valuation of barrier options. We consider only barriers that are constant in the stock price. One distinguishes between four basic forms; *down-and-out*, *down-and-in*, *up-and-out* or *up-and-in*, which indicates whether the right to exercise originates (*in*) or expires (*out*) on the barrier and whether the barrier is set above (*up*) or below (*down*) the spot price. Note that by arbitrage arguments, barrier options are cheaper than the corresponding plain vanilla option. As a consequence, barrier options are widely used both as hedging and as speculative instruments.

We illustrate the application of the binomial approach to options with constant barriers by considering a cash-or-nothing option with an up-and-out barrier. We have the following result:

Proposition 18. *Let $S^{(N)}$ be any binomial process (2.2) that (asymptotically) satisfies the moment matching conditions (2.5) and (2.6), so that $S^{(c,N)} \Rightarrow_w S$. Let $B > s_0$ and consider the payoff*

$$g(S) = 1_{\{S_t \le B\, \forall t\in[0,T]\}}. \tag{2.65}$$

Then the corresponding sequence of binomial option prices converges to the exact price; i.e.

$$E_{P^{(N)}}\left(e^{-rT} g\left(S^{(c,N)}\right)\right) \to E_Q\left(e^{-rT} g(S)\right) \qquad \text{as } N \to \infty.$$

Proof. Note first that payoff function g given in (2.65) is bounded. Due to Proposition 17, it therefore suffices to show that g is continuous almost everywhere; i.e. $Q \circ S^{-1}(D_g) = 0$. Note that its set of discontinuities D_g contains exactly those functions $w \in C[0,T]$ that hit the barrier B at some time $t \leq T$, but do not cross it; i.e.

$$Q \circ S^{-1}(D_g) = Q\left(\max_{t \leq T} S_t = B\right)$$

We can now write

$$\{\max_{t \leq T} S_t = B\} = \{\max_{t \leq T} \tilde{S}_t = \tilde{B}\},$$

where

$$\tilde{S}_t := (\tfrac{r}{\sigma} - \tfrac{1}{2}\sigma)t + W_t \qquad \text{and} \qquad \tilde{B} := \tfrac{1}{\sigma} \ln B/s_0.$$

We know from the reflection principle of Brownian motions (compare e.g. [KS98], Proposition 2.6.19) that $\max_{t \leq T} W_t$ is continuously distributed; thus, $Q(\max_{t \leq T} W_t = \tilde{B}) = 0$. Consequently, it remains to discuss whether the presence of the drift $(r/\sigma - \frac{1}{2}\sigma)$ causes any difficulties. But, in fact this is not the case because as we know from the Girsanov-Cameron-Martin Theorem (compare e.g. [KS98], Proposition 3.5.1), the law of $\tilde{S}$ is absolutely continuous with respect to the law of the Brownian motion W. As a result, we see that

$$Q \circ \tilde{S}^{-1}(D_g) = Q(\max_{t \leq T} \tilde{S}_t = \tilde{B}) = 0,$$

which completes the proof. □

Remark 12. *According to the above result, binomial option valuation can be applied to barrier options. However, as we will show below, the corresponding binomial prices exhibit an irregular convergence behaviour. Binomial prices for cash-or-nothing options suffer from similar difficulties. Cash-or-nothing options are constant in the terminal stock price, so that they have a single point of discontinuity at the strike value. Let us anticipate at this point that the advanced models described previously can significantly improve the convergence behaviour of the corresponding binomial prices in this case.*

Plain vanilla calls For plain vanilla puts, convergence of binomial prices to the Black-Scholes price follows directly from the definition of weak convergence. By contrast, the definition of weak convergence does not apply to plain vanilla calls because the corresponding payoff function is unbounded in the terminal stock price. In this section, we show that by Proposition 17, binomial option valuation can be applied to plain vanilla calls.

Since the payoff function of a plain vanilla call is continuous in the terminal stock price, it suffices to show that the corresponding sequence $(g(S^{(c,N)}))_N$ is UI. This requires to distinguish between the particular discretisation scheme under consideration. We will first consider a binomial process $S^{(N)}$ with risk-neutral transition probabilities. In this case, we can use the fact that $S_k^{(N)} e^{-rkT/N}$ is a discrete martingale.

Proposition 19. *Let $S^{(N)}$ be the binomial process (2.2) with risk-neutral transition probabilities. Suppose that $S^{(c,N)} \Rightarrow_w S$ (i.e. we have $\beta = \sigma$). We consider a plain vanilla call; i.e.*

$$g(S) = (S_T - K)^+$$

for some strike value $K > 0$. Then the corresponding sequence of binomial option prices converges to the exact price; i.e.

$$E_{P^{(N)}}\left(e^{-rT} g\left(S^{(c,N)}\right)\right) \to E_Q\left(e^{-rT} g(S)\right) \qquad \text{as } N \to \infty.$$

Proof. We use the de la Vallée-Poussin criterion with $f(t) = t^\delta$, $\delta > 1$. That is, we need to show that $E_{Q^{(N)}}(g(S^{(c,N)}))^\delta$ is uniformly bounded. Note first that

$$E_{Q^{(N)}}\left(g\left(S^{(c,N)}\right)\right)^\delta \le E_{Q^{(N)}}\left(S_T^{(c,N)}\right)^\delta = E_{Q^{(N)}}\left(S_N^{(N)}\right)^\delta.$$

Under the risk-neutral measure, $S_k^{(N)} e^{-rkT/N}$ is a discrete martingale, which implies that

$$\sup_{N \in \mathbb{N}} E_{Q^{(N)}}\left(S_N^{(N)}\right) = e^{rT} s_0.$$

For $\delta > 1$, we obtain that

$$\begin{aligned} \sup_{N\in\mathbb{N}} E_{Q^{(N)}}\left(S_N^{(N)}\right)^\delta &= \sup_{N\in\mathbb{N}} E_{Q^{(N)}}\left(s_0 e^{\alpha(N)T + \sigma\sqrt{T/N}\sum_{k=1}^N Z_k^{(N)}}\right)^\delta \\ &= \sup_{N\in\mathbb{N}} \left\{ s_0^{\delta-1} e^{\alpha(N)(\delta-\delta^2)T} E_{Q^{(N)}}\left(s_0 e^{\alpha(N)\delta^2 T + \sigma\sqrt{\frac{\delta^2 T}{N}}\sum_{k=1}^N Z_k^{(N)}}\right)\right\} \\ &= \sup_{N\in\mathbb{N}} \left\{ s_0^\delta e^{\alpha(N)(\delta-\delta^2)T} e^{r\delta^2 T}\right\}, \end{aligned}$$

where the last equality above follows from the martingale property of $S_k^{(N)} e^{-rkT/N}$ together with a re-scaling of time by δ^2 (for the re-scaling argument compare also

[AK94]). It then follows from boundedness of $(\alpha(N))_N$ that

$$\sup_{N \in \mathbb{N}} E_{Q^{(N)}} \left(S_N^{(N)} \right)^\delta < \infty, \tag{2.66}$$

and thus

$$\sup_{N \in \mathbb{N}} E_{Q^{(N)}} \left(g \left(S^{(c,N)} \right) \right)^\delta < \infty.$$

The assertion then follows from Proposition 17 together with the de la Vallée-Poussin criterion. □

Note in particular that the above result justifies binomial valuation of a plain vanilla call with the advanced models considered previously. Yet it remains to discuss the application of the methods suggested by CRR and by RB, respectively. For these models, the transition probabilities are not chosen according to the risk-neutral measure, so that the corresponding process $S_k^{(N)} e^{-rkT/N}$ is no longer a martingale. However, as observed in Corollary 3, if the number of periods N is sufficiently large, the probability for an up-movement $p_u(N)$ is smaller than the corresponding risk-neutral probability $q_u(N)$. This suggests that the corresponding sequence $(E_{P^{(N)}}(g(S^{(c,N)}))^\delta)_N$, $\delta > 1$, is uniformly bounded because this condition is already satisfied for the risk-neutral case.

Proposition 20. *Let $S^{(N)}$ be the binomial process (2.2) suggested by CRR or by RB, respectively. We consider a plain vanilla call with strike $K > 0$; i.e. $g(S) = (S_T - K)^+$. Then the corresponding sequence of binomial option prices converges to the exact price; i.e.*

$$E_{P^{(N)}} \left(e^{-rT} g \left(S^{(c,N)} \right) \right) \to E_Q \left(e^{-rT} g(S) \right) \qquad \text{as } N \to \infty.$$

Proof. As before, we use the de la Vallée-Poussin criterion with $f(t) = t^\delta$, $\delta > 1$. We know from Corollary 3 that for both the CRR model and the RB model, there exists some $N_0 \in \mathbb{N}$ such that

$$q_u(N) \geq p_u(N) \qquad \text{for all } N \geq N_0.$$

We hence obtain that for $N \geq N_0$,

$$\begin{aligned} E_{P^{(N)}} \left(S_T^{(c,N)} \right)^\delta &= E_{P^{(N)}} \left(S_N^{(N)} \right)^\delta \\ &= \prod_{k=1}^{N} C E_{P^{(N)}} \left(e^{\sigma \delta \sqrt{T/N} Z_k^{(N)}} \right) \end{aligned}$$

$$
\begin{aligned}
&= C^N \left(p_u(N) e^{\sigma\delta\sqrt{T/N}} + (1-p_u(N)) e^{-\sigma\delta\sqrt{T/N}} \right)^N \\
&\leq C^N \left(q_u(N) e^{\sigma\delta\sqrt{T/N}} + (1-q_u(N)) e^{-\sigma\delta\sqrt{T/N}} \right)^N \\
&= E_{Q^{(N)}} \left(S_N^{(N)} \right)^\delta ,
\end{aligned}
$$

where $C = s_0^\delta e^{\delta\alpha T}$ with $\alpha = 0$ for the CRR model and $\alpha = r - 1/2\sigma^2$ for the RB model. By (2.66), the family $(S_N^{(N)})_N$ is UI with respect to the risk-neutral measures $Q^{(N)}$. Thus, we have

$$
\sup_{N \geq N_0} E_{P^{(N)}} \left(S_N^{(N)} \right)^\delta \leq \sup_{N \geq N_0} E_{Q^{(N)}} \left(S_N^{(N)} \right)^\delta < \infty. \tag{2.67}
$$

We see that the family $(g(S^{(c,N)}))_N$ is UI with respect to the chosen probability measure $P^{(N)}$, which completes the proof by Proposition 17 and the de la Vallée-Poussin criterion. □

Remark 13. *Clearly, in practical applications the binomial method is not used for the valuation of a plain vanilla call because its price is readily available from the Black-Scholes formula. However as we show in the next section, the above arguments can be used to generalise the convergence result.*

Polynomially bounded payoff functionals In this section, we show that for the models under consideration the above results on the application of binomial option valuation to plain vanilla calls can be generalised to any type of option for which the payoff functional is polynomially bounded and continuous almost everywhere.

For the special case that the binomial model exhibits risk-neutral transition probabilities, the above result is shown in Amin and Khanna (1994) ([AK94], Section 5). In fact, the result extends to any binomial process (2.2) that (asymptotically) satisfies the moment matching conditions. As a consequence, the assumption of Proposition 21 covers, in particular, the schemes suggested by CRR and by RB, which are not considered by Amin and Khanna. As before, the essential difference is that for these methods, the corresponding process $S_k^{(N)} e^{-rkT/N}$ is not a martingale.

Proposition 21. *Let $S^{(N)}$ be any binomial process (2.2) that (asymptotically) satisfies the moment matching conditions (2.5) and (2.6), so that $S^{(c,N)} \Rightarrow_w S$. We consider any type of option whose payoff functional is continuous almost everywhere and bounded above by a polynomial; i.e. there are constants $C > 0$ and $p > 1$ such that*

$$
g(S) \leq C \left(1 + \sup_{t \leq T} |S_t| \right)^p , \qquad S \in C[0,T].
$$

Then the corresponding sequence of binomial option prices converges to the exact price; i.e.

$$E_{P^{(N)}}\left(e^{-rT}g\left(S^{(c,N)}\right)\right) \to E_Q\left(e^{-rT}g(S)\right) \qquad \text{as } N \to \infty.$$

Proof. As before, let $\delta > 1$. By the assumption on g,

$$\begin{aligned} E_{P^{(N)}}\left(g\left(S^{(c,N)}\right)\right)^{\delta} &\leq E_{P^{(N)}}\left(C^{\delta}\left(1+\sup_{t\leq T} S_t^{(c,N)}\right)^{\gamma}\right) \\ &\leq C^{\delta}2^{\gamma}\left(1+E_{P^{(N)}}\left(\sup_{t\leq T} S_t^{(c,N)}\right)^{\gamma}\right), \end{aligned}$$

where $\gamma := \delta p > 1$. Let us re-write $S^{(c,N)}$ in terms of the normalised RVs $Y_k^{(N)}$, i.e.

$$S^{(c,N)}(t) = s_0 e^{\mu(N)t+|\sigma(N)|\sqrt{T}\,Y^{(c,N)}\left(\frac{t}{T}\right)}, \qquad t \in [0,T],$$

where

$$Y^{(c,N)}(s) = \frac{1}{\sqrt{N}}\left(\sum_{k=1}^{[Ns]} Y_k^{(N)} + (Ns-[Ns])\,Y_{[Ns]+1}^{(N)}\right), \qquad s \in [0,1].$$

(compare (2.23)). First, we consider the case $(r-1/2\sigma^2) \geq 0$. Then by the asymptotic moment matching condition (2.5) on the logreturns, there exists some $N_0 \in \mathbb{N}$ such that $\mu(N) \geq 0$ for $N \geq N_0$. It follows that for $N \geq N_0$,

$$\sup_{t\leq T} S_t^{(c,N)} \leq s_0 e^{\mu(N)T} \sup_{t\leq T} e^{|\sigma(N)|\sqrt{T}\,Y^{(c,N)}\left(\frac{t}{T}\right)}.$$

We hence obtain that for $N \geq N_0$,

$$\begin{aligned} E_{P^{(N)}}\left(\sup_{t\leq T} S_t^{(c,N)}\right)^{\gamma} &\leq s_0^{\gamma} e^{\gamma\mu(N)T} E_{P^{(N)}}\left(\sup_{t\leq T} e^{|\sigma(N)|\sqrt{T}Y^{(c,N)}\left(\frac{t}{T}\right)}\right)^{\gamma} \\ &= s_0^{\gamma} e^{\gamma\mu(N)T} E_{P^{(N)}}\left(\max_{k=1,\ldots,N} e^{|\sigma(N)|\sqrt{T}M_k^{(N)}}\right)^{\gamma}, \end{aligned}$$

where $M^{(N)}$ is the discrete martingale defined in (2.26); i.e. $M_k^{(N)} = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} Y_i^{(N)}$, $1 \leq k \leq N$, and $M_0^{(N)} = 0$. In particular, the discrete process defined by $e^{|\sigma(N)|\sqrt{T}M_k^{(N)}}$, $0 \leq k \leq N$ is a submartingale. Hence, it follows from Doob's L^p inequality (compare e.g. [Dur05], Theorem 4.4.3)) that

$$E_{P^{(N)}}\left(\sup_{t\le T} S_t^{(c,N)}\right)^\gamma \le \left(\frac{\gamma}{\gamma-1}\right)^\gamma s_0^\gamma e^{\gamma\mu(N)T} E_{P^{(N)}}\left(e^{|\sigma(N)|\sqrt{T}M_N^{(N)}}\right)^\gamma = \left(\frac{\gamma}{\gamma-1}\right)^\gamma E_{P^{(N)}}\left(S_N^{(N)}\right)^\gamma. \tag{2.68}$$

It remains to consider the case $(r-1/2\sigma^2)<0$. By the (asymptotic) moment matching condition, there exists some $N_0 \in \mathbb{N}$ such that

$$2(r-1/2\sigma^2) < \mu(N) < 0 \qquad \text{for } N \ge N_0. \tag{2.69}$$

Then for $N \ge N_0$, the drift $\mu(N)$ is negative, which implies that

$$\sup_{t\le T} S_t^{(c,N)} \le s_0 \sup_{t\le T} e^{|\sigma(N)|\sqrt{T}Y^{(c,N)}\left(\frac{t}{T}\right)} \qquad \text{for } N \ge N_0.$$

Then by similar arguments as above, we have

$$E_{P^{(N)}}\left(\sup_{t\le T} S^{(c,N)}\right)^\gamma \le \left(\frac{\gamma}{\gamma-1}\right)^\gamma e^{-\gamma\mu(N)T} E_{P^{(N)}}\left(S_N^{(N)}\right)^\gamma \qquad \text{for } N \ge N_0.$$

Further by (2.69), the factor $e^{-\gamma\mu(N)T}$ on the right-hand side of the above inequality can be bounded by $e^{-2\gamma(r-1/2\sigma^2)T}$, which implies

$$E_{P^{(N)}}\left(\sup_{t\le T} S^{(c,N)}\right)^\gamma \le \left(\frac{\gamma}{\gamma-1}\right)^\gamma e^{-2\gamma(r-1/2\sigma^2)T} E_{P^{(N)}}\left(S_N^{(N)}\right)^\gamma. \tag{2.70}$$

Combining the results in (2.68) and in (2.70) for $(r-1/2\sigma^2)\ge 0$ and $(r-1/2\sigma^2)<0$, respectively, shows that there is always some constant $K \ge 0$ such that

$$\sup_{N\ge N_0} E_{P^{(N)}}\left(g\left(S^{(c,N)}\right)\right)^\delta \le C^\delta 2^\gamma \left(1+\sup_{N\ge N_0} E_{P^{(N)}}\left(\sup_{t\le T} S^{(c,N)}\right)^\gamma\right) \le C^\delta 2^\gamma\left(1+\left(\frac{\gamma}{\gamma-1}\right)^\gamma K \sup_{N\ge N_0} E_{P^{(N)}}\left(S_N^{(N)}\right)^\gamma\right).$$

As the family $(E_{P^{(N)}}(S_N^{(N)})^\gamma)_N$, $\gamma>1$, has already been shown to be uniformly bounded (compare 2.67), the above result implies that

$$\sup_{N\in\mathbb{N}} E_{P^{(N)}}\left(g\left(S^{(c,N)}\right)\right)^{\delta} < \infty.$$

Consequently, the assertion follows again from Proposition 17 together with the de la Vallée-Poussin criterion. □

Remark 14. *With respect to practical applications, binomial option valuation is thus justified for most common types of options.*

A counterexample In principle, we can of course think of payoff functionals for which the binomial option prices do not converge to the exact price. For example, consider the payoff function

$$g(S) = \begin{cases} 1 & S_T \in \left\{S_0 \exp\left(\left(r - \frac{1}{2}\sigma^2\right)T + \sigma\frac{k}{\sqrt{N}}\right) \mid N \in \mathbb{N}, -N \leq k \leq N\right\} \\ 0 & \text{else} \end{cases}$$

Then for any number of periods N, the RB model suggests the price $E_{P^{(N)}}\left(e^{-rT}g(S^{(c,N)})\right) = e^{-rT}$, while the exact price is given by $E_Q\left(e^{-rT}g(S)\right) = 0$. In this case, the binomial prices do not converge to the option price. Note that in accordance with Proposition 16, the limit e^{-rT} obtained along the sequence of RB models is greater than the correct price.

Apparently, the above example is artificial as the payoff function is constructed in accordance with the specific distribution of the RB model. However, the example shall serve as a warning to stress that weak convergence to the stock price process S does not always imply convergence to the expected value of some functional of S. However, as discussed above, in practical applications the binomial approach can be applied to most common valuation problems.

2.3.2 American-Type Options

The above results focus on European options. For American-style options, similar results are not readily available because the American option valuation problem involves a non-trivial timing decision. We briefly sketch the main ideas.

For a given payoff function g, the exact price of the corresponding American option is $E_Q(e^{-r\tau_{\text{opt}}}g(S(\tau_{\text{opt}})))$, where τ_{opt} is the optimal stopping time for the valuation problem under consideration. Similarly, the binomial price is given by $E_{P^{(N)}}(e^{-r\tau_*^{(N)}}g(S^{(c,N)}(\tau_*^{(N)})))$, where $\tau_*^{(N)}$ is the optimal stopping time associated with the binomial valuation problem.

If $S^{(c,N)}$ converges weakly to S, the sequence $(S^{(c,N)}, \tau_*^{(N)})_{N\in\mathbb{N}}$ is tight in $C[0,T]\times[0,T]$. Thus due to Prohorov's Theorem (compare e.g. [KS98], Theorem 2.4.7), any subsequence $(S^{(c,N_k)}, \tau_*^{(N_k)})_{N_k}$ has a further subsequence $(S^{(c,N_{k_l})}, \tau_*^{(N_{k_l})})_{N_{k_l}}$ that converges to some weak limit. Let (S,τ) be the limit of one such subsequence. Then S is the original stock price. However, τ depends on the particular subsequence chosen. Further, it is not clear whether τ is a stopping with respect to the filtration generated by S. This makes the analysis of binomial option valuation more complicated than in the case of European-style options.

In order to obtain an upper bound for the binomial option prices, we next follow Amin and Khanna (1994). For detailed arguments we refer to [AK94]. Amin and Khanna show that τ can be identified, in "some appropriate sense", with a legitimate stopping time with respect to the filtration generated by S. Their arguments are essentially based on the result by Kushner (1977) that there is a suitable filtered probability space $(\Omega,\mathscr{F},\mathscr{F}_t,P)$ and a pair (Y,τ_Y) defined on $((\Omega,\mathscr{F},P))$ such that Y is a geometric Brownian motion with respect to $\mathscr{F}_t$, $(Y,\tau_Y)\sim(S,\tau)$ and τ_Y is a stopping time with respect to $\mathscr{F}_t$ (compare [Kus77], Theorem 8.2.4). Amin and Khanna then show that if the payoff satisfies the assumptions of continuity and uniform integrability, then

$$\lim_{l\to\infty} E_{P^{(N_{k_l})}}\left(e^{-r\tau_*^{(N_{k_l})}} g\left(S^{(c,N_{k_l})}(\tau_*^{(N_{k_l})})\right)\right) \le E_Q\left(e^{-r\tau_{\text{opt}}} g\left(S(\tau_{\text{opt}})\right)\right).$$

As the subsequence $(S^{(c,N_k)}, \tau_*^{(N_k)})_{N_k}$ is chosen arbitrarily, it follows that

$$\limsup_{N\to\infty} E_{P^{(N)}}\left(e^{-r\tau_*^{(N)}} g(S^{(c,N)}(\tau_*^{(N)}))\right) \le E_Q(e^{-r\tau_{\text{opt}}} g(S,\tau_{\text{opt}})). \tag{2.71}$$

The reverse implication[4]

$$\liminf_{N\to\infty} E_{P^{(N)}}\left(e^{-r\tau_*^{(N)}} g\left(S^{(c,N)}\right)\right) \ge E_Q\left(e^{-r\tau_{\text{opt}}} g\left(S(\tau_{\text{opt}})\right)\right) \tag{2.72}$$

follows from the above argument and a direct extension of Proposition 16. The result in (2.71) then implies that the assertion of Proposition 17 extends to American-type options.

[4] Amin and Khanna refer to a convergence result of Kushner (1977) to establish (2.72).

2.4 Tree Procedures

As discussed in the previous section, the binomial method can be applied to numerical valuation of most common types of European and American options. The corresponding algorithm is called a *tree procedure* because the possible realisations of the binomial process $S^{(N)}$ can be identified with a tree structure: The initial value s_0 is designated the root of the tree, the terminal values of $S^{(N)}$ can be identified with the leaves of the tree and each node in the interior is connected to two successor nodes.

In Section 2.4.1 we describe the tree algorithm in detail. We see that for many standard types of options, computational effort is of order $O(N^2)$; and thus, in particular polynomial in N. In Section 2.4.2 we focus on numerical valuation of path-independent options. We find a conceptual link between binomial tree algorithms and explicit finite difference approximations to the Black-Scholes pricing PDE. Both the method suggested by CRR and by RB can be identified with an explicit finite difference approximation to the appropriately transformed Black-Scholes PDE. Of course, the above result on numerical valuation techniques appears as a consequence of the natural link between the martingale approach and the PDE approach to continuous-time option pricing.

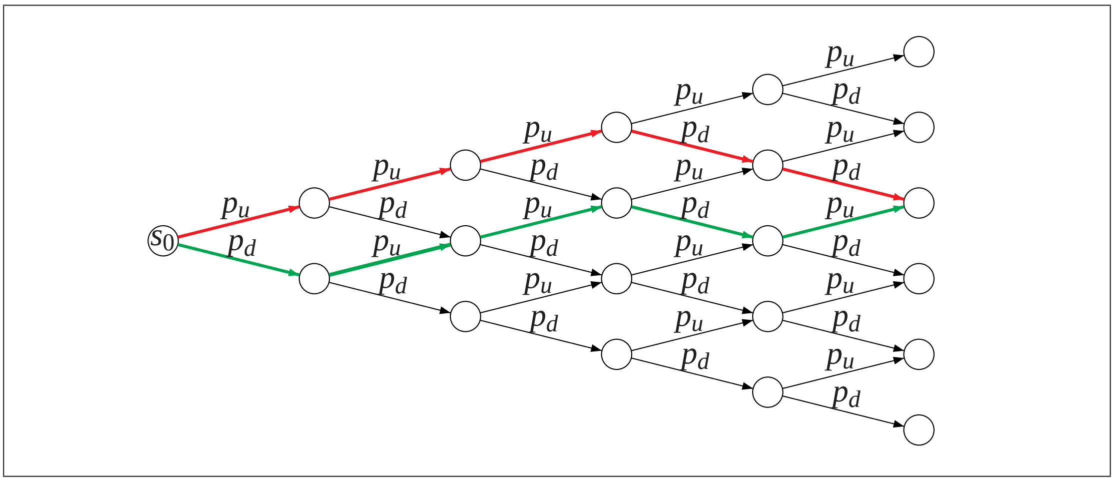
Before we consider the tree algorithm in detail, we wish to stress that in practical applications, binomial option valuation is of course only applied, if an analytic formula for the exact price is not readily available. Then the error in the binomial option price is not known; one may only have an intuition. Hence, how to choose the number of periods such that the pricing error in the corresponding binomial model is sufficiently small for the valuation problem of interest? In fact, in practical applications, an appropriate number of periods is often determined by computing binomial prices on gradually finer grids *until subsequent estimates vary less than some threshold*.

2.4.1 Variants of the Tree Algorithm

Binomial option valuation allows only for a finite number of possible payoff scenarios. In principle, there can be 2^N distinct scenarios, one for each path through the tree. In this case, the computational effort required to compute the expected payoff is non-polynomial in N. However, for many types of options different paths can realise the same payoff. This is due to *the re-combining structure of the tree*. That is, by definition of $S^{(N)}$, paths with the same number of up- and down-movements end at the same node, independently of the order in which the up- and down-movements have occurred.

The recombining structure of binomial trees is illustrated in Figure 2.1. In the example, the red and the green path end at the same node, but differ in the order of the up- and down-movements.

Fig. 2.1: A five-period binomial tree



For path-independent options, the re-combining structure of the tree implies that there are $N+1$ distinct payoff scenarios only, each belonging to a terminal node of the tree. Computational effort is therefore only of order $O(N^2)$. For path-dependent options, however, it depends on the specific payoff functional whether the re-combining structure of the tree can be used to reduce computational effort. In contrast to path-independent options, the realised payoff can depend on the order of the up-and down-movements. A prominent example are options on the average value of the stock price process over time; that is, $g(S) = g(1/T \int_0^T S_t \mathrm{d}t)$.

However, for many types of path-dependent options computational effort can still be reduced to order $O(N^2)$. This will be verified for barrier options and for American options. We will first set up the tree algorithm for the valuation of path-independent options. Afterwards we will discuss how to adapt the algorithm to the path-dependent case.

Path-independent options For path-independent options, it suffices to compute the expectation of possible terminal values. Due to the Markov property of the binomial process $S^{(N)}$, we can compute the expected terminal payoff by ”stepping backwards through the time layers of the re-combining tree”, which is formally based on the equality:

$$E_{P^{(N)}}\left(g(S_N^{(N)})\right) = E_{P^{(N)}}\left(E_{P^{(N)}}\left(\dots E_{P^{(N)}}\left(g(S_N^{(N)})\middle| S_{N-1}^{(N)}\right)\dots\middle| S_1^{(N)}\right)\right)$$

Hence, we obtain the following recursion: We start at the final time by assigning the payoff scenarios to the terminal nodes. We then step backwards through the time layers of the tree by computing the weighted sum of the values assigned to the successor nodes. The algorithm is given in pseudo-code (compare Algorithm 1).

Algorithm 1: Binomial tree for path-independent European options

---

Input: stock price parameters, risk-neutral rate r, payoff function g
Return: price estimate $= V[0] \times \exp(-r \times T)$

1. Forward Step

```
{initialise asset prices at maturity}
    Set SN[0] := s_0 d^N
for k = 1 to N do
        SN[k] := SN[k-1] × (u/d)
end for

{initialise option values at maturity}
for k = 0 to N do
        Set V[k] := g(SN[k])
end for
```

2. Backward Induction

```
{step backwards through the tree}
for k = N-1 to 0 do
   for l = 0 to k do
          V[l] := pu × V[l+1] + pd × V[l]

      {or under the RB scheme}
          V[l] := 0.5 × (V[l+1] + V[l])
   end for
end for
```

In Algorithm 1, $pu := p_u(N)$, $pd := p_d(N)$, $u := u(N)$, $d := d(N)$ are constants to be determined in advance. As for standard tree implementations, we ”do not span the tree”; i.e. we do not allocate memory for each node in the tree. Instead, the memory already reserved is overwritten at each step of the backward recursion in order to reduce the memory allocation required. Regarding computational effort, we have the following result:

Proposition 22. *Computational effort for Algorithm 1 is in general* $3/2N^2 + O(N)$. *For the discretisation scheme suggested by RB, computational effort reduces to* $N^2 + O(N)$, *which is optimal for the rate of growth of the tree (i.e. for the number of the successor nodes* $n = 2$*).*

Proof. In the backward step, the arithmetic mean has to be computed at each node of the tree, i.e. we have to consider $\sum_{i=0}^{N-1}(i+1) = N(N+1)/2$ nodes. This implies that computational effort for backward induction is $3/2N^2 + O(N)$ because it requires three operations in general (two multiplications and one addition) to compute the arithmetic mean. The forward step is negligible because the corresponding computational effort is only of order $O(N)$. With respect to the second part of the assertion, note that by the distributive law computing the arithmetic means is cheapest (one multiplication and one addition) for the case $p_u(N) = p_d(N) = 1/2$. □

Remark 15. *For the numerical examples on binomial valuation of single-asset options considered in Section 2.5, the above result does not lead to a significant difference in computing time. However, for our examples on multi-dimensional valuation problems analysed in Chapter 3 and in Chapter 4, computing time will be reduced significantly if the transition probabilities are chosen to be equal.*

For path-dependent options, it depends on the specific payoff functional whether there exists a suitable modification of the above tree algorithm. Next we show how barriers which are constant in the stock price can be incorporated into Algorithm 1.

Barrier options To apply the binomial approach to barrier options, we have to distinguish between knock-in and knock-out barriers. In the latter case, we are interested in events of the form $A := \{S_k^{(N)} < B \ \forall k = 1, \ldots, N\}$ with a barrier level $B > s_0$ for up-and-out options and in events of the form $A := \{S_k^{(N)} > B \ \forall k = 1, \ldots, N\}$ with a barrier level $B < s_0$ for down-and-out options. Let us first consider up-and-out options. Then the equality

$$1_{\{S_k^{(N)} < B \ \forall k=1,\ldots,N\}} = 1_{\{S_1^{(N)} < B\}} 1_{\{S_2^{(N)} < B\}} \cdots 1_{\{S_N^{(N)} < B\}} \tag{2.73}$$

and the Markovian structure of $S^{(N)}$ yield

$$E_{P^{(N)}}\left(g(S_N^{(N)}) 1_{\{S_k^{(N)} < B \ \forall k=1,\ldots,N\}}\right) = E_{P^{(N)}}\left(E_{P^{(N)}}\left(\ldots E_{P^{(N)}}\left(g(S_N^{(N)}) 1_{\{S_N^{(N)} < B\}} \middle| S_{N-1}^{(N)}\right) \ldots 1_{\{S_1^{(N)} < B\}} \middle| S_1^{(N)}\right)\right).$$

As a result, the backward induction step of Algorithm 1 can still be applied (compare Algorithm 2). Yet in addition to path-independent options, a zero option value is assigned to the nodes above the barrier level. Similarly, for down-and-out options a zero option value is assigned to the nodes below the barrier level. This means in particular that, while stepping backwards through the time layers of the tree, the array of transition states needs to be adjusted to the current time. The forward step remains unchanged. Clearly, the modified tree algorithm still requires computational effort of order $O(N^2)$.

For knock-in options, we are interested in events of the form $A := \{\exists k_0 \in \{1,\ldots,N\} :\ S^{(N)}(k_0) \geq B\}$ with a barrier level $B > s_0$ for up-and-in options and in events of the form $A := \{\exists k_0 \in \{1,\ldots,N\} :\ S^{(N)}(k_0) \leq B\}$ with a barrier level $B < s_0$ for down-and-in options. *In contrast to the knock-out case (compare (2.73)), the event A cannot be written as a simple product of one-step events.* Consequently, we cannot simply decide on the occurrence of the event A while stepping backwards through the tree. However, we have

$$P^{(N)}\Big(\exists k_0 \in \{1,\ldots,N\} :\ S^{(N)}(k_0) \geq B\Big) = 1 - P^{(N)}\Big(S^{(N)}(k) < B\ \ \forall k \in \{1,\ldots,N\}\Big).$$

Hence, the binomial price of a knock-in option can be obtained as the difference of the binomial price for the corresponding path-independent option and the corresponding knock-out option; i.e.

$$E^{(N)}\Big(g(S_N^{(N)})1_{\{\exists k_0\in\{1,\ldots,N\}:\, S^{(N)}(k_0)\geq B\}}\Big) = E^{(N)}\Big(g(S_N^{(N)})\Big) - E^{(N)}\Big(g(S_N^{(N)})1_{\{S^{(N)}(k)<B\ \ \forall k\in\{1,\ldots,N\}\}}\Big).$$

Using Algorithm 1 for the first term on the right-hand side of the above equality and Algorithm 2 for the second term leads to a total effort of $O(N^2)$.

Algorithm 2: Binomial tree for European options with a knock-out barrier

---

Input: stock price parameters, risk-neutral rate r, payoff function g,
barrier level $B \geq s_0$ ($B \leq s_0$) for up-and-out option (down-and-out)
Return: price estimate $= V[0] \times \exp(-r \times T)$

1. Forward Step

{remains unchanged}

2. Backward Induction

```
{step backwards through the tree checking whether the barrier is crossed}
for k = N − 1 to 0 do
   for l = 0 to k do
      {adjust the state array to the current time step}
            SN[l] := SN[l]/d
      {check whether the barrier level is crossed}
      if SN[l] ≥ B (SN[l] ≤ B) then {up-and-out (down-and-out)}
         {assign current option value}
               V[l] := 0
      else
               V[l] := (pu × V[l+1] + pd × V[l])
      end if
   end for
end for
```

We next show that Algorithm 1 can also be adapted to American valuation problems. In principle, this is due to the theorem on the Snell envelope of a discrete process. Compared to alternative techniques for numerical valuation of American options, the resulting tree algorithm is both conceptually easy and efficient, which explains the widespread use of the binomial approach to American option pricing in practical applications.

American Options For the valuation of American options, the backward induction step of Algorithm 1 has to be modified so as to allow for early exercise at each node of the tree. That is, the values assigned to the nodes in time layer $k < N$ are the realisations of the RV

$$V_k^{(N)} := \max\left\{E_{P^{(N)}}\left(e^{-rT/N}V_{k+1}^{(N)}\middle| S_k^{(N)}\right), g(S_k^{(N)})\right\} \tag{2.74}$$

with

$$V_N^{(N)} := g(S_N^{(N)}).$$

If $V_k^{(N)} = E_{P^{(N)}}(e^{-rT/N}V_{k+1}^{(N)}|S_k^{(N)})$ the option is not exercised, while $V_k^{(N)} = g(S_k^{(N)})$ corresponds to early exercise. The modified backward induction is justified by the Markov property of the binomial process $S^{(N)}$ together with the following proposition on the *Snell envelope* of a discrete process:

Proposition 23. *Let X_k, $k = 0,\ldots,N$, be an adapted process on a filtered probability space $(\Omega,\mathscr{F},\mathscr{F}_k,P)$ and assume that X_k is integrable for all $k = 0,\ldots,N$. Define an $\mathscr{F}(.)$-adapted process Z_k, $k = 0,\ldots,N$, by backward induction, letting*

$$Z_N := X_N; \qquad Z_k := \max\{X_k, E(Z_{k+1}|\mathscr{F}_k)\},\ k = 0,\ldots,N-1.$$

Then Z is an $\mathscr{F}(.)$-supermartingale with $Z_k \geq X_k$ a.s. for all $k = 0,\ldots,N$. Moreover,

$$\tau_* := \min\{k \in \{0,\ldots,N\}:\ Z_k = X_k\}$$

is a discrete $\mathscr{F}(.)$-stopping time such that $Z_{.\wedge\tau_*}$ is an $\mathscr{F}(.)$-martingale. In particular, τ_* solves the optimal stopping problem for the process X; i.e.

$$E(X(\tau_*)) = \sup_{\tau\in\Sigma_{0,N}} E(X(\tau)),$$

where $\Sigma_{0,N}$ is the class of $\mathscr{F}(.)$-stopping times taking values in $\{0,\ldots,N\}$ (compare e.g. [CRS71], Theorem 3.2).

In our context, we choose $X_k := e^{(T-t_k)r}g(S_k^{(N)})$, $k = 0,\ldots,N$, with $t_k = kT/N$, which implies that $V_k^{(N)} = e^{-(T-t_k)r}Z_k$. As $V_0^{(N)} = e^{-rT}Z_0$, it follows that with $t_\tau = \tau T/N$,

$$V_0^{(N)} = e^{-rT}\sup_{\tau\in\Sigma_{0,N}} E_{P^{(N)}}\left(e^{(T-t_\tau)r}g(S_\tau^{(N)})\right) = \sup_{\tau\in\Sigma_{0,N}} E_{P^{(N)}}\left(e^{-t_\tau r}g(S_\tau^{(N)})\right).$$

This argument justifies the modified backward induction suggested in (2.74). The tree procedure is attached in pseudo-code (compare Algorithm 3).

Algorithm 3: Binomial tree for American options

Input: stock price parameters, risk-neutral rate r, payoff function g
Return: price estimate $= V[0]$

Pre-step

{incorporate the discount factor into transition probabilities}
Set $dpu := pu \times \exp(-r \times T/N)$
Set $dpd := pd \times \exp(-r \times T/N)$

1. Forward Step

{remains unchanged}

2. Backward Induction

{step backwards through the tree applying the early exercise condition}
for $k = N-1$ to 0 **do**
  for $l = 0$ to k **do**
    {adjust the state array to the current time step}
      $\mathrm{SN}[l] := \mathrm{SN}[l]/d$
    {assign current option value}
      $V[l] := (dpu \times V[l+1] + dpd \times V[l])$
    {apply the early exercise condition}
      $V[l] := \max(V[l], g(\mathrm{SN}[l]))$
  end for
end for

Let us add some remarks concerning implementation: We see that binomial pricing of American options results in a tree algorithm based on a backward induction step that requires computational effort of order $O(N^2)$. As for barrier options, the array of transition states needs to be adjusted to the current time layer of the tree. However, discounting must not be delayed to the end of tree procedure; rather, the option value has to be discounted while stepping backwards through the tree. In order to improve efficiency, the transition probabilities can be pre-multiplied by the one-period discount factor $\exp(-r \times T/N)$. This saves one multiplication at every node of the tree.

For American knock-out options, we can combine Algorithm 2 and Algorithm 3 (compare Algorithm 4). Let us illustrate this with an up-and-out option. The values assigned to the nodes in time layer $k < N$ are the realisations of the RV

$$V_k^{(N)} := \max\left\{E_{P^{(N)}}\left(e^{-rT/N} V_{k+1}^{(N)} \middle| S_k^{(N)}\right), g(S_k^{(N)})\right\} 1_{\{S_k^{(N)} < B\}} \tag{2.75}$$

with

$$V_N^{(N)} := g(S_N^{(N)}) 1_{\{S_N^{(N)} < B\}}.$$

The suggested backward induction (2.75) can again be justified by Proposition 23 on the Snell envelope. Here we choose

$$X_k := e^{(T-t_k)r} g(S_k^{(N)}) 1_{\{S_l^{(N)} < B \, \forall l=1,\ldots,k\}}, \qquad \text{for } k = 0, \ldots, N,$$

which implies that

$$V_k^{(N)} = e^{-(T-t_k)r} Z_k \qquad \text{on } \{S_l^{(N)} < B, \, \forall l = 1, \ldots, k-1\},$$

where, as before, $t_k = kT/N$. Trivially, $V_0^{(N)} = e^{-rT} Z_0$, which implies that the backward induction (2.75) leads to

$$V_0^{(N)} = e^{-rT} Z_0 = \sup_{\tau \in \Sigma_{0,N}} E_{P^{(N)}}\left(e^{-rt_\tau} g(S_\tau^{(N)}) 1_{\{S_l^{(N)} < B \, \forall l=1,\ldots,\tau\}}\right).$$

Algorithm 4: Binomial tree for American knock-out options

Input: stock price parameters, risk-neutral rate r, payoff function g
Return: price estimate $= V[0]$

Pre-step

{remains unchanged (as in Algorithm 3)}

1. Forward Step

{remains unchanged}

2. Backward Induction

{step backwards through the tree checking whether the barrier is crossed and applying the early exercise condition}
for $k = N-1$ to 0 **do**
  for $l = 0$ to k **do**
    {adjust the state array to the current time step}
      $\mathrm{SN}[l] := \mathrm{SN}[l]/d$
    {check whether the barrier level is crossed}
    if $\mathrm{SN}[l] \geq B$ ($\mathrm{SN}[l] \leq B$) **then** {up-and-out (down-and-out)}
      $V[l] := 0$
    else
      $V[l] := (dpu \times V[l+1] + dpd \times V[l])$
     {apply the early exercise condition}
      $V[l] := \max(V[l], g(\mathrm{SN}[l]))$
    end if
  end for
end for

Binomial valuation of American knock-in options is more involved than for European-type options because the sum of the American knock-in and knock-out price does not equal the price of the corresponding standard American option. This is due to the fact that the options differ in the optimal exercise time. As a consequence, for

American options we can no longer value a knock-in option by the difference between the standard option price and the knock-out price. To apply the binomial method to American knock-in options, AitSahlia, Imhof and Lai (2004) suggest an approach that involves the first passage density of a Brownian motion, which is known explicitly. They use the following representation of the price V_{in} of an American knock-in put option:

$$V_{\text{in}} = \int_0^T e^{-rt} V(t,B)\, f_S(t)\, \mathrm{d}t,$$

where $V(t,B)$ is the time-t price of the corresponding standard American option if the time-t stock price is $S_t = B$. Further, f_S is the first passage density of the stock price S associated with the barrier level B. Approximating the integral by a Riemann sum over M equally spaced time intervals $[t_k, t_{k+1}]$ leads to

$$V_{\text{in}} \approx T/M \sum_{k=1}^{M} e^{-rt_k^{(M)}} V(t_k^{(M)}, B)\, f_S(t_k^{(M)}), \tag{2.76}$$

where $t_k^{(M)} = kT/M$. Consequently, the price of an American knock-in option can be approximated by computing the binomial prices corresponding to $V(t_1^{(M)}, B), \ldots, V(t_M^{(M)}, B)$. This requires M calls to Algorithm 3. However, by an appropriate choice of the grid, i.e. an appropriate choice of the binomial process $S^{(N)}$, computational effort can be reduced to a single run of an appropriate backward induction algorithm. For details we refer to [AIL04].

In the above, we have discussed how to adapt the binomial tree algorithm to the valuation of specific types of options. Next we focus on numerical valuation of path-independent options. We wish to demonstrate an important conceptual property of the tree algorithm: its connection to explicit finite difference methods (for short: explicit FDMs).

2.4.2 Connection to Explicit Finite Difference Methods

This section deals with the connection between tree algorithms and explicit finite difference schemes for numerical valuation of path-independent options, which was first observed by Brennan and Schwartz (1978) [BS78]. Of course, the above relation can be anticipated from the fact that in the Black-Scholes setting, the martingale approach is naturally linked to the PDE approach via the Feynman-Kac Theorem. Hence, we first wish to recall some main results on the connection between these two approaches to option pricing in a Black-Scholes market. Afterwards, we transfer results to binomial option pricing.

The martingale approach and the PDE approach We consider a path-independent option with payoff[5] $g:(0,\infty)\to[0,\infty)$. Let us define $V:[0,T]\times(0,\infty)\to\mathbb{R}$ by

$$V(t,S) := E_Q^{t,S}\left(e^{-r(T-t)}g(S_T)\right) = E_Q\left(e^{-r(T-t)}g(S_T)\,\middle|\,S_t=S\right)$$
$$= E_Q\left(e^{-r(T-t)}g\left(Se^{\left(r-1/2\sigma^2\right)(T-t)+\sigma W_{T-t}}\right)\right). \quad (2.77)$$

Then $V(t,S)$ is the time t-price of the option if the stock trades at S at time t. As a fundamental result in mathematical finance, $V(t,S)$ is the solution to an appropriate Cauchy problem [6]:

Proposition 24. *Assume that g is polynomially bounded, i.e.*

$$g(S)\leq C\left(1+S^{\beta}\right) \qquad \textit{for all } S\in(0,\infty),$$

where C, β are positive constants.
Then $V(t,S)\in C^{\infty,\infty}([0,T)\times(0,\infty))\cap C([0,T]\times(0,\infty))$ and $V(t,S)$ is the unique polynomially bounded solution to the Cauchy problem

$$\begin{aligned} V_t+\tfrac{1}{2}\sigma^2S^2V_{SS}+rSV_S-rV &= 0, \qquad (t,S)\in[0,T)\times(0,\infty) \\ V(T,S) &= g(S) \qquad\quad S\in(0,\infty). \end{aligned} \quad (2.78)$$

For completeness, we give a proof of Proposition 24 below. We will refer to the following result:

Lemma 5. *Let X and Y be independent RVs. Let φ be a function with $E\,|\,\varphi(X,Y)|<\infty$ and let $g(x):=E(\varphi(x,Y))$. Then*

$$E(\varphi(X,Y)\,|\,X)=g(X)$$

(compare e.g. [Dur05], Example 4.1.5).

---

[5] Note that in contrast to previous notations, the domain of g has been changed to $(0,\infty)$ as we consider only path-independent options.

[6] In this thesis, we assume the martingale representation of the option price to be given (compare (2.77)) and we derive a pricing PDE (compare (2.78)) from the stochastic representation of the price. Let us emphasise that in the history of option pricing, the two approaches appeared in opposite order. In the groundbreaking work of Black and Scholes and Merton (1973), the risk-neutral valuation principle was introduced. Based on this principle, they identified the valuation problem with a Cauchy problem. Their ansatz is often referred to as the ”delta-hedging approach”. The ”martingale approach” to option pricing was later suggested by Harrison and Kreps (1979) and Harrison and Pliska (1981) who showed that option pricing is naturally linked with martingale theory [HK79], [HP81]. Since then, the martingale approach has played a dominating role.

Proof of Proposition 24. Note first that

$$V(t,S) \to g(S) \qquad \text{as } t \uparrow T.$$

As we will see, smoothness of the option price directly follows from smoothness of the Gauss kernel. In order to prove existence of the partial derivatives with respect to S, we hence consider the following representation for $V(t,S)$:

$$V(t,S) = \frac{1}{\sqrt{2\pi}} e^{-r(T-t)} \int_{\mathbb{R}} e^{-\frac{1}{2}\left(y - \frac{\ln(S)}{\sigma\sqrt{T-t}}\right)^2} g\left(e^{\left(r-1/2\sigma^2\right)(T-t)+\sigma\sqrt{T-t}y}\right) \mathrm{d}y.$$

As g is polynomially bounded, it follows from the Differentiation Lemma (compare e.g. [Bau92], Lemma 16.2) that for any $p \in \mathbb{N}$,

$$\begin{aligned}\left|\frac{\partial^p V}{\partial S^p}(t,S)\right| &= \frac{1}{\sqrt{2\pi}} e^{-r(T-t)} \int_{\mathbb{R}} \left|\frac{\partial^p}{\partial S^p}\left(e^{-\frac{1}{2}\left(y - \frac{\ln(S)}{\sigma\sqrt{T-t}}\right)^2}\right)\right| g\left(e^{\left(r-1/2\sigma^2\right)(T-t)+\sigma\sqrt{T-t}y}\right) \mathrm{d}y \\ &\leq C \int_{\mathbb{R}} e^{-\frac{1}{2}\left(y - \frac{\ln(S)}{\sigma\sqrt{T-t}}\right)^2} \left|P_p(S,y)\right| \left(1 + e^{\beta\sigma\sqrt{T-t}y}\right) \mathrm{d}y\end{aligned} \tag{2.79}$$

for some positive constants C and β. For S fixed, $P_p(S,.)$ is a polynomial. As the kernel decreases essentially by e^{-y^2}, the integral on the right-hand side of inequality (2.79) exists, which shows that $V(t,S)$ is infinitely often differentiable with respect to S. Similarly, we can use the representation

$$V(t,S) = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}\sigma\sqrt{T-t}} e^{-r(T-t)} e^{-\frac{\left(y-\left(r-1/2\sigma^2\right)(T-t)\right)^2}{2\sigma^2(T-t)}} g\left(Se^y\right) \mathrm{d}y$$

in order to show that $V(t,S)$ is infinitely often differentiable with respect to t for $t < T$. Thus, $V(t,S) \in C^{\infty,\infty}([0,T) \times (0,\infty)) \cap C([0,T] \times (0,\infty))$. Let us now show that $V(t,S)$ solves the Cauchy problem (2.78). Note first that the terminal condition is trivially satisfied. For the dynamics, we use standard tools on stochastic processes: Let us define the process

$$M_t := E_Q\left(g\left(S_T\right) \middle| \mathscr{F}_t^S\right)$$

where $\mathscr{F}_t^S := \sigma(S_s; 0 \le s \le t)$. Then by the Markov property of S,

$$M_t = E_Q\left(g\left(S_t e^{\left(r-1/2\sigma^2\right)(T-t)+\sigma(W_T-W_t)} \right) \middle| S_t \right).$$

Applying Lemma 5 with $X = S_t$, $Y = W_T - W_t$ and

$$\varphi(S_t, W_T - W_t) = g\left(S_t e^{\left(r-1/2\sigma^2\right)(T-t)+\sigma(W_T-W_t)} \right)$$

shows that $M_t = e^{r(T-t)} V(t, S_t)$. As $V(t,S) \in C^{1,2}$, Ito's Formula leads to the following dynamics for M_t:

$$\begin{aligned} \mathrm{d}M_t = e^{r(T-t)} \left(V_t(t,S_t) + \tfrac{1}{2}\sigma^2 S_t^2 V_{SS}(t,S_t) + rS_t V_S(t,S_t) - rV(t,S_t) \right) \mathrm{d}t \\ + e^{r(T-t)} \sigma S_t V_S(t,S_t)\, \mathrm{d}W_t \end{aligned} \qquad (2.80)$$

Define

$$h(t) := V_t(t,S_t) + \tfrac{1}{2}\sigma^2 S_t^2 V_{SS}(t,S_t) + rS_t V_S(t,S_t) - rV(t,S_t) \qquad 0 \le t < T.$$

Recalling that every continuous local martingale of finite variation is a.s. constant (compare e.g. [Kal01], Proposition 17.2, p. 330), we see from (2.80) that

$$\int_0^t h(s)\mathrm{d}s = 0 \qquad a.s. \qquad \text{for all } t \in [0,T).$$

As h is continuous, applying the Fundamental Theorem of Calculus shows that

$$h \equiv 0 \qquad a.s.$$

Now it follows again by continuity of h and the fact that the distribution of S_t, $0 \le t < T$, has support $(0,\infty)$ that

$$V_t + \tfrac{1}{2}\sigma^2 S^2 V_{SS} + rSV_S - rV = 0, \qquad (t,S) \in [0,T) \times (0,\infty).$$

It remains to show uniqueness: Since g is assumed to be polynomially bounded, we see from the definition of V that it is polynomially bounded with respect to S uniformly in $t \in [0,T]$, i.e.

$$\max_{0 \le t \le T} |V(t,S)| \le C\left(1 + S^\beta\right) \qquad \text{for all } S \in (0,\infty) \qquad (2.81)$$

for some positive constants C and β. Consequently, it follows from the Feynman-Kac Theorem (compare e.g. [KS98], Theorem 5.7.6) that $E_Q^{t,S}\left(e^{-r(T-t)}g(S_T)\right)$ is the unique solution to the PDE (2.78) within the class of functions that satisfy the polynomial growth condition (2.81). □

We now present the implications of the above result to binomial option pricing. As shown e.g. in Heston and Zhou (2000), there exists a suitable transformation of variables so that the tree algorithm associated with the RB model can be identified with an explicit finite difference approximation to the transformed Cauchy problem [HZ00]. We follow the approach of Heston and Zhou, but we keep notations general so that the result additionally covers the CRR model. Clearly, the variable transform then depends on the particular choice of the drift α.

As suggested by the form of the binomial schemes, we use the transformation of variables

$$S = s_0 e^{\alpha t + \sigma x}$$

with $\alpha = r - 1/2\sigma^2$ for the RB scheme and $\alpha = 0$ for the CRR scheme. Further, we introduce the functions

$$u(t,x) := e^{r(T-t)} V(t,S) \tag{2.82}$$

and

$$\tilde{g}(x) := g\left(s_0 e^{\alpha T + \sigma x}\right). \tag{2.83}$$

If the payoff function g is polynomially bounded, it follows from Proposition 24 that u is of class $C^{\infty,\infty}([0,T) \times (0,\infty)) \cap C([0,T] \times (0,\infty))$. Moreover, it is straightforward to verify that u solves the transformed Cauchy problem[7]

$$\begin{aligned} u_t + \tfrac{r - 1/2\sigma^2 - \alpha}{\sigma} u_x + \tfrac{1}{2} u_{xx} &= 0 \qquad (t,x) \in [0,T) \times (-\infty,\infty) \\ u(T,x) &= \tilde{g}(x) \qquad x \in (-\infty,\infty). \end{aligned} \tag{2.84}$$

By means of the transformed Cauchy problem (2.84), we can now link the binomial tree algorithm to an explicit FDM.

The binomial approach and the explicit FDM The link between the martingale approach and the PDE approach leads to the following result on numerical option pricing

[7] In the original paper by Heston and Zhou, the choice $\alpha = r - 1/2\sigma^2$ results in the heat equation.

techniques:

Proposition 25. *Let $S^{(N)}$ be the binomial process suggested by RB or by CRR, respectively. Then the corresponding tree algorithm for the valuation of path-independent options can be identified with an explicit FDM that approximates the solution to the transformed Cauchy problem (2.84). Here we use finite differences of first order in t and second order in x.*

Proof. Note first that if we replace the derivatives in (2.84) by finite differences of first order in t and second order in x, and neglect the error terms, we obtain the approximation

$$-\left(\frac{u(t,x)-u(t-\Delta t,x)}{\Delta t}\right) \approx \frac{r-1/2\sigma^2-\alpha}{\sigma}\left(\frac{u(t,x+\Delta x)-u(t,x-\Delta x)}{2\Delta x}\right) + \frac{1}{2}\left(\frac{u(t,x+\Delta x)-2u(t,x)+u(t,x-\Delta x)}{(\Delta x)^2}\right), \quad (2.85)$$

where $\Delta t > 0$ and $\Delta x > 0$ are the increments in the time and in the space domain, respectively. In order to identify the binomial method with an explicit FDM based on the approximation above, we have to fix the following grid:

$$(t_j, x_i) := (j\,\Delta t, i\,\Delta x) \qquad j = 0,\ldots,N; \qquad i = -j, -j+2, \ldots, j$$

with grid size

$$\Delta t = (\Delta x)^2 = T/N. \quad (2.86)$$

Then (2.85) leads to an explicit FDM that approximates the theoretical solution to the Cauchy problem (2.84) at the grid points specified above. That is, starting from the terminal values

$$\hat{u}_{N,i} = \tilde{g}\left(i\sqrt{\tfrac{T}{N}}\right), \qquad \text{for } i = -N, -(N-2), \ldots, N,$$

we assign an appropriate value to each grid point by the following backward recursion: For all $j = N, \ldots, 1$ and for all $i = -j+1, -j+3, \ldots, j-1$, we set

$$\hat{u}_{j-1,i} = p_1 \hat{u}_{j,i+1} + p_2 \hat{u}_{j,i-1}, \quad (2.87)$$

where

$$p_1(N) := \tfrac{1}{2} + \tfrac{1}{2\sigma}\left(r - \tfrac{1}{2}\sigma^2 - \alpha\right)\sqrt{\tfrac{T}{N}}$$
$$p_2(N) := \tfrac{1}{2} - \tfrac{1}{2\sigma}\left(r - \tfrac{1}{2}\sigma^2 - \alpha\right)\sqrt{\tfrac{T}{N}}.$$

Here $\hat{u}_{j,i}$ approximates the theoretical solution u at the grid point (t_j, x_i) In particular, $\hat{V}_{0,0} := e^{-rT}\hat{u}_{0,0}$ is the finite difference approximation to the option price $V(0, s_0)$. As we note by inspection, the backward recursion (2.87) obtained from finite differences is identical to that specified by the binomial algorithm. In particular, we have $p_u(N) = p_1(N)$ and $p_d(N) = p_2(N)$. □

Remark 16. *If we consider the above binomial schemes under the risk-neutral measure, there is only an asymptotic equivalence between the backward recursion (2.87) and the backward recursion specified by the binomial method in the sense that*

$$q_u(N) = p_1(N) + O\left(\frac{1}{N^{3/2}}\right) \qquad q_d(N) = p_2(N) + O\left(\frac{1}{N^{3/2}}\right)$$

(compare Proposition 3).

To conclude, the tree procedure suggested by CRR or by RB can be identified with an explicit finite difference approximation to the Cauchy problem (2.84). The corresponding explicit finite difference scheme is special in the sense that *it incorporates distributional information that specifies the underlying grid*. By contrast, for a pure PDE approach, it is necessary to specify the underlying grid exogenously, i.e. one has to decide how to truncate the S domain, how to choose grid points, how to relate the grid sizes in the time and in the space domain, etc. For the binomial tree algorithm, these issues are settled endogenously by definition of the discrete model, which is of course advantageous for practical applications.

2.5 The Convergence Behaviour of Binomial Option Prices

This section deals with the convergence behaviour of binomial option prices for the different discretisation schemes considered previously. As we will demonstrate both theoretically and by many numerical examples, the optimal drift model shows superior performance compared to the alternative methods.

Let us first stress that the convergence behaviour of binomial option prices depends crucially on the valuation problem under consideration, and in particular on the specific payoff function. Hence, we analyse the convergence behaviour of binomial trees separately for the two most common payoff structures; first, for payoffs that are constant in the terminal value $S(T)$ (i.e. cash-or-nothing options); and second, for payoffs that are linear in $S(T)$ (i.e. plain vanilla options). For the European case, these simple types of options admit an explicit pricing formula. Hence, in practical applications numerical option pricing would not be applied to these options. However, their simple payoff structure allows to derive an asymptotic expansion of the pricing error around the Black-Scholes value, so that the convergence behaviour of the corresponding binomial prices can be analysed theoretically. Importantly, *the convergence behaviour observed often carries over to related, but more complex types of options for which an explicit pricing formula is not known.* Hence regarding practical applications, one can benefit a lot from analysing these simple payoff structures. This is illustrated for American plain vanilla puts. In this case, the convergence behaviour of binomial prices can be (partially) anticipated from the asymptotic expansion of the pricing error in its European counterpart.

As for any discrete model, the rate of convergence is a central property of the convergence behaviour of binomial option prices. It measures the (asymptotic) speed/accuracy trade-off of a numerical method. Regarding practical applications, it is important to know the order of convergence [HZ00]: Firstly, the rate of convergence helps to rank competing numerical methods. Almost any method can give fast inaccurate results, and, given enough computing time, many methods can give arbitrarily accurate results. Knowing the order of convergence helps to decide which of the competing models should be preferred. Secondly, the rate of convergence indicates whether extrapolation is useful. That is, extrapolation techniques can increase accuracy, but they are only applicable provided convergence is smooth; i.e. if the rate of convergence is known, if the coefficient of the leading error term is a fixed constant and if oscillations of higher order terms are known to be negligible. Consequently, our analysis on the convergence behaviour of binomial prices is mainly focused on the rate of convergence achieved for the different schemes.

For cash-or-nothing options, the convergence behaviour of binomial prices can be deduced directly from our results on the fit of the binomial distribution to the lognormal distribution (compare section 2.2.2). Hence, the binomial prices obtained from conventional schemes converge in general no faster than $1/\sqrt{N}$, as suggested by the Berry-Esséen inequality (compare Corollary 10). For plain vanilla options, the payoff function exhibits a kink in the terminal stock price, i.e. a discontinuity in the first derivative. As shown by Diener and Diener and by Chang and Palmer, this payoff structure leads to cancellation effects in the asymptotic expansion of the pricing error. We will see that in this case conventional schemes admit convergence of order $1/N$, which is in particular above the Berry-Esséen bound.

As mentioned above, the order of convergence can be increased by extrapolation provided convergence is smooth. However, as discussed in Section 2.2.2, *the discretisation error in conventional binomial schemes converges non-smoothly. As a consequence, for non-smooth payoff functionals, the convergence behaviour of the corresponding sequence of binomial option prices is also oscillatory and non-monotone.* Thus, there is low-frequency shrinking according to the rate of convergence, but in addition, there are high-frequency oscillations. In principle, the presence of oscillations can be traced back to the fact that when grid size changes, the position of nodes in the tree varies in relation to some fixed discontinuity or kink in the payoff. A prominent example is the so-called *sawtooth effect* which was first observed for barrier options by Boyle and Lau (1994) [BL94]. In the following, we analyse the non-monotone convergence behaviour of conventional schemes both theoretically and in practical examples.

The irregular convergence behaviour of conventional schemes is a serious issue for their application to option valuation. Obvious problems caused by oscillations are the facts that a finer discretisation does not necessarily provide a better estimate, and that the option price cannot be obtained by extrapolation methods.

There are many advanced schemes in the literature that are especially designed and optimised to a specific valuation problem in order to improve the convergence behaviour of the pricing error. In Section 2.2.3, we presented advanced binomial schemes that can be adapted to a given point x, so that the discretisation error around x exhibits a superior asymptotic behaviour. That is, they establish smooth convergence or they even increase the rate of convergence. This suggests that these schemes are advantageous for the practical application of binomial option valuation. In the following, this is discussed in detail. We will see that the Tian model and the Chang and Palmer model can be adapted to the strike value of interest so as to improve the convergence behaviour for both cash-or-nothing options and plain vanilla options. In particular, extrapolation methods can be applied to increase the order of convergence. However, the optimal drift model we suggest can admit convergence of order $o(1/N)$ without extrapolation for both types of options. We will demonstrate that *by virtue of its superior rate of convergence, the optimal drift model is advantageous compared to the conventional schemes and to the advanced schemes presented*.

2.5.1 Constant Payoff Structures (Cash-or-Nothing Options)

This paragraph deals with the convergence behaviour of binomial prices for cash-or-nothing options, i.e. options that pay a constant amount of money. In this case, the convergence behaviour of binomial prices is only affected by the discontinuity in the payoff function.

A cash-or-nothing call (put) with strike $K > 0$ pays a cash amount $G > 0$ if the terminal value lies above K (below K); i.e.

$$g(S) = G\,1_{\{S_T \geq K\}} \qquad (g(S) = G\,1_{\{S_T < K\}}).$$

With $V^{\text{cash}}(K)$ denoting the Black-Scholes price and $V^{(N),\text{cash}}(K)$ denoting the binomial price, we have

$$\left|V^{\text{cash}}(K) - V^{(N),\text{cash}}(K)\right| = Ge^{-rT}\left|P^{(N)}\left(S_N^{(N)} \geq K\right) - Q\left(S_T \geq K\right)\right|.$$

Consequently, the pricing error is readily available from the distributional fit of the binomial model to a Brownian motion (compare Section 2.2.2). In particular, according to the Berry-Esséen inequality, the sequence of price estimates converges in order $1/\sqrt{N}$

(compare Corollary 4):

Proposition 26. *Let $S^{(N)}$ be the process suggested by CRR, the process suggested by RB or any binomial process (2.2) with $\beta = \sigma$ and with risk-neutral transition probabilities. Then*

$$\sup_K \left| V^{cash}(K) - V^{(N),cash}(K) \right| = O\left(\frac{1}{\sqrt{N}}\right) \tag{2.88}$$

Remark 17. *Note that if the payoff function is piecewise constant with a finite number of discontinuities, the corresponding binomial prices converge also in order $1/\sqrt{N}$.*

We first discuss the convergence behaviour of conventional schemes; i.e. binomial schemes with constant drift α. This will be illustrated with the schemes suggested by RB and by CRR. For conventional schemes, the Berry-Esséen inequality is tight. Furthermore, these schemes suffer heavily from an irregular convergence behaviour. This will be discussed in detail. Afterwards, we demonstrate the superior convergence behaviour of the advanced methods discussed in Section 2.2.3.

Conventional Binomial Models

According to Proposition 7, the Berry-Esséen bound (2.88) is tight for the conventional models; to be precise, we have

Proposition 27. *Let $S^{(N)}$ be the process suggested by CRR, the process suggested by RB or any binomial process (2.2) with $\beta = \sigma$, $\alpha(N) \equiv \alpha$ constant in N and with risk-neutral transition probabilities. Then with $C^{cash}(K)$ ($P^{cash}(K)$) denoting the Black-Scholes price for the cash-or-noting call (put) and $C^{(N),cash}(K)$ ($P^{(N),cash}(K)$) denoting the corresponding binomial price,*

$$C^{(N),cash}(K) = C^{cash}(K) + Ge^{-rT}\frac{e^{-\frac{1}{2}d_2^2}}{\sqrt{2\pi T}} b(N)\left(\frac{T}{N}\right)^{1/2} + O\left(\frac{1}{N}\right)$$

and

$$P^{(N),cash}(K) = P^{cash}(K) - Ge^{-rT}\frac{e^{-\frac{1}{2}d_2^2}}{\sqrt{2\pi T}} b(N)\left(\frac{T}{N}\right)^{1/2} + O\left(\frac{1}{N}\right)$$

with

$$b(N) = 1 - 2\{-a(N)\} = 1 - 2\{-\tfrac{1}{2}N + \frac{\ln(s_0/K) + \alpha T}{2\sigma\sqrt{T}}\sqrt{N}\}, \tag{2.89}$$

where, as before, $\{.\}$ denotes the fractional part.

Remark 18. *Note that the point x at which the binomial distribution function is evaluated throughout Section 2.2.2 and Section 2.2.3 is now interpreted as the strike value of the option. In particular, the advanced schemes will later be optimised according to strike value of the specific valuation problem.*

According to our previous analysis, we see from Proposition 27 that convergence is non-smooth because the oscillating factor $b(N)$ enters the coefficient of the leading error term. As explained above, this has undesirable consequences for practical applications. We will see next that the irregular convergence behaviour for cash-or-nothing options suffers from two main effects; *the sawtooth effect* and *the even-odd problem*.

Before, we wish to add that as $-1 < b(N) \leq 1$, the oscillations of the leading error term are bounded by

$$-Ge^{-rT}\varphi(d_2)\tfrac{1}{\sqrt{N}} < c_1(N)\tfrac{1}{\sqrt{N}} \leq Ge^{-rT}\varphi(d_2)\tfrac{1}{\sqrt{N}}, \tag{2.90}$$

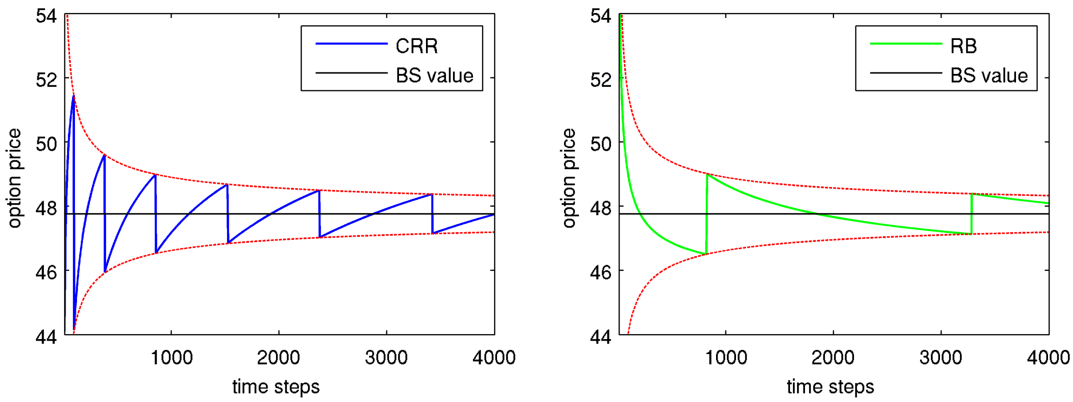
where $\varphi(.)$ is the lognormal density (compare Corollary 5). We can hence formulate the following result on the amplitude of the oscillations:

Corollary 11. *The smaller $e^{-rT}\varphi(d_2)$, the tighter the bounds on the leading error term, i.e. the smaller the amplitude of the oscillations.*

The sawtooth effect In this section, we analyse the convergence behaviour of binomial prices along values of N that are of the same parity. The general case is analysed in the next paragraph on the even-odd problem.

Figure 2.2 shows the binomial prices for a cash-or-nothing call obtained from the CRR tree and from the RB tree, respectively. The estimates are computed for even values of N; that is, $N = 10 : 2 : 4000$. The dashed red lines indicate the bounds on the leading error term observed in (2.90).

Fig. 2.2: Convergence pattern for a cash-or-nothing call (N even)
$s_0 = 95$, $\sigma = 0.25$, $r = 0.1$, $T = 1$, $G = 100$, strike $K = 100$



As we see in Figure 2.2, the binomial prices are neither consistently greater nor less than the Black-Scholes price; rather, they alternate between over- and underestimation with some highly accurate values in between. The behaviour observed is called the *sawtooth effect*. Let us remark that we will see a similar pattern if we restrict the sequence of price estimates to odd values of N. We next explain this pattern by the asymptotics of the pricing error derived in Propositition 27. Further, these rigorous results are illustrated with intuitive arguments on the specific allocation of probability mass in the binomial model under consideration; i.e. the location of terminal nodes in the corresponding tree.

For any number of periods N, there is some integer $l(N)$ such that the strike value K falls between the terminal node corresponding to $l(N)-1$ up-movements and that corresponding to $l(N)$ up-movements; i.e.

$$s_N^{(N)}(l-1) = s_0 u^{l-1}(N) d^{N-l+1}(N) < K \leq s_0 u^l(N) d^{N-l}(N) = s_N^{(N)}(l).$$

Assume first that the strike value K is close to $s_N^{(N)}(l)$; that is, there is a terminal node in the corresponding binomial tree which is just above the strike. Intuitively, in this case the probability to end up in the money is expected to be too high compared to the continuous-time model, which results in an overestimation of the exact price. In fact, the above argument can be verified from the asymptotics of the discretisation error because if K is close to $s_N^{(N)}(l)$, we have that $b(N) \approx 1$ (compare (2.56)). Consequently, the leading error term is close to its upper bound $Ge^{-rT}\varphi(d_2)/\sqrt{N}$. By contrast, in case that K is close to $s_N^{(N)}(l-1)$, we have that $b(N) \approx -1$, so that the leading error term is close to its lower bound $-Ge^{-rT}\varphi(d_2)/\sqrt{N}$. Hence, in the latter case the asymptotic expansion of the discretisation error indicates that the binomial model underestimates the exact price. This, too, is coherent with intuition: For K close to $s_N^{(N)}(l-1)$, we anticipate that the likelihood to end up in the money is too low compared to the continuous-time situation, which then implies that the option price is underestimated. Let us finally consider the case $K = (s_N^{(N)}(l-1)s_N^{(N)}(l))^{1/2}$. Then according to (2.56), the leading term in the asymptotic expansion of the pricing error is equal to zero. Hence, this situation leads to a highly accurate price estimate. In this case, the strike value falls on the geometric average between the two neighbouring nodes.

In the above, we have considered the number of periods N to be fixed. *If we now assume that the step size changes, the position of nodes in tree varies in relation to the strike value. This leads to the observed oscillations in the convergence pattern.* In particular, the above situations occur as the three extreme cases; that is, the price estimate touches the upper bound, it is highly accurate or it touches the lower bound.

To further illustrate the above effect, let us define m as the effective number of up-movements, i.e.

$$m = \sharp\,\text{up-movements} - \sharp\,\text{down-movements}.$$

In our example, the CRR model with $N = 94$ periods has a terminal node at 100.028. Hence, the strike value ($K = 100$) falls just below that node. The node corresponds to the paths with an effective number of up-movements given by $m = 2$ ($l = 48$). If the number of periods is increased to $N = 96$, the distance between adjacent possible realisations shrinks. Here a path with $m = 2$ ($l = 49$) effective up-movements ends at 99.9738, which is now below the strike value K. Consequently, the location of nodes is such that the risk-neutral probability to end up in the money (.5278) is overestimated for $N = 94$ (.5677), but it is underestimated for $N = 96$ (.4904). We therefore observe a sudden drop in the corresponding binomial prices from 47.7604 £ to 44.3732 £.

The specific oscillations observed in our example can be interpreted as follows: As the step size increases, the CRR model leads to price estimates that increase in absolute value until a sudden downward drop. This is again followed by an increase in absolute value in successive discretisation steps. By contrast, the price estimates obtained from the RB tree decrease in absolute value until an abrupt rise. Let us stress that *the specific form of the sawtooth pattern observed for our example is not generic; rather, it depends on the parameter setting*. This is easy to see: Note first that for even values of N, each terminal node corresponds to an even number of effective up-movements. Let m be the effective number of up-movements and suppose that $F(m) > 0$ is such that

$$\exp\left(m\sigma\sqrt{\tfrac{T}{F(m)}}\right) = \tfrac{K}{s_0 e^{\alpha T}}. \tag{2.91}$$

In our example, the input parameters are such that $K/s_0 > 1$, while $K/s_0 e^{\left(r-1/2\sigma^2\right)T} < 1$. Hence, we see from (2.91) that for the CRR model (i.e.

$\alpha = 0$), the strike always lies between two nodes that exhibit a *positive* effective number of up-movements. By contrast, for the RB model (i.e. $\alpha = r - 1/2\sigma^2$), the strike always lies between two nodes that exhibit a *negative* effective number of up-movements. The specific form of the sawtooth effect observed for our example can now be explained by the fact that while

$$e^{|m|\sigma\sqrt{T/N+2}} < e^{|m|\sigma\sqrt{T/N}},$$

we have

$$e^{-|m|\sigma\sqrt{T/N}} < e^{-|m|\sigma\sqrt{T/N+2}}.$$

Consequently, if N approaches $F(|m|)$ (along the sequence of even integers), $s_0 e^{|m|\sigma\sqrt{T/N}}$ approaches K from above. Yet, if N approaches $F(-|m|)$, $s_0 e^{\left(r-1/2\sigma^2\right)T} e^{-|m|\sigma\sqrt{T/N}}$ approaches K from below. This explains why in our example, the binomial prices obtained from the CRR model are piecewise increasing, while those obtained from the RB model are piecewise decreasing. If N rises above $F(|m|)$ or $F(-|m|)$, respectively, the sequence of price estimates faces a sudden downward drop or an abrupt rise. We observe from (2.91) that

$$F(-|m|) = F(|m|) = m^2 \left(\frac{\sigma\sqrt{T}}{\ln(K/s_0)-\alpha T}\right)^2.$$

This shows that in our example the oscillation frequency is higher for the CRR model than for RB model because $\sigma\sqrt{T}/|\ln(K/s_0)| \approx 4.9$, while $\sigma\sqrt{T}/|\ln(K/s_0) - (r - 1/2\sigma^2 T)| \approx 14.3$.

In general, the above arguments imply that the specific form of the sawtooth effect is determined according to the following result:

Proposition 28. *Suppose we limit the sequence of price estimates to values of the same parity (either even or odd). Then the leading error term is piecewise increasing if $s_0 e^{\alpha T} < K$ and piecewise decreasing for $s_0 e^{\alpha T} > K$. Moreover, the greater $\sigma\sqrt{T}/|\ln(K/s_0) - \alpha T|$, the lower the frequency of the oscillations.*

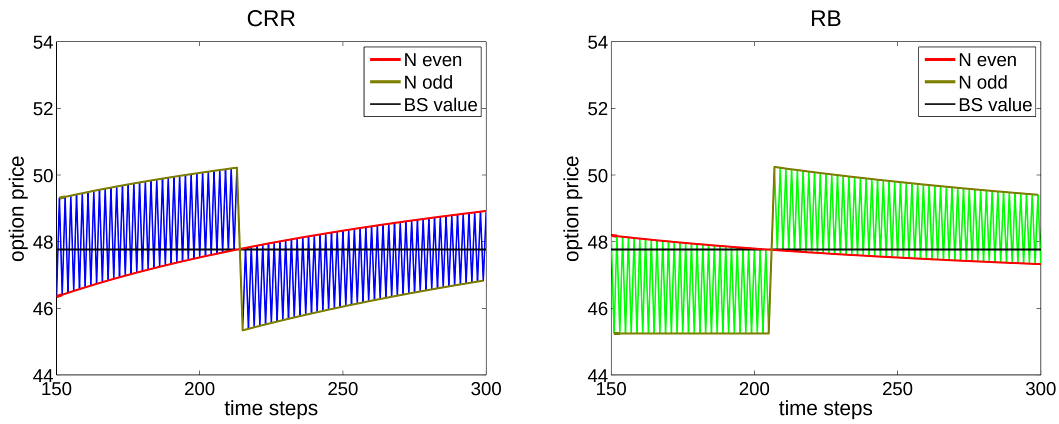
Remark 19. *Clearly, the above result can alternatively be deduced from the formula (2.89) for $b(N)$.*

Corollary 12. *If the stock price parameters are fixed, the frequency of oscillations depends on the strike value only. Then, the smaller the distance between K and $s_0 e^{\alpha T}$, the lower the frequency of the oscillations.*

We anticipate from Corollary 12 that the conventional schemes admit smooth convergence for the limit case $K = s_0 e^{\alpha T}$. This will be verified later. Before, we wish to demonstrate a second main effect of irregular convergence behaviour other than the sawtooth-effect: the even-odd effect.

The even-odd problem In the above, we have limited our analysis to values of N that are of the same parity. We next investigate the convergence behaviour of binomial prices along integers of alternating parity. We will see that in this case, the convergence pattern exhibits *micro oscillations* between even and odd values of N. This effect is often called the even-odd problem. The micro oscillations are superimposed on the *marco oscillations* investigated previously; i.e. they are superimposed on the sawtooth pattern along the even integers and the sawtooth pattern along the odd integers. This is illustrated for $N = 150:1:300$ in Figure 2.3.

Fig. 2.3: Convergence pattern for a cash-or-nothing call: Micro oscillations
$s_0 = 95$, $\sigma = 0.25$, $r = 0.1$, $T = 1$, $G = 100$, strike $K = 100$



The presence of micro oscillations can be deduced from the asymptotic expansion of the pricing error because $b(N)$ involves the fractional part of $-1/2N$ (compare (2.89)). To give an intuitive argument on the presence of micro oscillations, let us recall from (2.39) that the root of $s_0 u(N)^a d(N)^{N-a} = K$ is of the form

$$a(N) = \frac{1}{2}N + c\sqrt{N}$$

for some appropriate constant c. Hence for reasonable large values of N, the root $a(N)$ increases approximately according to $a(N+1) \approx a(N) + 1/2$. However, in the discrete model an increase by $1/2$ is not possible. As a consequence, the relative number of terminal nodes in the in-the-money region is approximately the same for N and $N+2$, while it is significantly different for $N+1$. This difference affects the probability mass assigned to the in-the-money region and hence causes the even-odd problem (compare Table 2.1).

Table 2.1: The even-odd problem for the RB tree
$s_0 = 95$, $\sigma = 0.25$, $r = 0.1$, $T = 1$, $G = 100$, strike $K = 100$

N	200	201	202	BS value
$a(N)$	99.5062	100.0050	100.5038	—
$b(N)$	0.0125	-0.9900	0.0075	—
$l(N)$	100	101	101	—
Nodes in the money/ total nodes	0.50249	0.50000	0.50246	—
Probability to end up in the money	0.5282	0.5000	0.5280	0.5278
Binomial price	47.7912	45.2419	47.7786	47.7604

It remains to consider the convergence behaviour of cash-or-nothing calls for the border case $K = s_0 e^{\alpha T}$. As discussed above, we expect that in this situation, the conventional schemes behave differently than in the general case. In particular, we anticipate that convergence is smooth. This is demonstrated next.

The border case $\mathbf{K = s_0 e^{\alpha T}}$ Note first that the border case associated with the CRR model occurs for the at-the-money situation; i.e. $K = s_0$ (for an analysis of this case compare also [DD04]). By contrast, the border case associated with the RB model corresponds to the situation $K = s_0 e^{(r-1/2\sigma)T}$.

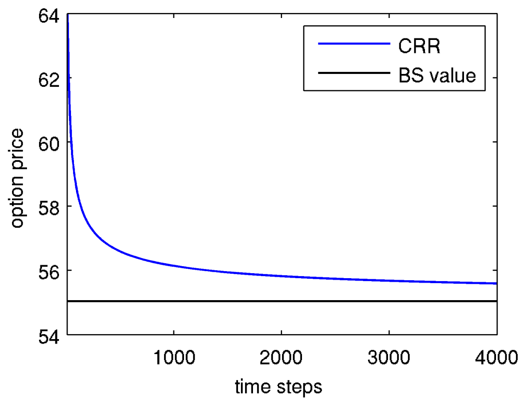
We first investigate the corresponding binomial prices along even integers. In this case, as we observe from the asymptotic expansion of the discretisation error, we have $b(N) = 1$, so that the coefficient of the leading error term always coincides with its upper bound. Consequently, the price estimates converge smoothly according to

$$C^{(N),\mathrm{cash}}(s_0 e^{\alpha T}) = C^{\mathrm{cash}}\left(s_0 e^{\alpha T}\right) + G e^{-rT} \frac{e^{-\frac{1}{2}d_2^2}}{\sqrt{2\pi T}} \left(\frac{T}{N}\right)^{1/2} + O\left(\frac{1}{N}\right), \qquad N \text{ even},$$

where d_2 is evaluated at $s_0 e^{\alpha T}$.

Smooth convergence is illustrated for the CRR tree in Figure 2.4. The plot shows the convergence pattern for an at-the-money cash-or-nothing call with $N = 10 : 2 : 4000$. Except for the strike value, the parameters are kept as before.

Fig. 2.4: Convergence pattern for an at-the-money cash-or-nothing call **(N even)**
$s_0 = 95$, $\sigma = 0.25$, $r = 0.1$, $T = 1$, $G = 100$, **strike K = 95**



As in the above, we wish to add an intuitive argument on smooth convergence in the border case by means of the specific allocation of probability mass. Since we limit our analysis to even values of N, it follows from symmetry that the terminal node $S^{(N)}(N/2)$ corresponding to zero effective up-movements (i.e. the centre of tree) coincides with $K = s_0 e^{\alpha T}$; that is, $K = S^{(N)}(N/2)$. As a result, the centred node always contributes to the probability mass assigned to the in-the-money region. Consequently, for any even number of periods N, the binomial price overestimates the exact price to the maximum extent. Let us emphasise that the benefits due to smooth convergence overcompensate the fact that the price estimates are at the maximal distance to the exact price. In particular, smooth convergence allows for extrapolation methods.

Let us now consider the border case for any odd value of N. Then the strike $K = s_0 e^{\alpha T}$ is again located at the centre of the tree; yet, as N is odd, the tree is centred around the geometric average of the terminal nodes $S^{(N)}(\frac{N+1}{2})$ and $S^{(N)}(\frac{N-1}{2})$ corresponding to the effective number of up-movements $m = 1$ and $m = -1$, respectively. Hence, we have

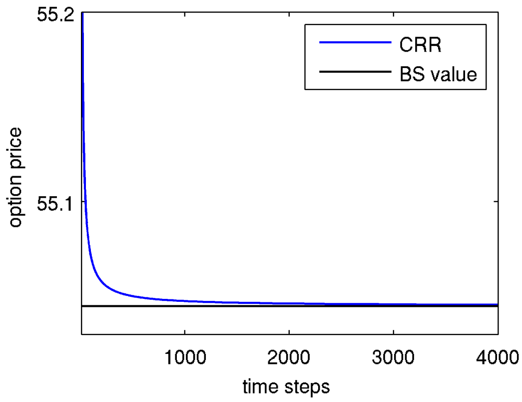
$$K = \left(S^{(N)}\left(\tfrac{N+1}{2} - 1\right) S^{(N)}\left(\tfrac{N+1}{2}\right) \right)^{1/2}. \tag{2.92}$$

Consequently, the strike value is always optimally located in relation to its neighbouring nodes, so that we expect a higher order of convergence. To verify the above conjecture by the asymptotic expansion of the pricing error, note that according to (2.92), $b(N) = 0$ for any odd value of N. As a result, the first error term cancels out, so that the rate of convergence is now $1/N$. Further, it is clear from previous results that the leading term in the discretisation error converges monotonically (compare Corollary 8 and Corollary 9).

Figure 2.5 illustrates the price estimates obtained from the CRR tree along

$N = 9 : 2 : 3999$. Convergence is obviously faster than in the previous case (note that the scaling of the y-axis is not the same as in Figure 2.4). Apparently, convergence is smooth.

Fig. 2.5: Convergence pattern for an at-the-money cash-or-nothing call **(N odd)**
$s_0 = 95$, $\sigma = 0.25$, $r = 0.1$, $T = 1$, $G = 100$, **strike K = 95**



To conclude, we have seen that for the conventional schemes suggested by RB and by CRR, the binomial prices of cash-or-nothing options converge in general no faster than in order $1/\sqrt{N}$. Further, convergence is non-smooth; it suffers from the sawtooth effect and from the even-odd problem. The only exception is the situation $K = s_o e^{\alpha T}$. In this case, both schemes admit smooth convergence along integers of the same parity. In particular, if the binomial prices are evaluated along odd values of N, we achieve convergence of order $1/N$.

Let us stress that we obtain superior convergence properties for the border case $K = s_o e^{\alpha T}$ because the corresponding valuation problem matches *coincidentally* with the definition of the binomial model under consideration. We next illustrate the convergence behaviour of cash-or-nothing prices obtained from the advanced binomial schemes presented in Section 2.2.3; that is, the models suggested by Tian and by Chang and Palmer as well as the optimal drift model we suggest. These models are advantageous for the practical application of binomial option pricing because they can be adapted to the strike value of interest. For cash-or-nothing options, they achieve smooth convergence or increase the rate of convergence *for any specific strike value*.

Advanced Binomial Models from Literature

The advanced binomial schemes presented in Section 2.2.3 can be adapted to a given point x so that the discretisation error around x exhibits a superior asymptotic behaviour. In principle, this is based on the following idea: We start from some binomial process with constant drift $\alpha(N) = \alpha$ and $\beta = \sigma$. For any number of periods N, the drift is then individually corrected for in order to improve the position of the point x in relation to its neighbouring nodes. Clearly, the resulting model does no longer exhibit a constant drift; yet, it achieves a superior asymptotic behaviour of the discretisation error around x. For application to numerical option pricing, *we simply identify the point x with the strike value of interest*. This is illustrated in the following.

The Tian model Let $S_\alpha^{(N)}$ be the binomial process (2.2) with $\beta = \sigma$ and $\alpha(N) \equiv \alpha$ constant in N. Further, let $l_\alpha(N)$ denote the number of up-movements such that the strike values K lies between the nodes $s_N^{(N)}(l_\alpha(N)-1)$ and $s_N^{(N)}(l_\alpha(N))$. According to the results from Section 2.2.3, the Tian model can be adapted to the strike value of interest so that the strike always falls onto the neighbouring upper node $s_N^{(N)}(l_\alpha(N))$ (compare (2.58)). The Tian model hence achieves $b(N) = 1$ for any number of periods N, so that the leading error term always admits its upper bound. As a consequence, the pricing error converges smoothly (compare Proposition 11):

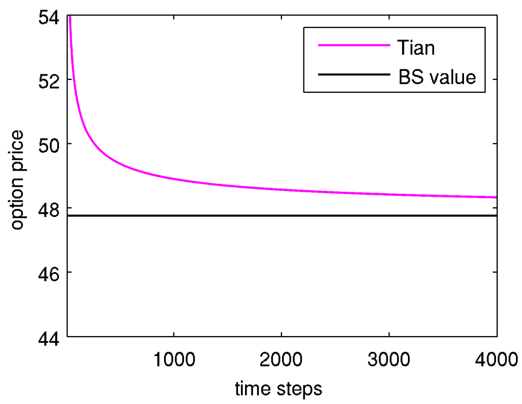
Proposition 29. *For the Tian model, the binomial price of a cash-or-nothing call admits the following asymptotic behaviour:*

$$C^{(N),cash}(K) = C^{cash}(K) + Ge^{-rT}\frac{e^{-\frac{1}{2}d_2^2}}{\sqrt{2\pi T}}\left(\frac{T}{N}\right)^{1/2} + O\left(\frac{1}{N}\right).$$

Remark 20. *Note that the original drift α does not enter the leading term of the pricing error.*

Figure 2.6 illustrates smooth convergence of the Tian tree for $N = 10:1:4000$. As in the original paper, the Tian tree is superimposed on the CRR tree (i.e. $\alpha = 0$); yet as mentioned above, the drift does not significantly influence the convergence pattern.

Fig. 2.6: Convergence pattern for a cash-or-nothing call: The Tian tree
$s_0 = 95$, $\sigma = 0.25$, $r = 0.1$, $T = 1$, $G = 100$, strike $K = 100$



Due to smooth convergence, extrapolation methods can be applied to the Tian model. This is illustrated next. Before, let us give a short reminder on Richardson extrapolation: Suppose that $C(N)$ converges smoothly to the Black-Scholes price C_{BS}, i.e. there are some order r and some constant a_r such that

$$C(N) = C_{\text{BS}} + a_r \tfrac{1}{N^r} + O\left(\tfrac{1}{N^s}\right), \qquad s > r.$$

Then the error ratio is of the form

$$\rho(2N) = \tfrac{C(N)-C_{\text{BS}}}{C(2N)-C_{\text{BS}}} = 2^r + O\left(\tfrac{1}{N^s}\right).$$

It hence converges to $\rho = 2^r$. Extrapolating the observed values (2-point Richardson extrapolation) leads to the aggregated price estimate

$$\hat{C}(2N) := \tfrac{\rho C(2N)-C(N)}{\rho-1},$$

which admits a pricing error of order $O(1/N^s)$. This means that the aggregated price estimate is of *higher order* of accuracy than the original price estimates (for details on Richardson extrapolation see [Tia99]).

According to Proposition 11, applying Richardson extrapolation to the Tian model leads to aggregated estimates of the form

$$C^{(2N),\text{cash}}(K) = C^{\text{cash}}(K) + \tfrac{\sqrt{2}-2}{\sqrt{2}-1} a_1 \tfrac{T}{2N} + o\left(\tfrac{1}{N}\right), \tag{2.93}$$

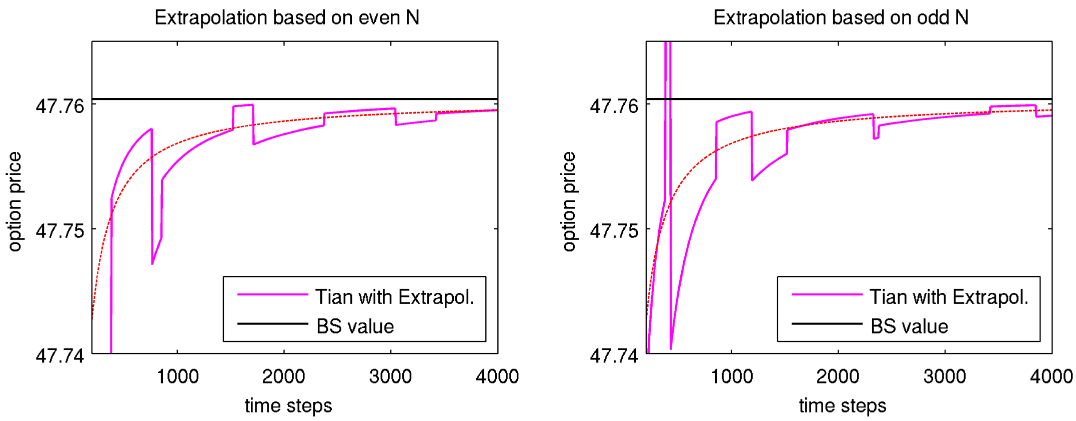
where

$$a_1 = Ge^{-rT} \tfrac{e^{-\frac{1}{2}d_2^2}}{\sqrt{2\pi}} \left(\tfrac{1}{2\sigma^2} d_1 (r-\alpha)^2 + \tfrac{2-d_1 d_2 - d_1^2}{6\sigma\sqrt{T}} (r-\alpha) + \tfrac{d_1^3 + d_1 d_2^2 + 2d_2 - 4d_1}{24T} - \tfrac{d_2}{2T} \right).$$

The aggregated estimates hence converge in order $1/N$. However, as we illustrate next, convergence is no longer smooth. The leading error term converges monotonically; yet the fluctuations of order $o(1/N)$ are not negligible. Consequently, a further extrapolation step cannot be applied.

Figure 2.7 illustrates the sequence of aggregated estimates $\hat{C}(2N)$. Recall that in our numerical example, the Tian tree is superimposed on the CRR model (i.e. $\alpha = 0$). In contrast to the original estimates, for the aggregated estimates the particular choice of α enters the coefficient of the leading error term. The left plot shows the aggregated estimates along even values of N; i.e. $2N = 200:4:4000$ ($N = 100:2:2000$). The plot on the right-hand side illustrates the aggregated estimates along odd values of N; i.e. $2N = 202:4:3998$ ($N = 101:2:1999$). Apparently, the aggregated estimates are exposed to an even-odd effect.

Fig. 2.7: Convergence pattern for a cash-or-nothing call: The Tian tree with 2-point Richardson extrapolation (superimposed on CRR)
$s_0 = 95$, $\sigma = 0.25$, $r = 0.1$, $T = 1$, $G = 100$, strike $K = 100$



The dashed red curve illustrates the leading error term of order $1/N$. The aggregated price estimates oscillate around that curve in order $o(1/N)$. The oscillations observed involve the difference between the constant drift α and the modified drift $\tilde{\alpha}(N)$. With the notations from Section 2.2.3, we have

$$\tilde{\alpha}(N) = \frac{2\sigma\sqrt{T/N}}{T}(a_\alpha(N) - l_\alpha(N)) + \alpha$$

(compare (2.59)). In particular, due to the factor $a_\alpha(N) - l_\alpha(N)$, we anticipate an even-odd effect.

We next consider the convergence behaviour of cash-or-nothing calls for the model suggested by Chang and Palmer. The CP model can be adapted to the strike value of interest so that convergence of order $1/N$ is achieved without extrapolation.

The Chang-Palmer model As in the above, let $S_\alpha^{(N)}$ be the binomial process (2.2) with $\beta = \sigma$ and $\alpha(N) \equiv \alpha$ constant in N. Further, let $l_\alpha(N)$ denote the number of up-movements for which $K \in (s_N^{(N)}(l_\alpha(N) - 1), s_N^{(N)}(l_\alpha(N))]$. According to previous results, the CP model can be adapted to the strike value of interest so that the strike always falls onto the geometric average of the two neighbouring nodes (compare (2.62)). Consequently, for any number of periods N, we have $b(N) = 0$. This implies that the first error term in the asymptotic expansion of the pricing error cancels out (compare Proposition 13):

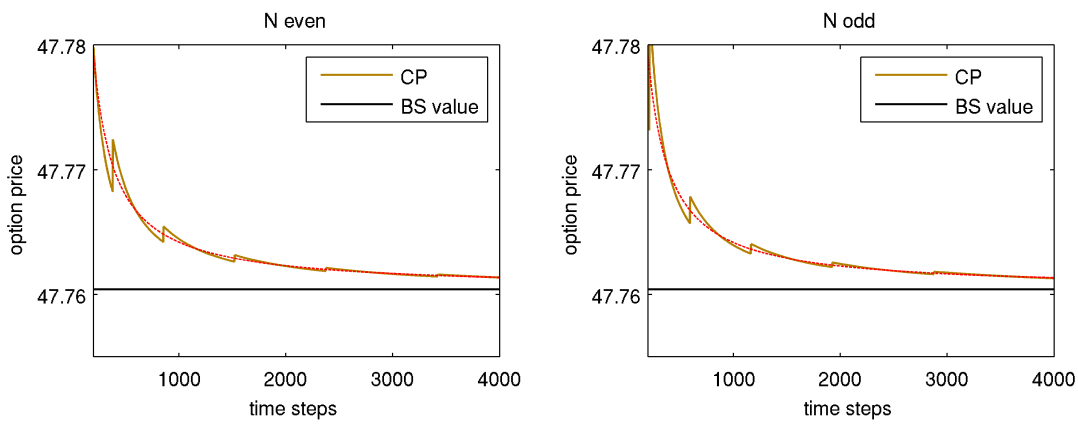
Proposition 30. *Let $S^{(N)}$ be any binomial process (2.2) with $\alpha(N) \equiv \alpha$ constant and $\beta = \sigma$. For the superimposed CP model, the binomial price of a cash-or-nothing call admits the following asymptotic expansion:*

$$C^{(N),cash}(K) = C^{cash}(K) +$$
$$Ge^{-rT} \frac{e^{-\frac{1}{2}d_2^2}}{\sqrt{2\pi}} \left(\frac{1}{2\sigma^2} d_1 (r-\alpha)^2 + \frac{2 - d_1 d_2 - d_1^2}{6\sigma\sqrt{T}} (r-\alpha) + \frac{d_1^3 + d_1 d_2^2 + 2d_2 - 4d_1}{24T} \right) \frac{T}{N} + o\left(\frac{1}{N}\right).$$

Remark 21. *As for the extrapolated Tian model, convergence of cash-or-nothing prices obtained from the CP model is of order $1/N$. However, for the CP model, this order is achieved without extrapolation.*

Figure 2.8 illustrates the convergence behaviour of the CP tree for even ($N = 200 : 2 : 4000$) and for odd values of N ($N = 201 : 2 : 3999$). As in the original paper by Chang and Palmer, the CP tree is superimposed on the CRR tree (i.e. $\alpha = 0$).

Fig. 2.8: Convergence pattern for a cash-or-nothing call: The CP tree (superimposed on CRR) $s_0 = 95$, $\sigma = 0.25$, $r = 0.1$, $T = 1$, $G = 100$, strike $K = 100$



Apparently, we observe oscillations and an even-odd effect. As in the above, we anticipate the presence of oscillations from the fact that the distance between the original drift α and the modified drift $\underline{\alpha}(N)$ varies in N. With the notations from Section 2.2.3, we have

$$\tilde{\alpha}(N) = \frac{2\sigma\sqrt{T/N}}{T} \left(a_\alpha(N) - l_\alpha(N) + \frac{1}{2} \right) + \alpha$$

(compare (2.63)).

We next investigate the convergence behaviour of cash-or-nothing call prices obtained from the OD model we suggest. If we adapt the OD tree to the strike value of interest, we can achieve convergence of order $o(1/N)$.

The Optimal Drift Model

Let $S_\alpha^{(N)}$ be the binomial process (2.2) with $\beta = \sigma$ and $\alpha(N) \equiv \alpha$ constant in N. As explained previously, the optimal drift model is based on the CP model. In contrast to the CP model, the drift α of the embedded process $S_\alpha^{(N)}$ is adapted to the valuation problem under consideration. This further improves the convergence behaviour of cash-or-nothing call prices. In particular, the OD model can achieve convergence of order $o(1/N)$ (compare Proposition 14):

Proposition 31. *For the optimal drift model, the binomial price of a cash-or-nothing call admits the following asymptotic behaviour: If*

$$D(K) = -d_1^4(K) + \sigma\sqrt{T}d_1^3(K) - d_1^2\left(1+\sigma^2 T\right) + 5\sigma\sqrt{T}d_1(K) + 2 \geq 0,$$

we have

$$C^{(N),cash}(K) = C^{cash}(K) + o\left(\tfrac{1}{N}\right);$$

otherwise, we have

$$C^{(N),cash}(K) = C^{cash}(K) + Ge^{-rT}\frac{e^{-\frac{1}{2}d_2^2}}{\sqrt{2\pi}}\frac{1}{36Td_1}\left(d_1^4 + d_1^2 d_2^2 + 5d_1 d_2 - 4d_1^2 - 2 - d_1^3 d_2\right)\frac{1}{N} + o\left(\tfrac{1}{N}\right).$$

Remark 22. *The OD tree converges in order $o(1/N)$ if $d_1(K)$ is reasonably small in absolute value. In essence, this condition excludes deep-in-the-money and deep-out-of-the money situations only. Further, even if the convergence rate cannot be improved, the OD tree is still advantageous compared to the CP model. In this case, the OD model always exhibits a smaller constant of the leading error term: Recall that in our generalisation of the Chang and Palmer model, the coefficient of the leading error term is a quadratic function in α. In the OD model, the constant of leading error term is defined as the vertex of the corresponding parabola.*

Figure 2.9 illustrates the convergence behaviour of the OD tree for even ($N = 200:2:4000$) and for odd values of N ($N = 201:2:3999$). For our numerical example, the strike value is close to the present stock price; consequently, convergence of the OD tree is of order $o(1/N)$. While the estimates obtained from the CP tree oscillate around a monotonically decreasing function of order $1/N$, those obtained from the OD tree oscillate around the Black-Scholes value. Consequently, amongst the competing models, the OD tree is clearly the preferred one by virtue of its superior rate of convergence (compare Table 2.2).

Fig. 2.9: Convergence pattern for a cash-or-nothing call: The OD tree
$s_0 = 95$, $\sigma = 0.25$, $r = 0.1$, $T = 1$, $G = 100$, strike $K = 100$

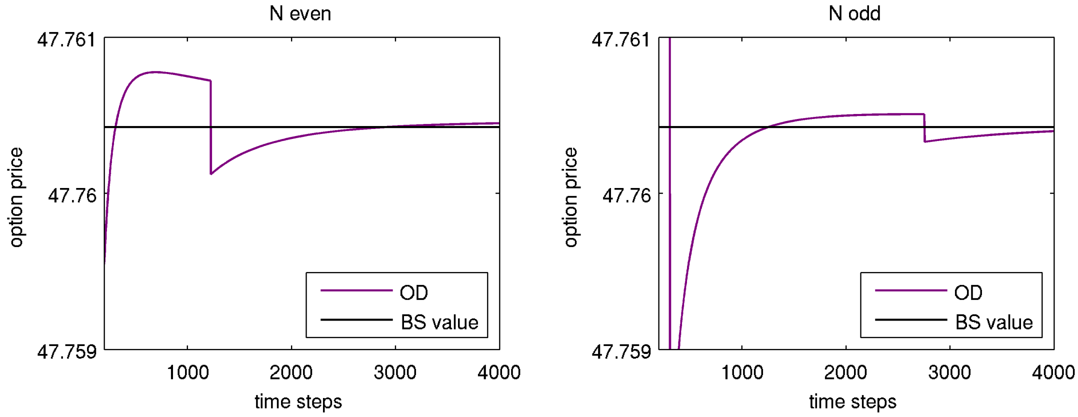


Table 2.2: Cash-or-nothing call prices under conventional and advanced binomial schemes: $s_0 = 95$, $\sigma = 0.25$, $r = 0.1$, $T = 1$, $G = 100$, strike $K = 100$

N	CRR tree	RB tree	Tian	CP tree	OD tree	Tian with extrapolation
200	47.5257	47.7912	50.3228	47.7798	47.7596	47.6913
300	48.9204	47.3242	49.8476	47.7713	47.760391	47.7253
400	46.1524	47.0456	49.5701	47.7717	47.7607	47.7529
500	47.1027	46.8554	49.3772	47.7685	47.7607	47.7556
640	48.0365	46.6682	49.1881	47.7661	47.7608	47.7574
820	48.8635	46.5021	49.0207	47.7644	47.7608	47.7484
1000	47.1805	48.6610	48.9022	47.7644	47.7608	47.7555
2000	47.9034	47.6615	48.5669	47.7623	47.760355	47.7574
3000	47.9178	47.2180	48.4187	47.7617	47.760428	47.7596
4000	47.7477	48.0926	48.3304	47.7614	47.760450	47.7594
5000	47.5104	47.7922	48.2702	47.7612	47.760384	47.7594
10000	47.5869	47.7666	48.1208	47.7608	47.760427	47.7601
15000	47.7984	47.8921	48.0546	47.7607	47.760418	47.7602
BS value	47.760425					

2.5.2 Linear Payoff Structures (Plain Vanilla Options)

In the following, we consider the convergence behaviour of binomial prices for plain vanilla options; i.e. in case of a call (put), we have

$$g(S) = (S_T - K)^+ \qquad\qquad (g(S) = (K - S_T)^+)$$

with strike value K. While cash-or-nothing options exhibit a discontinuity at the strike value, plain vanilla options exhibit a kink; i.e. a discontinuity in the first derivative with respect to the terminal stock price $S(T)$. In the following, we will demonstrate how the specific structure of a plain vanilla option influences the asymptotic behaviour of the discretisation error in the corresponding binomial prices.

Our theroretical analysis of the pricing error will be limited to the risk-neutral case. In this case (and only in this case), the binomial price of a plain vanilla option can be represented as the weighted difference of two binomial distribution functions. As a result, the Berry-Esséen inequality remains applicable. This will be explained in detail. Yet, the two binomial distribution functions representing the price of a plain vanilla option are related to one another, so that cancellation effects occur. As analysed by Chang and Palmer and by Diener and Diener, the leading term in the pricing error cancels out. Consequently, *any risk-neutral binomial model admits convergence of order* $1/N$. In particular, in contrast to cash-or-nothing options, the Berry-Esséen inequality is no longer tight for a constant drift α. However, conventional methods still suffer from an irregular convergence behaviour.

We will further see that the advanced schemes from literature remain superior to conventional methods: For plain vanilla options, they do not increase the rate of convergence, but they can achieve smooth convergence by adapting the tree to the strike value of interest. This will be explained in detail.

In contrast to the advanced schemes suggested by Tian and by Chang and Palmer, the optimal drift model improves the rate of convergence. We will demonstrate that *the optimal drift model can again admit convergence of order* $o(1/N)$. Further, we will briefly discuss the application of the OD model to the American case. For an American put, an explicit pricing formula is not available. The above results suggest that the OD model will be of great practical value for American option valuation by virtue of its superior rate of convergence. This issue will be illustrated numerically.

The Pricing Error for Risk-Neutral Transition Probabilities

In the following, we restrict our attention to binomial models with risk-neutral transition probabilities. Thus, in this case, the binomial call price can be written as the weighted sum of two binomial distribution functions. Consequently, the pricing error can be analysed using previous results.

Recall first that according to the Black-Scholes formula, the call price in the continuous-time setting is of the form

$$\begin{aligned} C(K) &= e^{-rT} E_Q \left(S_T 1_{\{S_T \geq K\}} \right) - K e^{-rT} Q(S_T \geq K) \\ &= s_0 \Phi \left(\frac{\ln(s_0/K) + \left(r + 1/2\sigma^2\right)T}{\sigma\sqrt{T}} \right) - K e^{-rT} \Phi \left(\frac{\ln(s_0/K) + \left(r - 1/2\sigma^2\right)T}{\sigma\sqrt{T}} \right) \end{aligned}$$

(compare e.g. [KK01], p. 101) or equivalently

$$C(K) = s_0 \tilde{Q}(S_T \geq K) - K e^{-rT} Q(S_T \geq K),$$

where $\tilde{Q}$ is the unique equivalent martingale measure if prices are expressed in units of S_t (compare e.g. [KK01], Theorem III.38 on numeraire invariance). It is well-known that we have a similar result for the call price in the binomial model[8]: The binomial call price is given by

$$C^{(N)}(K) = e^{-rT} E_{Q^{(N)}} \left(S_N^{(N)} 1_{\{S_N^{(N)} \geq K\}} \right) - K e^{-rT} Q^{(N)} \left(S_N^{(N)} \geq K \right). \tag{2.94}$$

There exists a suitable probability measure $\tilde{Q}^{(N)}$ under which the first term on the right-hand side of equation (2.94) can be determined from the distribution of $S^{(N)}$ evaluated at the strike K. Similarly as in the continuous-time situation, the required change-of-measure involves introducing $S^{(N)}$ as the numeraire. To be precise on the above arguments, note that

$$\begin{aligned} & e^{-rT} E_{Q^{(N)}} \left(S_N^{(N)} 1_{\{S_N^{(N)} \geq K\}} \right) = \\ & \qquad e^{-rT} \sum_{j=l(N)}^{N} \binom{N}{j} q_u(N)^j \left(1 - q_u(N)\right)^{N-j} s_0 u(N)^j d(N)^{N-j} = \\ & \qquad s_0 \sum_{j=l(N)}^{N} \binom{N}{j} \left(q_u(N) u(N) e^{-rT/N} \right)^j \left((1 - q_u(N)) d(N) e^{-rT/N} \right)^{N-j}, \end{aligned}$$

where as before $q_u(N)$ is the risk-neutral probability for an up-movement, and $l(N)$ is the smallest integer l such that $s_0 u(N)^l d^{N-l}(N) \geq K$ (compare [CRR79]). We now observe that

$$e^{-rT} E_{Q^{(N)}} \left(S_N^{(N)} 1_{\{S_N^{(N)} \geq K\}} \right) = s_0 \tilde{Q}^{(N)} \left(S_N^{(N)} \geq K \right), \tag{2.95}$$

[8] In the above, we have agreed to use the term binomial price ”in a loose sense”. Yet in this section, we have to be more precise using this term.

where $\tilde{Q}^{(N)} := \bigotimes_{k=1}^{N} \tilde{Q}^{(1,N)}$ with

$$\tilde{Q}^{(1,N)}(1) := \tilde{q}_u(N) := q_u(N)e^{-rT/N}u(N) \tag{2.96}$$

and

$$\tilde{Q}^{(1,N)}(-1) := \tilde{q}_d(N) := (1-q_u(N))\, e^{-rT/N}d(N).$$

Note that $\tilde{q}_u(N)+\tilde{q}_d(N)=1$. Further, we have absence of arbitrage opportunities for N sufficiently large (compare Corollary 1). By (AAO), we have $d(N) < e^{rT/N} < u(N)$, which implies that $0 \leq \tilde{q}_u(N) \leq 1$. Consequently, if the number of periods N is sufficiently large, $\tilde{Q}^{(N)}$ is a well-defined probability measure. The binomial call price can hence be represented as the weighted difference of two binomial distribution functions; that is,

$$C^{(N)}(K) = s_0\, \tilde{Q}^{(N)}\left(S_N^{(N)} \geq K\right) - Ke^{-rT} Q^{(N)}\left(S_N^{(N)} \geq K\right). \tag{2.97}$$

Remark 23. *Let us stress that the above representation of the binomial call price depends crucially on the assumption of risk-neutrality. For other probability measures $P^{(N)}$ different from the risk-neutral measure $Q^{(N)}$, we cannot give an analogue to the representation (2.97) because the definitions*

$$\tilde{P}^{(1,N)}(1) := \tilde{p}_u(N) := p_u(N)e^{-rT/N}u(N) \tag{2.98}$$

and

$$\tilde{P}^{(1,N)}(-1) := \tilde{p}_d(N) := (1-p_u(N))\, e^{-rT/N}d(N)$$

do not result in a well-defined probability measure $\tilde{P}^{(N)} := \bigotimes_{k=1}^{N} \tilde{P}^{(1,N)}$. We always have $\tilde{p}_u(N)+\tilde{p}_d(N) \neq 1$.

The above result suggests in particular that the Berry-Esséen bound[9] remains applicable for plain vanilla options in the risk-neutral case; i.e. binomial prices converge in order $O(1/\sqrt{N})$:

Proposition 32. *Let $S^{(N)}$ be any binomial process (2.2) with $\beta=\sigma$ and with risk-neutral transition probabilities. Then for any strike value K, the binomial price of a*

[9] For plain vanilla options, the minimal convergence rate is of course not uniform in the strike value.

plain vanilla call converges in order $1/\sqrt{N}$; i.e.

$$\left|C(K) - C^{(N)}(K)\right| = O\left(\tfrac{1}{\sqrt{N}}\right).$$

Proof. The following proof is based on the representation (2.97) for the binomial call price. We consider the two terms in the above representation separately. Clearly, the second term $Ke^{-rT}\,Q^{(N)}(S_N^{(N)} \geq K)$ is the binomial price of a cash-or-nothing call with strike K and promised cash amount K. Hence, for any strike value K,

$$Ke^{-rT}\left|Q(S_T \geq K) - Q^{(N)}\left(S_N^{(N)} \geq K\right)\right| = O\left(\tfrac{1}{\sqrt{N}}\right).$$

To deal with the first term on the right-hand side of equation (2.97), we show that if the measure is changed from $Q^{(N)}$ to $\tilde{Q}^{(N)}$, the assertion of Proposition 5 is satisfied with $\mu = \left(r + 1/2\sigma^2\right)$ (instead of $\mu = r - 1/2\sigma^2$). Consequently, we need to show that

$$\left|\mu_{\tilde{Q}^{(N)}}(N) - \left(r + 1/2\sigma^2\right)\right| \stackrel{!}{=} O\left(\tfrac{1}{\sqrt{N}}\right) \tag{2.99}$$

and

$$\left|\frac{\sigma^2}{\sigma^2_{\tilde{Q}^{(N)}}(N)} - 1\right| \stackrel{!}{=} O\left(\tfrac{1}{\sqrt{N}}\right) \tag{2.100}$$

where the characteristics $\mu_{\tilde{Q}^{(N)}}(N)$ and $\sigma^2_{\tilde{Q}^{(N)}}(N)$ are computed with respect to $\tilde{Q}^{(N)}$ following (2.8) and (2.9), respectively. Note from the asymptotic expansion of $q_u(N)$ (see (2.15)) that

$$\tilde{q}_u(N) = \tfrac{1}{2} + \tfrac{1}{2\sigma}\left(r - \alpha(N) + 1/2\sigma^2\right)\left(\tfrac{T}{N}\right)^{1/2} + O\left(\tfrac{1}{N^{3/2}}\right)$$

(compare also [DD04], Prop. 3.3). Consequently,

$$\mu_{\tilde{Q}^{(N)}}(N) \stackrel{(2.8)}{=} \alpha(N) + \sigma\left(\tfrac{N}{T}\right)^{1/2}(2\tilde{q}_u(N) - 1) = \left(r + \tfrac{1}{2}\sigma^2\right) + O\left(\tfrac{1}{N}\right)$$

and

$$\sigma^2_{\tilde{Q}^{(N)}}(N) \stackrel{(2.9)}{=} 4\sigma^2\tilde{q}_u(N)(1 - \tilde{q}_u(N)) = \sigma^2 + O\left(\tfrac{1}{N}\right),$$

which shows that the requirements (2.99) and (2.100) are satisfied. Hence, it follows from Proposition 5 that

$$\sup_K \left| \tilde{Q}^{(N)} \left(S_N^{(N)} \geq K \right) - \Phi \left(\frac{\ln\left(\frac{s_0}{K}\right) + \left(r + \frac{1}{2}\sigma^2\right)T}{\sigma\sqrt{T}} \right) \right| = \sup_K \left| \tilde{Q}^{(N)} \left(S_N^{(N)} \geq K \right) - \tilde{Q} \left(S_T \geq K \right) \right| = O\left(\frac{1}{\sqrt{N}} \right),$$

which completes the proof. □

Remark 24. *It is clear that Proposition 32 generalises to any piecewise linear payoff structure and in particular to plain vanilla puts.*

As discussed previously, the above result follows essentially from the fact that the binomial call price can be written as a weighted sum of two binomial distribution functions both evaluated at $l(N)$, which is the smallest integer such that $s_0 u(N)^l d(N)^{N-l} \geq K$. The corresponding success probabilities are related to each other according to definition (2.96). We expect this relationship to cause cancellation effects if we expand the discretisation error of the two distribution functions. In fact, as shown by Diener and Diener and by Chang and Palmer, in total the leading term of the pricing error cancels out. As a result, binomial prices of plain vanilla options converge in order $1/N$ (compare [DD04] Thm. 2.1., [CP07], p. 93):

Proposition 33. *Let $S^{(N)}$ be any binomial process (2.2) with $\beta = \sigma$ and risk-neutral transition probabilities. Then the binomial price of a plain vanilla option (call or put) admits the following asymptotic behaviour:*

$$V^{(N)}(K) = V(K) + \frac{s_0 e^{-\frac{1}{2}d_1^2}}{24\sigma\sqrt{2\pi}\sqrt{T}} \left(f(N) - 12\sigma^2 \left(b^2(N) - 1 \right) \right) \frac{T}{N} + o\left(\frac{1}{N} \right),$$

where

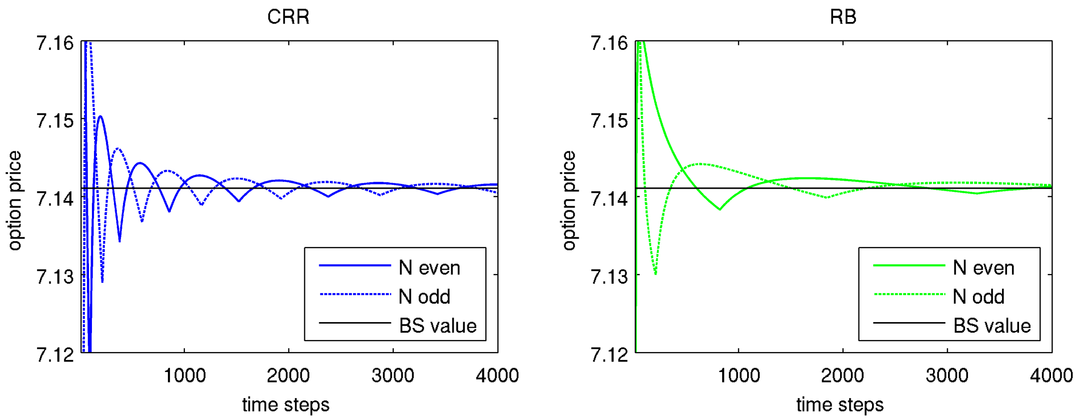
$$f(N) = -12T(r - \alpha(N))^2 + 4\left(d_1^2 - d_2^2\right)(r - \alpha(N)) - \sigma^2 \left(6 + d_1^2 + d_2^2\right).$$

Conventional Schemes

According to Proposition 33, binomial prices of plain vanilla options converge faster than suggested by the Berry-Esséen inequality. Note that the above result also applies to binomial prices obtained from conventional schemes; i.e. for schemes with constant drift α. By contrast, we have seen that for cash-or-nothing options, convergence of order $1/N$ can only be achieved if the drift α is non-constant and adapted to the valuation problem of interest, as is done in the advanced models considered previously. On the other hand, Proposition 33 also indicates that if the drift α is constant, binomial prices of plain vanilla options still suffer from non-monotone convergence and the presence of an even-odd effect.

Let us stress that the above result does not cover the schemes suggested by CRR and by RB because these schemes do not assume risk-neutral transition probabilities (compare Remark 23). However, numerical results suggest that the corresponding sequence of price estimates converges also non-smoothly in order $1/N$ (compare Figure 2.10).

Fig. 2.10: Convergence pattern for a plain vanilla put: The CRR tree and the RB tree
$s_0 = 95$, $\sigma = 0.25$, $r = 0.1$, $T = 1$, strike $K = 100$



Apparently, in contrast to cash-or-nothings options, a sawtooth pattern is not present. Rather, the CRR and RB price estimates oscillate in form of scallops. In fact, the presence of scallops suggests that the sequence $b(N)$ enters the leading error term *quadratically*, so that the price estimates oscillate in form of parabolas. Note that this observation matches with the asymptotic expansion of the pricing error obtained for the risk-neutral case. By contrast, for cash-or-nothing options, the sawtooth effect is due to the fact that $b(N)$ enters the leading term of the pricing error *linearly*.

Advanced Schemes from Literature

For the advanced schemes suggested by Tian and by Chang and Palmer, Proposition 33 leads to the following asymptotic expansion of the pricing error:

Proposition 34. *Let $S^{(N)}$ be any binomial process (2.2) with $\alpha(N) \equiv \alpha$ constant and $\beta = \sigma$. For the superimposed Tian model, the binomial price of a plain vanilla option admits the following asymptotic expansion:*

$$V^{(N)}(K) = V(K) + \frac{s_0 e^{-\frac{1}{2}d_1^2}}{24\sigma\sqrt{2\pi}\sqrt{T}} f \frac{T}{N} + o\left(\frac{1}{N}\right)$$

with

$$f = -12T(r-\alpha)^2 + 4\left(d_1^2 - d_2^2\right)(r-\alpha) - \sigma^2\left(6 + d_1^2 + d_2^2\right).$$

For the superimposed CP model, we have

$$V^{(N)}(K) = V(K) + \frac{s_0 e^{-\frac{1}{2}d_1^2}}{24\sigma\sqrt{2\pi}\sqrt{T}}\tilde{f}\,\frac{T}{N} + o\left(\frac{1}{N}\right),$$

where $\tilde{f} = f + 12\sigma^2$.

Remark 25. *The above result is a generalisation of the results in Chang and Palmer for the case* $\alpha \neq 0$ *(compare [CP07], Corollary 1 and 2). Hence, Proposition 34 allows to apply the optimal drift model to the valuation of plain vanilla options by optimising the drift* α *of the embedded binomial process.*

According to Proposition 34, the Tian model and the CP model do not differ qualitatively when applied to plain vanilla options. For both models, the rate of convergence is not improved compared to conventional schemes, i.e. the rate of convergence is in general no faster than $1/N$. However, the leading term of the pricing error converges monotonically.

Figure 2.11 illustrates the put prices obtained from the two models for our example ($N = 10:2:4000$). We see that higher order oscillations are essentially negligible, so that Richardson extrapolation can be applied. Figure 2.12 shows the corresponding aggregated estimates for $2N = 300:4:4000$.

Fig. 2.11: Convergence pattern for a plain vanilla put: The Tian tree (superimposed on CRR) $s_0 = 95$, $\sigma = 0.25$, $r = 0.1$, $T = 1$, strike $K = 100$

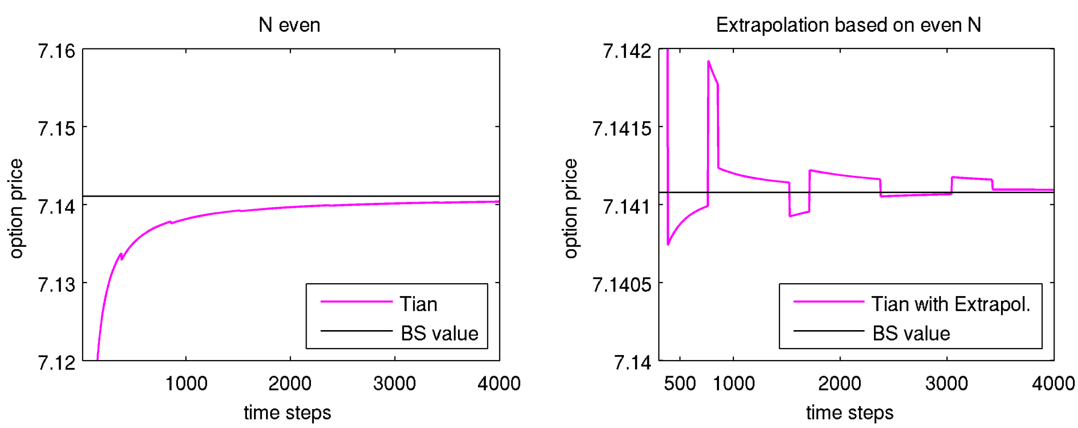
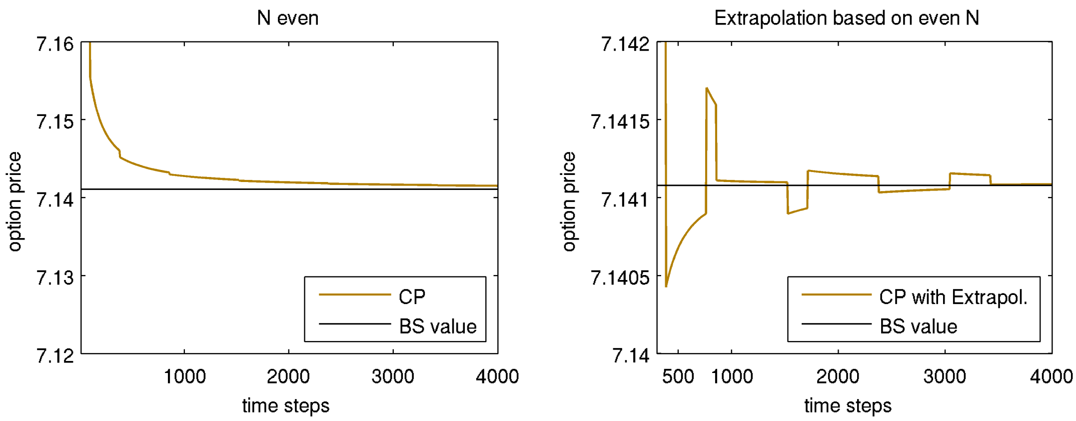


Fig. 2.12: Convergence pattern for a plain vanilla put: The CP tree (superimposed on CRR) $s_0 = 95$, $\sigma = 0.25$, $r = 0.1$, $T = 1$, strike $K = 100$



The Optimal Drift Model

As discussed in the above, the OD model is based on the CP model. In contrast to the latter, it optimises the drift of the embedded binomial model. It follows from Proposition 34 that the OD model can be adapted to the valuation of plain vanilla options so that the convergence rate is improved to $o(1/N)$:

Proposition 35. *For the optimal drift model, we have*

$$V^{(N)}(K) = V(K) + o\left(\tfrac{1}{N}\right)$$

if

$$D(K) = 9 - d_1^2(K) + \sigma\sqrt{T}d_2(K) \geq 0.$$

Otherwise, we have

$$V^{(N)}(K) = V(K) + \tfrac{2}{3T}\sigma^2\left(9 - d_1^2(K) + \sigma\sqrt{T}d_2(K)\right)\tfrac{T}{N} + o\left(\tfrac{1}{N}\right).$$

According to Proposition 35, when applied to the valuation of plain vanilla options, the OD model admits a superior rate of convergence compared to both conventional schemes and the advanced schemes suggested Tian and by Chang and Palmer.

Remark 26.

1. *Note that the condition on convergence of order $o(1/N)$ is even weaker than for cash-or-nothing options. In fact, we may say that with respect to practical appli-*

cations, the condition is satisfied for most interesting cases. In our example with parameters $s_0 = 95$, $\sigma = 0.25$, $r = 0.1$, $T = 1$, *the condition only excludes the cases* $K \leq 49$ *and* $K \geq 222$.

2. *For plain vanilla options, the OD model could in principle also be based on the Tian model. However, in this case, the resulting model would obtain convergence of order* $o(1/N)$ *only if* $D_{Tian}(K) = -9 - d_1^2(K) + \sigma\sqrt{T}d_2(K) \geq 0$.

Figure 2.13 illustrates the put price obtained from the OD model along even and odd values of N for our example. According to the above remark, the rate of convergence is known to be faster than $1/N$ as the strike value is set to $K = 100$. In fact, as we observe from the plot, the pricing error is approximately of the same magnitude as that obtained from extrapolation of the CP or the Tian model. However, for these methods, the application of Richardson extrapolation clearly requires additional computing time. Consequently, the OD tree is again the most advantageous choice amongst the competing methods; it possesses the best time/accuracy trade-off. This is illustrated in Table 2.13. Here in each row, computing time for the binomial tree algorithms is set to 100% and computing time for the two methods involving extrapolation is given as a percentage.

Fig. 2.13: Convergence pattern for a plain vanilla put: The OD tree
$s_0 = 95$, $\sigma = 0.25$, $r = 0.1$, $T = 1$, strike $K = 100$

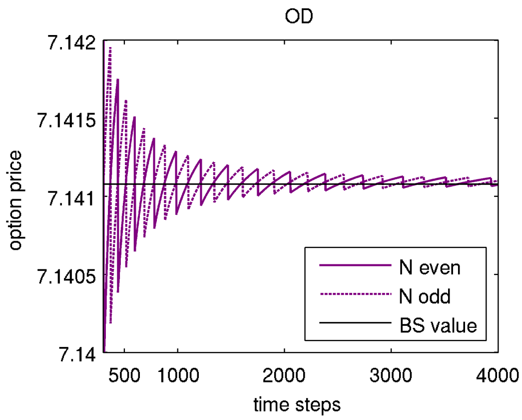


Table 2.3: Plain vanilla put prices under conventional and advanced binomial schemes: $s_0 = 95$, $\sigma = 0.25$, $r = 0.1$, $T = 1$, $G = 100$, strike $K = 100$

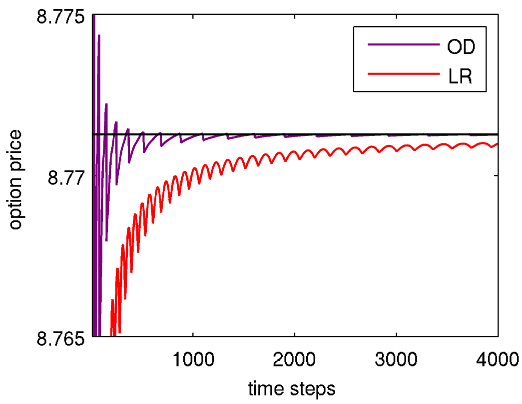
N	CRR tree	RB tree	Tian	CP tree	OD tree	Tian (CP) with extrapolation	
200	7.15025	7.15222	7.12590	7.14961	7.13902	7.14608 (7.14414)	-
300	7.14268	7.14787	7.13157	7.14715	7.14219	7.14344 (7.14255)	-
400	7.13674	7.14491	7.13333	7.14504	7.14146	7.14077 (7.14047)	-
500	7.14332	7.14272	7.13512	7.14443	7.14100	7.14088 (7.14068)	-
640	7.14375	7.14045	7.13660	7.14382	7.14090	7.14095 (7.14083)	-
820	7.13916	7.13832	7.13769	7.14330	7.14094	7.14182 (7.14163)	-
1000	7.14179	7.14049	7.13816	7.14277	7.14092	7.14120 (7.14111)	-
2000	7.14196	7.14217	7.13967	7.14196	7.14114	7.14119 (7.14115)	125 %
3000	7.14163	7.14081	7.14015	7.14167	7.14111	7.141066 (7.14105)	118 %
4000	7.14156	7.14126	7.14039	7.14152	7.141074	7.14110 (7.14109)	128 %
5000	7.14125	7.14153	7.14052	7.14143	7.14109	7.14111 (7.14110)	126 %
10000	7.14117	7.14131	7.14081	7.14126	7.14109	7.14110 (7.14109)	124 %
15000	7.14121	7.14118	7.14091	7.14121	7.14109	7.14109 (7.14109)	124 %
BS Value	7.141079564						

American Put Prices A basic approach to the valuation of American options is the decomposition technique proposed by MacMillan (1986) and Barone-Adesi and Whaley (1987). Here the American option price is divided into that of a similar European option plus the early exercise premium [Mac86], [BAW87]. The decomposition approach suggests that the methods preferred for the European case may often be advantageous for the valuation of the corresponding American option. Consequently, we anticipate that the OD model will also admit strong performance properties when applied to the valuation of American puts.

To confirm the above conjecture by means of our numerical example, Table 2.4 shows the corresponding binomial put prices obtained from the methods under consideration. In fact, as for the European case, the OD tree is the preferred choice amongst the competing binomial methods. Only the methods that include extrapolation lead to a comparable discretisation error, but they require additional computing time.

Let us remark that the Leisen-Reimer tree is currently amongst the most efficient methods for American option pricing (compare e.g. [Sta04], [Sta05]). We wish to stress that for our numerical example, the OD tree also outperforms the LR tree[10] (compare Figure 2.14).

Fig. 2.14: Convergence pattern for an American plain vanilla put: The OD tree vs the LR tree
$s_0 = 95$, $\sigma = 0.25$, $r = 0.1$, $T = 1$, strike $K = 100$



To conclude, due to the widespread use of the binomial approach in American option pricing, the above numerical results promise a great potential of the OD model for practical applications. A profound analysis of this issue is left for future research.

[10] Here (as in Table 2.4), results from the LR tree are obtained with the Preizer-Pratt method 2 inversion [LR96].

Table 2.4: American plain vanilla put prices under conventional and advanced binomial schemes: $s_0 = 95$, $\sigma = 0.25$, $r = 0.1$, $T = 1$, $G = 100$, strike $K = 100$

N	CRR tree	RB tree	Tian	CP tree	OD tree	Tian (CP) with extrapolation		
200	8.77498	8.77567	8.76571	8.77438	8.76929	8.77330	(8.77139)	-
300	8.77133	8.77406	8.76782	8.77350	8.77152	8.77190	(8.77152)	-
400	8.77016	8.77297	8.76866	8.77309	8.77119	8.77162	(8.77180)	-
500	8.77247	8.77115	8.76918	8.77264	8.77099	8.77136	(8.77139)	-
640	8.77235	8.77079	8.76973	8.77240	8.77100	8.77125	(8.77107)	-
820	8.77048	8.76971	8.77010	8.77217	8.77107	8.77146	(8.77134)	133 %
1000	8.77176	8.77050	8.77030	8.77202	8.77109	8.77143	(8.77140)	120 %
2000	8.77167	8.77165	8.77081	8.77167	8.771277	8.77132	(8.77131)	129 %
3000	8.77153	8.77110	8.77098	8.77154	8.771277	8.771294	(8.771290)	128 %
4000	8.77151	8.77131	8.77106	8.77148	8.77126	8.77130	(8.77130)	126 %
5000	8.77139	8.77143	8.77110	8.77145	8.771275	8.77131	(8.77130)	126 %
10000	8.77134	8.77136	8.77120	8.77137	8.771289	8.77130	(8.77130)	126 %
15000	8.77135	8.77132	8.77123	8.77135	8.771288	8.771294	(8.771294)	120 %
LR tree (N=100.001)	8.771281982							

2.6 Conclusion

To conclude, we finally wish to summarise the main aspects of the optimal drift model:

- In the OD model, *the transition probabilities are defined with respect to the risk-neutral measure*.
- The OD model can be adapted to the strike value of interest. In contrast to the Tian model and to the Chang and Palmer model, we optimise the drift of the embedded binomial process. Consequently, the shape of the tree constructed by the OD model further exploits the structure of the valuation problem of interest. This leads to a *superior convergence behaviour of the corresponding binomial option prices.*
- Both for cash-or-nothing options and for plain vanilla options, the prices obtained from the optimal drift model can exhibit convergence of order $o(1/N)$. For these two common payoff structures, *the superior convergence rate of the OD model has been verified rigorously.*
- We anticipate that the OD model will also show strong performance for the valuation of American options. As a major use of binomial methods is in the valuation of American options, we wish to stress *the significance of the strong performance of the OD method for practical applications.*

Chapter 3

The Standard Approach to Multi-Dimensional Trees

We consider an m-dimensional Black-Scholes model with stock price dynamics under the risk-neutral measure Q given by

$$dS_i(t) = S_i(t)(r\mathrm{d}t + \sigma_i \mathrm{d}W_t^i), \quad S_i(0) = s_{i,0} > 0 \qquad \text{for } i = 1, \dots, m \tag{3.1}$$

for Brownian motions W^i and W^j with correlation ρ_{ij} for $i \neq j$. Then the instantaneous returns of stock i and j satisfy

$$\text{Corr}\left[\frac{\mathrm{d}S_i(t)}{S_i(t)}, \frac{\mathrm{d}S_j(t)}{S_j(t)}\right] = \rho_{ij}\,\mathrm{d}t.$$

The correlations are assumed to be such that the corresponding variance-covariance matrix is positive-definite.

In the last two chapters of this thesis, we focus on binomial pricing of multi-asset options. The underlying stocks are assumed to follow the m-dimensional Black-Scholes dynamics defined above.

3.1 Introduction

This chapter deals with the standard approach to valuing options on m correlated (log-normally distributed) stocks by multivariate binomial trees. In principle, this approach works as follows: One approximates the *joint evolution* of the m stocks by an m-dimensional Markov chain appropriate in the sense that its transition states and probabilities (asymptotically) match the drifts, variances and correlations of the increments of the price processes.

In addition to the one-dimensional (for short: 1D) setting, the entire correlation structure of the continuous-time model has to be matched. This often leads to difficulties in the construction of the m-dimensional tree, and it is sometimes even impossible. More precisely, the main drawbacks of standard multi-dimensional tree procedures are as follows:

- Since the number of moment matching conditions grows quadratically in the dimension, setting up an appropriate binomial model soon gets tedious.
- To ensure that the correlation structure between the stocks is matched, correlation parameters typically enter the transition probabilities. This has the effect that transition probabilities can become negative for certain model parameters. Theoretically, application of the tree procedure cannot be justified in this situation.
- Research on 1D trees is not directly applicable.

Although there is a vast amount of literature on 1D trees, there is not so much literature on the standard approach to multi-dimensional trees; a widely known example in 2D is Boyle (1988) (see [Boy88]). Boyle, Evnine and Gibbs (1989) (for short: BEG) suggest an approximation of an m-dimensional geometric Brownian motion by a 2^m-step Markov chain that can be seen as the canonical extension of the 1D CRR tree [BEG89]. However, the transition probabilities in the BEG model are not necessarily well-defined, and in contrast to the 1D CRR model, this problem cannot always be fixed by choosing a sufficiently large number of periods. Kamrad and Ritchken (1991) hence modify the BEG model by introducing horizontal jumps [KR91]. This leads to an additional degree of freedom that can be used to ensure non-negative transition probabilities, but it also increases complexity of the model.

In Chapter 4, we suggest the decoupling method as an alternative approach to multi-dimensional trees that does not suffer from the drawbacks listed above. Let us stress that this approach is conceptually different to the advanced multi-dimensional model suggested by Kamrad and Ritchken. The decoupling approach results in well-defined multi-dimensional trees by separation of the correlation structure from the tree structure. As a consequence, the correlation structure enters transition states rather than transition probabilities. In particular, this approach does not increase the number of degrees of freedom.

In the following, we illustrate the standard approach to multi-dimensional trees with multi-dimensional variants of both the CRR tree and the RB tree. We also highlight the main drawbacks of these models when applied in a multi-dimensional setting.

3.2 Discretisation of the Stock Price

The basic idea of the BEG model is that transition states are made up of components, each of which describes the possible evolution of a component of the discretised m-dimensional stock price process. The joint distribution of the m-dimensional discrete process is such that projecting onto a component results in the 1D discretisation scheme suggested by CRR [BEG89]. As we see in the following, this construction technique can also be used to extend the 1D RB tree to an m-dimensional framework. We next present the resulting m-dimensional variant of the RB tree and compare it to the BEG model.

To guarantee weak convergence to the continuous-time price process, the first two moments of the log-returns must be (asymptotically) matched; i.e. the approximating process $S^{(N)}$ has to be defined on some probability space $(\Omega^{(N)}, \mathscr{F}^{(N)}, P^{(N)})$ so that as grid size tends to zero,

$$\mu_i(N) := \tfrac{1}{\Delta t}\left[\mathrm{E}_{P^{(N)}}\left(\ln\left(\frac{S_{k,i}^{(N)}}{S_{k-1,i}^{(N)}}\right)\middle|\, S_{k-1,i}^{(N)}\right)\right] \to r - \tfrac{1}{2}\sigma_i^2 \qquad \text{for } i = 1,\dots,m \tag{3.2}$$

$$\sigma_i^2(N) := \tfrac{1}{\Delta t}\left[\mathrm{Var}_{P^{(N)}}\left(\ln\left(\frac{S_{k,i}^{(N)}}{S_{k-1,i}^{(N)}}\right)\middle|\, S_{k-1,i}^{(N)}\right)\right] \to \sigma_i^2 \qquad \text{for } i = 1,\dots,m \tag{3.3}$$

and

$$c_{ij}(N) := \tfrac{1}{\Delta t}\left[\mathrm{Cov}_{P^{(N)}}\left(\ln\left(\frac{S_{k,i}^{(N)}}{S_{k-1,i}^{(N)}}\right), \ln\left(\frac{S_{k,j}^{(N)}}{S_{k-1,j}^{(N)}}\right)\middle|\, S_{k-1,i}^{(N)}, S_{k-1,j}^{(N)}\right)\right] \to \sigma_i\sigma_j\rho_{ij}$$
$$\text{for } i = 1,\dots,m,\ j < i. \tag{3.4}$$

The i^{th} component process coincides with the RB model if

$$\begin{aligned} S_{0,i}^{(N)} &= s_{0,i} \\ S_{k,i}^{(N)} &= \begin{cases} S_{k-1,i}^{(N)} e^{(r-\frac{1}{2}\sigma_i^2)\Delta t + \sigma_i\sqrt{\Delta t}} & \text{with prob. } \frac{1}{2} \\ S_{k-1,i}^{(N)} e^{(r-\frac{1}{2}\sigma_i^2)\Delta t - \sigma_i\sqrt{\Delta t}} & \text{with prob. } \frac{1}{2} \end{cases} \end{aligned} \tag{3.5}$$

It coincides with the CRR model if

$$\begin{aligned} S_{0,i}^{(N)} &= s_{0,i} \\ S_{k,i}^{(N)} &= \begin{cases} S_{k-1,i}^{(N)} e^{\sigma_i\sqrt{\Delta t}} & \text{with prob. } \frac{1}{2}\left(1 + \left(r - \frac{1}{2}\sigma_i^2\right)\frac{\sqrt{\Delta t}}{\sigma}\right) \\ S_{k-1,i}^{(N)} e^{-\sigma_i\sqrt{\Delta t}} & \text{with prob. } \frac{1}{2}\left(1 - \left(r - \frac{1}{2}\sigma_i^2\right)\frac{\sqrt{\Delta t}}{\sigma}\right) \end{cases} \end{aligned}$$

As we recall from Section 2.2, for both models the discrete components (asymptotically) satisfy the moment matching conditions (3.2) and (3.3). These conditions are satisfied exactly for the RB model. For the CRR model, the variance is only matched when ignoring some term of order Δt (compare (2.19)).

If we take a simple product of the embedded 1D trees, the correlation condition (3.4) will in general not be satisfied. Rather, we have to define appropriate one-step transitions that take into account the correlation structure of the continuous-time model. This is discussed next.

Let us consider some period $k \leq N$. Since each component can either increase or decrease, we introduce the set of all possible *up-down-scenarios*, i.e.

$$\mathscr{E}_k = \left\{ \underline{\omega_k} = \left(\omega_{k,1}, \ldots, \omega_{k,m}\right) \mid \omega_{k,i} \in \{-1,1\} \quad \forall\, i = 1, \ldots, m \right\}.$$

To obtain appropriate transitions in the m-dimensional RB model, we define $P_k : \mathscr{E}_k \to \mathbb{R}$ by

$$P_k\left(\underline{\omega_k}\right) := \tfrac{1}{2^m} \left(1 + \sum_{\substack{i,j=1 \\ i<j}}^{m} \rho_{ij} \delta_{ij}\left(\underline{\omega_k}\right) \right), \tag{3.6}$$

where

$$\delta_{ij}\left(\underline{\omega_k}\right) = \begin{cases} 1 \text{ if } \omega_{k,i} = \omega_{k,j} \\ -1 \text{ if } \omega_{k,i} \neq \omega_{k,j} \end{cases} \tag{3.7}$$

(a similar construction appears in [Ami91], [HW94]). Note that by symmetry

$$\sum_{\underline{\omega_k} \in \mathscr{E}_k} \sum_{\substack{i,j=1 \\ i<j}}^{m} \rho_{ij} \delta_{ij}\left(\underline{\omega_k}\right) = 0 \tag{3.8}$$

and hence,

$$P_k(\mathscr{E}_k) = \sum_{\underline{\omega_k} \in \mathscr{E}_k} P_k\left(\underline{\omega_k}\right) \overset{(3.6)}{=} 1 + \tfrac{1}{2^m} \sum_{\underline{\omega_k} \in \mathscr{E}_k} \sum_{\substack{i,j=1 \\ i<j}}^{m} \rho_{ij} \delta_{ij}\left(\underline{\omega_k}\right) \overset{(3.8)}{=} 1.$$

Apparently, as the set $\mathscr{E}_k$ is finite, P_k can be extended to a probability measure on the corresponding discrete probability space if the correlation structure is such that P_k is non-negative.

While P_k is always non-negative in dimension $m = 1$ and $m = 2$, there are restrictions on the correlation structure in higher dimensions. As an example, consider an option on three underlyings S_1, S_2, S_3 with $\rho_{12} = -0.7$, $\rho_{23} = 0.1$ and $\rho_{13} = -0.5$; then

$$P_k\left(\omega_{k,1} = \omega_{k,2} = \omega_{k,3} = -1\right) = -0.0125 < 0.$$

Clearly, *for these model parameters the construction technique does not result in a well-defined discrete model.*

It is known from BEG that we obtain appropriate transitions in a m-dimensional CRR model if we set

$$P_k\left(\underline{\omega}_k\right) := \tfrac{1}{2^m}\left(1 + \sum_{\substack{i,j=1\\ i<j}}^{m} \rho_{ij}\delta_{ij}\left(\underline{\omega}_k\right) + \sqrt{\Delta t}\sum_{i=1}^{m}\delta_i\left(\underline{\omega}_k\right)\frac{r-\frac{1}{2}\sigma_i^2}{\sigma_i}\right) \tag{3.9}$$

with $\delta_{ij}(.)$ defined as in (3.7) and $\delta_i\left(\underline{\omega}_k\right) := 1\ (-1)$ if $\omega_{k,i}$ is an up-scenario (down-scenario). As in the 1D setting, the transition probabilities chosen depend on the grid size. As the grid size tends to zero, each transition probability converges to the corresponding transition probability in the multi-dimensional variant of the RB tree, i.e.

$$P_k^{\mathrm{CRR}}\left(\underline{\omega}_k\right) \to P_k^{\mathrm{RB}}\left(\underline{\omega}_k\right) \qquad \forall \underline{\omega}_k \in \mathscr{E}_k.$$

Consequently, the two models will be applicable under the same restrictions on the model parameters, but on top of this, the BEG tree requires a sufficiently small grid size.

In the following, we assume that the model parameters are such that P_k can be extended to a well-defined probability measure. Let us introduce the generic element ω describing the up-down-behaviour of a path of the multi-dimensional discrete asset price

$$\omega = \left(\omega_{1,1}, ..., \omega_{1,m}, ..., \omega_{N,1}, \ldots, \omega_{N,m}\right) \in \mathscr{E}_1 \times \ldots \times \mathscr{E}_N$$

and let us introduce the coordinate mappings

$$Z_{k,i}\left(\omega\right) = \omega_{k,i}, \qquad \text{for } k = 1, ..., N,\ i = 1, ..., m.$$

Then we can define the one-step transition in the m-dimensional variant of the RB model by

$$S_k^{(N)} := \begin{pmatrix} S_{k,1}^{(N)} \\ \vdots \\ S_{k,m}^{(N)} \end{pmatrix} := \begin{pmatrix} S_{k-1,1}^{(N)} e^{(r-\frac{1}{2}\sigma_1^2)\Delta t + Z_{k,1}\sigma_1\sqrt{\Delta t}} \\ \vdots \\ S_{k-1,m}^{(N)} e^{(r-\frac{1}{2}\sigma_m^2)\Delta t + Z_{k,m}\sigma_m\sqrt{\Delta t}} \end{pmatrix}.$$

Finally, we equip the path space $\mathscr{E}$ with the probability measure

$$P^{(N)} = \bigotimes_{k=1}^{N} P_k$$

under which $Z_{k,i}$ and $Z_{l,j}$ are independent for $k \neq l$. In particular, $S^{(N)}$ is a Markov process.

By summing over all possible combinations with $\omega_{k,i}$ fixed, it follows from symmetry arguments that

$$\begin{aligned} P^{(N)}\left(S_{k,i}^{(N)} = S_{k-1,i}^{(N)} e^{(r-\frac{1}{2}\sigma_i^2)\Delta t + \sigma_i\sqrt{\Delta t}} \,\middle|\, S_{k-1,i}^{(N)}\right) &= \frac{1}{2} \\ P^{(N)}\left(S_{k,i}^{(N)} = S_{k-1,i}^{(N)} e^{(r-\frac{1}{2}\sigma_i^2)\Delta t - \sigma_i\sqrt{\Delta t}} \,\middle|\, S_{k-1,i}^{(N)}\right) &= \frac{1}{2} \end{aligned}$$

Hence, as desired, each component coincides with the one-step transition in the RB model (compare to (3.5)).

It remains to show that the covariance of the log-returns is matched under $P^{(N)}$: When we sum over the relevant elements of $\mathscr{E}_k$, it follows again by symmetry that

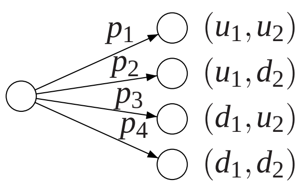
$$\begin{aligned} P^{(N)}(Z_{k,i} = 1, Z_{k,j} = 1) = P^{(N)}(Z_{k,i} = -1, Z_{k,j} = -1) = \tfrac{1}{4}(1 + \rho_{ij}) \\ P^{(N)}(Z_{k,i} = 1, Z_{k,j} = -1) = P^{(N)}(Z_{k,i} = -1, Z_{k,j} = 1) = \tfrac{1}{4}(1 - \rho_{ij}) \end{aligned}$$

which gives us

$$\mathrm{Cov}_{P^{(N)}}\left(Z_{k,i}, Z_{k,j}\right) = \rho_{ij}.$$

Figure 3.1 illustrates the one-step transitions for the 2D RB tree.

Fig. 3.1: One-step transition for the 2D RB model



$$\begin{aligned} u_i &= e^{(r-0.5\sigma_i^2)\Delta t + \sigma_i\sqrt{\Delta t}} \\ d_i &= e^{(r-0.5\sigma_i^2)\Delta t - \sigma_i\sqrt{\Delta t}} \end{aligned}$$

$p_1 = p_4 = \frac{1}{4}(1 + \rho)$; $\qquad p_2 = p_3 = \frac{1}{4}(1 - \rho)$

For the BEG model, one-step transitions are defined as

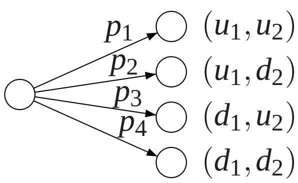
$$S_k^{(N)} := \begin{pmatrix} S_{k-1,1}^{(N)} e^{Z_{k,1}\sigma_1\sqrt{\Delta t}} \\ \vdots \\ S_{k-1,m}^{(N)} e^{Z_{k,m}\sigma_m\sqrt{\Delta t}} \end{pmatrix}.$$

As for the variances, the BEG model matches the correlation structure of the continuous-time model only asymptotically when ignoring some term of order Δt; to be precise,

$$c_{ij}(N) = \sigma_i\sigma_i\rho_{ij} - \left(r - \tfrac{1}{2}\sigma_i^2\right)\left(r - \tfrac{1}{2}\sigma_j^2\right)\tfrac{T}{N}.$$

Figure 3.2 shows one-step transitions for the 2D CRR model.

Fig. 3.2: One-step transition for the 2D CRR model (=2D BEG)



$$u_i = e^{\sigma_i\sqrt{\Delta t}}$$
$$d_i = e^{-\sigma_i\sqrt{\Delta t}}$$

$$p_1 = \tfrac{1}{4}\left(1+\rho+\left(\frac{r-\frac{1}{2}\sigma_i^2}{\sigma_i}+\frac{r-\frac{1}{2}\sigma_j^2}{\sigma_j}\right)\sqrt{\Delta t}\right); \quad p_2 = \tfrac{1}{4}\left(1-\rho+\left(\frac{r-\frac{1}{2}\sigma_i^2}{\sigma_i}-\frac{r-\frac{1}{2}\sigma_j^2}{\sigma_j}\right)\sqrt{\Delta t}\right)$$

$$p_3 = \tfrac{1}{4}\left(1-\rho+\left(-\frac{r-\frac{1}{2}\sigma_i^2}{\sigma_i}+\frac{r-\frac{1}{2}\sigma_j^2}{\sigma_j}\right)\sqrt{\Delta t}\right); \quad p_4 = \tfrac{1}{4}\left(1+\rho+\left(-\frac{r-\frac{1}{2}\sigma_i^2}{\sigma_i}-\frac{r-\frac{1}{2}\sigma_j^2}{\sigma_j}\right)\sqrt{\Delta t}\right)$$

Clearly, for the BEG tree and for the suggested m-variant of the RB tree, each node has 2^m successor nodes. The log-BEG tree inherits symmetry around the starting value from its 1D components, while the multi-dimensional variant of the RB tree possesses symmetry properties with respect to the transition probabilities. As a consequence, for the latter the number of distinct weights is reduced to 2^{m-1}.

Remark 27 (Incompleteness of the discrete market)**.** *Let us stress that the discrete market model associated with the binomial processes $S_1^{(N)}, \ldots, S_m^{(N)}$ defined as above and a bond with one-period return $e^{r\Delta t}$ is not complete (compare e.g. [Bjö04], Section 3). Consequently, in the discrete model there is no unique option price. However, in the* *context of binomial option pricing, there is no impact of the model being complete or incomplete on convergence to the exact price. For an approximation of the multi-dimensional Black-Scholes model by a complete multinomial model, we refer to He (1990) (see [He90]).*

To conclude, we have seen that as one-step transition probabilities are not simply products of marginal probabilities, it is tedious to define a law of the discrete process that suitably approximates the joint distribution of the continuous-time process. The correlation structure of the continuous-time model enters the one-step transition probabilities (compare (3.6), (3.9)). This implies that each set of model parameters leads to a particular specification of the measure $P^{(N)}$, which is not always well-defined in dimensions $m \geq 3$.

3.3 Option Valuation with Standard Multi-Dimensional Trees

As seen in the above, provided the BEG tree and the suggested multi-dimensional variant of the RB tree are well-defined, they (asymptotically) satisfy the moment matching conditions (3.2) - (3.4). As a consequence, due to Donsker's Theorem, the continuous process obtained from $S^{(N)}$ by linear interpolation and appropriate time-scaling converges weakly to the stock price process in an m-dimensional Black-Scholes setting. As explained previously, this provides the theoretical basis for numerical option valuation. Using a loose terminology (compare Remark 27), the resulting estimates for the exact option price are again referred to as "binomial prices".

In Section 3.3.1, we briefly discuss the main aspects of the corresponding tree algorithm for numerical option valuation. In Section 3.3.2, we investigate the convergence behaviour of binomial prices obtained from the standard multi-dimensional schemes considered above. Our analysis is focused on payoff functions that exhibit discontinuities. We will see that the multi-dimensional binomial schemes inherit the irregular convergence behaviour observed for their 1D variants.

3.3.1 The Tree Algorithm

For the standard multi-dimensional schemes considered above, the corresponding tree algorithm is conceptually the same as for the 1D case: First, we assign possible payoff scenarios to the terminal nodes. Afterwards, we step backwards through the tree, as suggested by the Markov property of the process $S^{(N)}$. The tree algorithm for the valuation of path-independent options with two underlyings can be found below (compare Algorithm 5). In accordance with Figures 3.1 and 3.2, p_1 denotes the weight assigned to state (u_1, u_2), p_2 denotes the weight assigned to state (u_1, d_2), etc. The weights are determined in advance. For path-dependent options, it depends again on the specific

Algorithm 5: The standard approach to binomial trees for a path-independent European option with two underlyings

Input: stock price parameters, risk-neutral rate r, payoff function g
Return: price estimate $= V[0][0] \times \exp(-r \times T)$

1. Forward Step

{initialise asset prices at maturity}
Set $\text{SN}[0][1] := s_{0,1} d_1^N$
Set $\text{SN}[0][2] := s_{0,2} d_2^N$
for $k = 1$ to N **do**
$\text{SN}[k][1] := \text{SN}[k-1][1] \times (u_1/d_1)$
$\text{SN}[k][2] := \text{SN}[k-1][2] \times (u_2/d_2)$
end for

{initialise option values at maturity}
for $k_1 = 0$ to N **do**
for $k_2 = 0$ to N **do**
Set $V[k_1][k_2] := g(\text{SN}[k_1][1], \text{SN}[k_2][2])$
end for
end for

2. Backward Induction

{step backwards through the tree}
for $k = N-1$ to 0 **do**
for $l_1 = 0$ to k **do**
for $l_2 = 0$ to k **do**

$$V[l_1][l_2] := p_1 \times V[l_1+1][l_2+1] + p_2 \times V[l_1+1][l_2] + p_3 \times V[l_1][l_2+1] + p_4 \times V[l_1][l_2]$$

{or under the 2D RB scheme}

$$V[l_1][l_2] := p_1 \times (V[l_1+1][l_2+1] + V[l_1][l_2]) + p_2 \times V[l_1+1][l_2] + V[l_1][l_2+1])$$

end for
end for
end for

payoff functional whether there exists a suitable modification of the above algorithm.

As for the 1D case, computational effort required for Algorithm 5 can essentially be attributed to the backward induction step. We have the following result:

Proposition 36. *In the general m-dimensional situation, Algorithm 5 requires computational effort of order $O(N^{m+1})$. The leading constant is in general $\frac{2^{m+1}-1}{m+1}$. For the m-dimensional RB tree, the leading constant is reduced to $\frac{3/2\, 2^m-1}{m+1}$.*

Proof. Excluding the terminal layer, we have to consider $\sum_{k=0}^{N-1}(k+1)^m$ nodes. Hence, the number of relevant nodes is of order $O(N^{m+1})$ with constant $1/(m+1)$. In the backward step, the arithmetic mean is computed at each of these nodes, which in general requires $2^{m+1}-1$ operation counts per node (2^m multiplications and 2^m-1 addition). However, as seen above, for the m-dimensional RB tree, the number of distinct weights is reduced to 2^{m-1}. It follows from the distributive law that the number of operation counts per node is reduced to $\frac{3}{2}2^m-1$. □

Remark 28.

- *According to Proposition 36, computational effort grows exponentially in the number of the underlying stocks. Therefore, the above multi-dimensional tree algorithm is currently not practically useful for high-dimensional valuation problems.*

- *We have seen that the tree algorithm associated with the multi-dimensional variant of the RB tree requires less computational effort than that associated with the BEG tree. As we illustrate below, the difference in operation counts is reflected in computing time. Note that by contrast, for the one-dimensional examples considered previously, the tree algorithms do not differ significantly in computing time. Nevertheless, computational effort required by the multi-dimensional RB tree is still suboptimal for the rate of growth of the tree.*

3.3.2 The Convergence Behaviour of Binomial Option Prices

As observed in the above, the standard approach to multi-dimensional trees suffers from several conceptual drawbacks: the construction of trees is tedious, its application is restricted by model parameters, etc. These drawbacks are specific to the multi-dimensional situation. On top of that, the irregular convergence behaviour of conventional 1D schemes is inherited to their multi-dimensional tree variants. This is illustrated next.

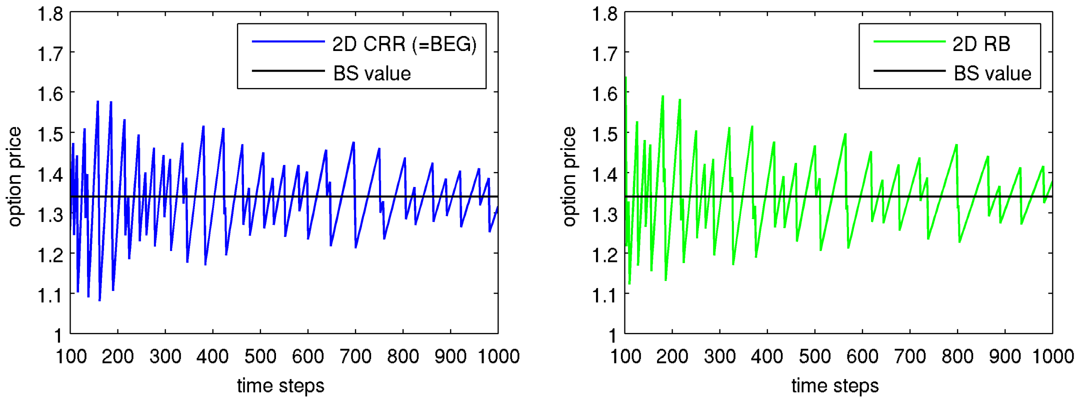
We investigate the convergence behaviour of the above schemes for payoff functions that exhibit discontinuities; first, for cash-or-nothing options; and second, for barrier options. For the examples considered below, an explicit pricing formula is known. As discussed previously, such simples examples will give an intuition on the convergence behaviour of similar types of options for which an explicit pricing formula is not known. Note in this context that the pricing formula we will use for an analytic valuation of barrier options is only valid under restrictions on the correlation structure.

Cash-or-Nothing Options We consider a two-asset cash-or-nothing call; i.e.

$$g(S_1, S_2) = G\,1_{\{S_1(T) \geq K_1\,,\, S_2(T) \geq K_2\}}$$

with strike values $K_1 > 0$ and $K_2 > 0$ and with a promised cash-amount $G > 0$. Figure 3.3 illustrates the convergence pattern obtained for the 2D RB tree and for the BEG tree for $N = 100 : 2 : 1000$.

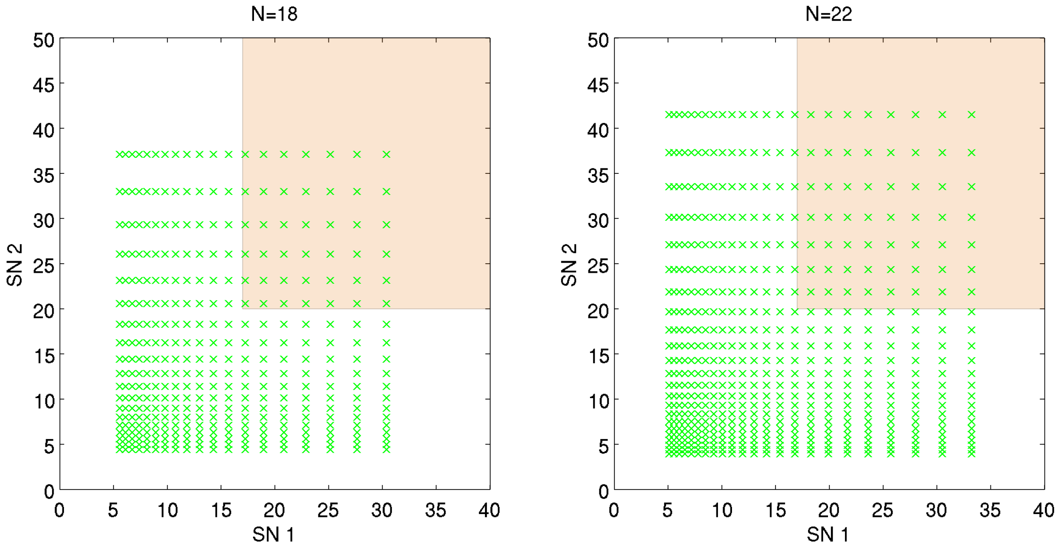
Fig. 3.3: Convergence pattern for a two-asset cash-or-nothing call;
$S_1(0) = 12.0$, $S_2(0) = 12.0$, $\sigma_1 = 0.2$, $\sigma_2 = 0.25$, $\rho = 0.5$, $T = 1.0$ $r = 0.1$, $K_1 = 17.0$, $K_2 = 20.0$, $G = 100$



For our numerical example, both methods suffer heavily from an irregular convergence behaviour. As for cash-or-nothing options with a single underlying, the price estimates are neither consistently greater nor less than the exact price; rather, they alternate in a sawtooth pattern between over- and underestimation with some highly accurate values in between (compare Figure 2.2). In contrast to the 1D case, the amplitude of the oscillations does not decrease monotonically in the number of periods N.

Similar to the 1D situation, the sawtooth effect can be traced back to the fact that when grid sizes changes, the position of nodes in the tree varies in relation to the strike values K_1 and K_2. As a consequence, there is typically either too much probability mass in the in-the-money or in the out-of-the money area, respectively. This is illustrated next.

Fig. 3.4: 2D RB tree: Realisations of $(S_1^{(N)}, S_2^{(N)})$;



As we see from Figure 3.4, the terminal values of the two-dimensional process $S^{(N)}$ form a rectangular grid. This is due to the fact that the transition states are defined as a Cartesian product of the two components. The rectangular grid structure of the terminal nodes is illustrated above for the 2D RB tree, but we will observe a similar pattern for the BEG tree. The coloured rectangle illustrates the in-the-money region. Its borderlines are given by the strike values $K_1 = 17$ and $K_2 = 20$. Due to the rectangular grid structure, there are "columns" and "rows" parallel to the borderlines of the in-the-money area. For $N = 18$, there is a column ($S_{N,1}^{(N)} = 17.2489$) and a row ($S_{N,2}^{(N)} = 20.5953$) right *above* the borderlines. In contrast, for $N = 22$ there is a column ($S_{N,1}^{(N)} = 16.7894$) and a row ($S_{N,2}^{(N)} = 19.689$) right *below* the borderlines. Consequently, if we count the nodes in the coloured rectangle and weight them with respect to $P^{(N)}$, the risk-neutral probability to end up in-the-money (1.4819%) is heavily overestimated for $N = 18$ (2.1433 %) and it is heavily underestimated for $N = 22$ (0.9354 %). As a result, the option price (1.34087 £) is heavily overpriced in the first case (1.93932 £) and heavily underpriced in the second case (0.84634 £).

For the number of periods N in the above example, both components simultaneously lead to an over- or underestimation of the probability mass in-the-money region. Yet this is not necessarily the case. Rather, for many values of N, the likelihood of the event that the first stock ends up in-the-money is overestimated, while the likelihood of the event that the second stock ends up in-the-money is underestimated, or vice versa. This is due to the fact that the corresponding 1D cash-or-nothing option on the first asset oscillates in general with a different frequency than that on the second asset. Due to the superposition of oscillations with different frequencies, the amplitude of the oscillation observed for a multi-asset cash-or-nothing option ceases to decrease monotonically.

As illustrated in the above, for cash-or-nothing options the convergence pattern observed for the multi-dimensional variant of the RB tree does not differ significantly from that observed for the BEG model. However, there is a significant difference in computing time. In fact, in the above example, *computing time required by the multi-dimensional RB tree is reduced to approximately 80% of that required by the BEG tree*. Let us stress that we also observe a reduction in computing time for other types of options. Note that by contrast, for our examples on options with a single underlying, the different tree algorithms do not differ significantly in computing time (a difference in computing time is observed only if an additonal extrapolation step is used).

Barrier Options In the following, we deal with the convergence behaviour of barrier options, as a special type of path-dependent options. Two examples will be presented. The first example is a barrier option with two knock-out barriers. In the second example, we consider two barriers of different type, a knock-in barrier on stock 1 and a knock-out barrier on stock 2. For both cases, the barriers are assumed to be constant in the underlying stocks. As for the cash-or-nothing option considered above, both examples are such that the option promises a cash amount of $G > 0$ paid at maturity.

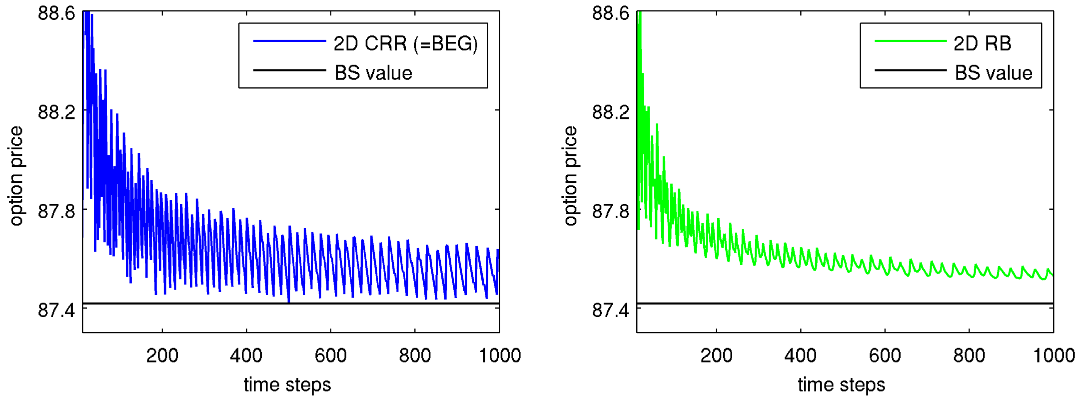
Example 1 Let us consider a knock-out barrier option on two stocks with payoff

$$g(S_1, S_2) = G\, 1_{\{S_1(t) < B_1\ \forall t \in [0,T]\ ,\ S_2(t) > B_2\ \forall t \in [0,T]\}}.$$

Figure 3.5 shows the convergence pattern of the methods under consideration for $N = 10:2:1000$. For the parameter setting given, the Black-Scholes value can be calculated explicitly as suggested in He et al. (1998) [1].

[1]The explicit formula suggested is applicable if the correlation is of the form $\rho = \cos(\Pi/n)$ with $n = 3, 4, \ldots$. Otherwise (in particular for a negative correlation!) there is only a semi-analytical formula involving an infinite sum of Bessel functions [HKR98].

Fig. 3.5: Convergence pattern for a barrier option with an up-and-out barrier on stock 1 and a down-and-out barrier on stock 2;
$S_1(0) = 20.0$, $S_2(0) = 30.0$, $\sigma_1 = 0.2$, $\sigma_2 = 0.25$, $T = 1.0$ $r = 0.1$, $B_1 = 33.0$, $B_2 = 15.0$, $G = 100$ and **correlation $\rho = 0.5$**



Apparently, both methods overestimate the exact price. Furthermore, convergence of binomial prices is again exposed to oscillations. For the BEG model, fluctuations exhibit a sawtooth pattern with some cusp points that are already reasonably accurate for small values of N. For the 2D RB tree, there are no optimal choices for the grid size. However, the oscillations are of lower amplitude. The patterns observed will be analysed next. Before, let us stress that for the above example, the 2D RB tree is again advantageous with respect to computing time: For fixed grid size, the computing time is reduced to approximately 90% of that required by the BEG tree. Apparently, time reduction is not as good as for the previous example. This is due to the knock-out feature, so that a zero value is assigned to the nodes above/below the corresponding barrier. For these nodes, neither the RB tree nor the BEG tree requires any computional effort. Consequently, compared to path-independent options, the RB tree saves computional effort only for the nodes for which the option is not knocked out.

We wish to add that the above patterns are generic for any knock-out option with barriers constant in the underlying; in particular, we observe similar patterns for the 1D case.

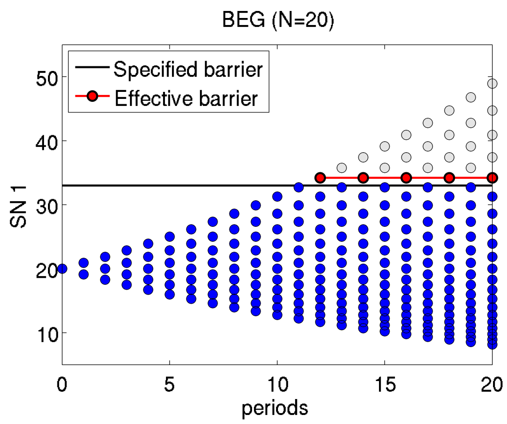
Let us first consider the pattern observed for the BEG tree. For any number of periods N, the specified barriers will in general lie between two horizontal layers of nodes in the corresponding tree. To be precise, the up-and-out barrier on stock 1 lies between the corresponding binomial process after $m-1$ effective up jumps and the binomial process after m effective up jumps, i.e.

$$s_{0,1}u_1^{m-1} < B_1 < s_{0,1}u_1^m,$$

where $u_1 = e^{\sigma\sqrt{T/N}}$ denotes the corresponding upward jump size in the BEG model. Consequently, the binomial tree first ”feels” the effect of the barrier at $s_{0,1}u_1^m$, which is therefore referred to as the effective barrier (compare [DKEB95]). Clearly, *the effective barrier is larger than the specified barrier* B_1.

Figure 3.6 illustrates the possible transition states of the first component of the BEG tree for $N = 20$ and the corresponding effective barrier. In the example, the effective barrier corresponds to $m = 12$ effective up-movements.

Fig. 3.6: The BEG tree: The specified up-and-out barrier B_1 and the effective barrier induced by the first component $S_1^{(N)}$;
$S_1(0) = 20.0$, $\sigma_1 = 0.2$, $T = 1.0$ $r = 0.1$, $B_1 = 33.0$



Similarly, the down-and-out barrier on stock 2 lies between the corresponding binomial process after k effective down jumps and after $k-1$ effective down jumps, i.e.

$$s_{0,2}d_2^{k-1} > B_2 > s_{0,2}d_2^k,$$

where $d_2 = e^{-\sigma\sqrt{T/N}}$ denotes the corresponding downward jump size. Hence, the effective barrier is given by $s_{0,2}d_2^k$, which is in particular *below the specified barrier* B_2.

The location of the effective barriers in relation to the specified barriers indicates that the binomial tree *underestimates the risk of being knocked out. As a result, the tree tends to overestimate the exact price of a knock-out option.*

Furthermore, the oscillations observed are due to the fact that the distance between the effective barriers and the specified barriers varies with grid size. In particular, if the grid size is such that the effective barriers are both close to the corresponding specified barrier, the BEG tree leads to accurate estimates. For the first component, the up-and-out barrier B_1 is just below a layer of horizontal nodes in the tree if N is the largest integer smaller than

$$F_1(m) := \frac{m^2\sigma_1^2 T}{\left(\ln(B_1/s_{0,1})\right)^2}, \qquad m = 1,2,\ldots$$

Similarly, for the second component, the down-and-out barrier B_2 is just above a layer of horizontal nodes in the tree if N is the largest integer smaller than

$$F_2(k) := \frac{k^2\sigma_2^2 T}{\left(\ln(s_{0,2}/B_2)\right)^2}, \qquad k = 1,2,\ldots$$

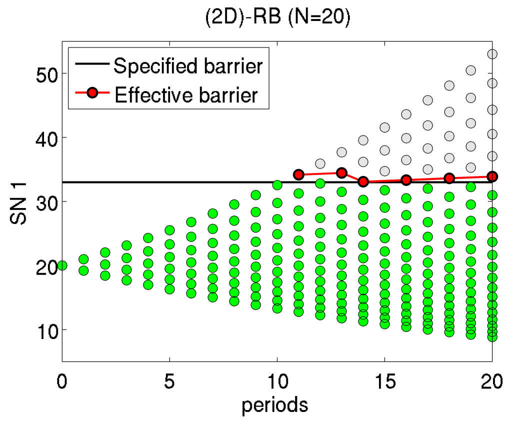
(compare [BL94]). Consequently, we obtain an accurate estimate if the corresponding number of periods N is such that there is some $m = 1,2,\ldots$ and some $k = 1,2,\ldots$ with

$$N = [F_1(m)] = [F_2(k)] .$$

For our numerical example, $N = 500$ is a preferred choice for the number of periods in the discrete model. Then the up-and-out barrier $B_1 = 33$ is just below the corresponding effective barrier (33.0034) and the down-out-barrier $B_2 = 15$ is just above the corresponding effective barrier (14.9995). In the example, the effective barriers correspond to $m = 28$ effective up-movements of the first component and $k = 31$ effective down-movements of the second component. In the above situation, the binomial price obtained for the BEG model (87.4211 £) is already close to the exact price (87.4192 £).

We now discuss the difference in the convergence behaviour observed for the 2D RB tree. By definition, the log-component processes are no longer symmetric around the corresponding starting value. Rather, they are tilted upwards or downwards (depending on the sign of $r - 1/2\sigma_i^2$). As the tilt increases in the number of performed transitions, the corresponding effective barrier is non-constant along the discretised asset path, while the specified barrier is. For our numerical example, Figure 3.7 illustrates the effective barrier for the first component of the 2D RB tree for $N = 20$. Apparently, the effective barrier is non-constant. The fact that the effective barrier is non-constant implies, on the one hand, that there are no specific small values of N for which we get accurate estimates; on the other hand, it also results in oscillations that are of lower amplitude than those obtained for the BEG tree. However, extrapolation methods can be used neither for the BEG tree nor for the RB tree.

Fig. 3.7: The (2D)-RB tree: The specified up-and-out barrier B_1 and the effective barrier induced by the first component $S_1^{(N)}$;
$S_1(0) = 20.0$, $\sigma_1 = 0.2$, $T = 1.0$ $r = 0.1$, $B_1 = 33.0$

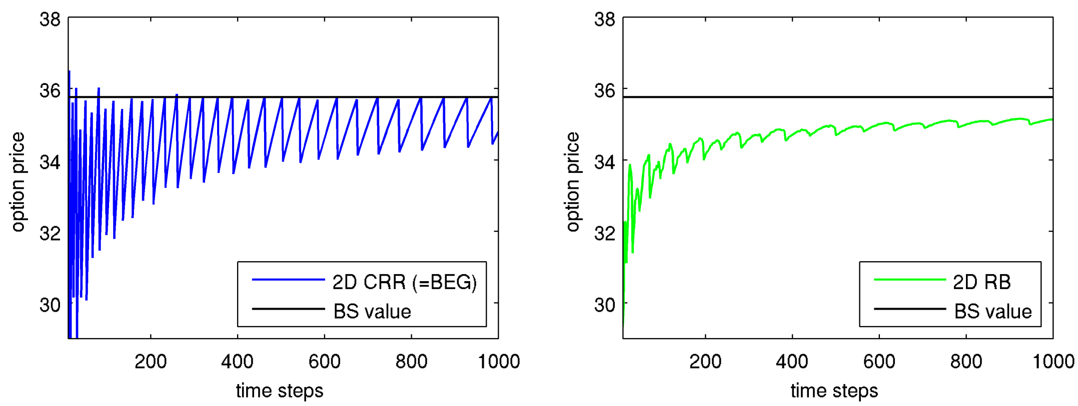


Example 2 We next consider a barrier option with a knock-in barrier on stock 1 and a knock-out barrier on stock 2; i.e. the payoff is given by

$$g(S_1, S_2) = G\,1_{\{S_1(t_0) \geq B_1 \text{ for some } t_0 \in [0,T]\ ,\ S_2(t) \geq B_2\ \forall t \in [0,T]\}}.$$

Figure 3.8 shows the binomial prices obtained for the two models for $N = 10 : 2 : 1000$. Parameters are kept unchanged, except for $B_1 = 25$.

Fig. 3.8: Convergence pattern for a cash-or-nothing option with an up-and-in barrier B_1 on stock 1 and a down-and-out barrier B_2 on stock 2;
$S_1(0) = 20.0$, $S_2(0) = 30.0$, $\sigma_1 = 0.2$, $\sigma_2 = 0.25$, $T = 1.0$ $r = 0.1$, $B_1 = 25.0$, $B_2 = 15.0$, $G = 100$ and **correlation** $\rho = \mathbf{0.5}$

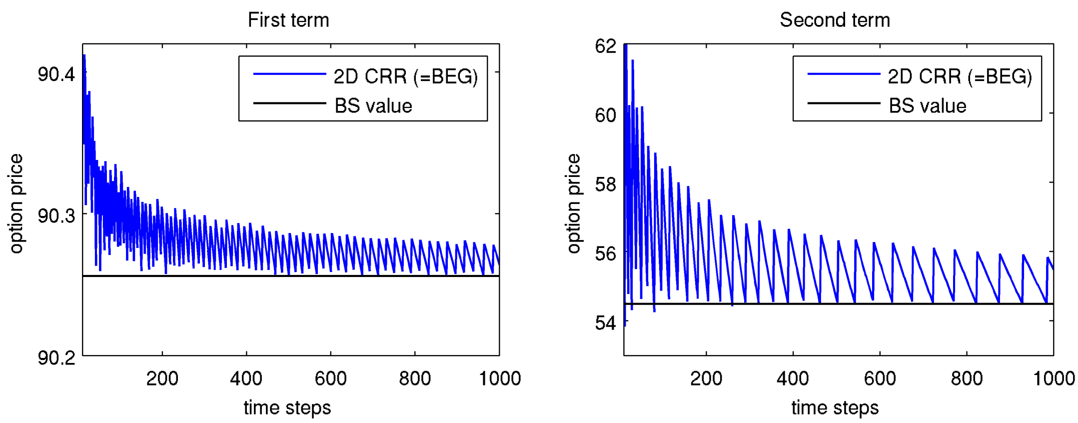


Apparently, the 2D RB tree again leads to a more regular convergence behaviour than the BEG tree. However, convergence remains non-monotone. Unlike the binomial price estimates obtained for a pure knock-out option, the price estimates obtained in the second example are typically smaller than the Black-Scholes price. Let us explain this in decomposing the option price V in the following way:

$$V\left(\text{'Up/in barrier } B_1 \text{ on } S_1\text{','Down/out barrier } B_2 \text{ on } S_2\text{'}\right) =$$
$$V\left(\text{'Down/out barrier } B_2 \text{ on } S_2\text{'}\right) - V\left(\text{'Up/out barrier } B_1 \text{ on } S_1\text{','Down/out barrier } B_2 \text{ on } S_2\text{'}\right) \quad (3.10)$$

Note that the options appearing on the right-hand side of equation (3.10) both exhibit pure knock-out features. Consequently, according to the results in the above, the two tree methods tend to overestimate each term. In total, the price estimates for the option on the left-hand side are typically smaller than the Black-Scholes price because mispricing is largely due to the second term. This is illustrated in Figure 3.9, in which the total error in the BEG price is decomposed according to (3.10).

Fig. 3.9: Convergence pattern for a cash-or-nothing option with an up-and-in barrier B_1 on stock 1 and a down-and-out barrier B_2 on stock 2: Decomposition of the total error



To summarise our numerical results, both for cash-or-nothing options and for the barrier options considered, the BEG tree and the multi-dimensional variant of the RB tree exhibit an irregular convergence behaviour. For the latter, the oscillations are of lower amplitude, yet extrapolation methods cannot be applied.

In the last chapter, we present the decoupling approach as an alternative to the standard approach to multi-dimensional trees investigated above. The decoupling approach is based on a transformation method that results in multi-dimensional binomial trees which are well-defined for an arbitrary correlation structure of the multi-dimensional Black-Scholes model. In addition, it will often yield superior performance of the corresponding trees. In particular, the convergence behaviour is more regular, which can even make the oscillations vanish so that extrapolation methods can be applied.

Chapter 4

The Decoupling Approach to Multi-Dimensional Trees

4.1 Introduction

In this section, we introduce the *decoupling approach* to binomial option pricing in an m-dimensional Black-Scholes setting. The model we suggest contains the 2D example by Hull as a special case [Hul06]. The main idea is to transform the original stock price process S to a process Y with independent component processes *before* the approximating binomial tree is constructed. This allows to define *a multi-dimensional tree that is in principle a product of 1D trees.*

The specific advantages of the decoupling approach are as follows:

- Due to the separation of the correlation structure from the tree structure, it is easy to guarantee non-negative transition probabilities.
- Easy construction of the tree (in particular, we present a special product form that easily allows the enlargement of the tree by a further stock with full use of the tree already constructed).
- The decoupling approach can be combined with any 1D discretisation scheme for the individual stocks. This implies that we can make full re-use of results obtained for 1D trees.
- Excellent numerical performance.

In the context of option pricing, decoupling of correlated processes goes back to Hull and White (1990). They consider the 2D log-asset price $\underline{\phi} = (\phi_1, \phi_2)$ with dynamics

$$\mathrm{d}\phi_i = \alpha_i \mathrm{d}t + \sigma_i \mathrm{d}W_t^i \qquad \text{for } i = 1, 2,$$

where W^1, W^2 are Brownian motions with correlation ρ. Hull and White suggest to transform the original process ϕ to a new process ψ with independent components via

$$\underline{\psi} := \begin{pmatrix} \sigma_2 & \sigma_1 \\ \sigma_2 & -\sigma_1 \end{pmatrix} \underline{\phi}.$$

In Hull and White (1990), this transformation is applied in combination with finite difference methods [HW90]. It is presented prior to 2D tree procedures in Hull's textbook *"Options, Futures and Other Derivatives"* [Hul06]. An extension to higher dimensions is not given.

Clewlow and Strickland (1998) present a transformation method based on the Spectral Theorem; i.e. they gain independence of the component processes by multiplication with a rotation matrix. The appropriate angle is determined by a spectral decomposition of the variance-covariance matrix [CS98]. Clearly, this transformation can easily be extended to arbitrary dimensions. Clewlow and Strickland use the transformation in combination with finite difference schemes. Natcheva (2006) applies their idea to tree procedures. She analyses the impact of a rotation prior to a quadrinomial tree method for pricing contingent claims on the interest rate in a two-factor setting. The interest rates are modelled as Ornstein-Uhlenbeck processes [Nat06]. In this case, the transformed component processes are driven by independent Brownian motions, but they are not mutually independent because they are still coupled via the drift vector. Amin (1991) suggests a discretisation scheme for m correlated assets, where each asset is driven by a vector of independent Brownian motions [Ami91].

In the following, we present a general decoupling method for an m-dimensional Black-Scholes model. The transformations suggested by Hull and White and by Clewlow and Strickland appear as special cases of our general method in $m = 2$.

4.2 A General Decoupling Method

Let us consider the m-dimensional stock price process S following the Black-Scholes dynamics (3.1). In this section, we focus on how to transform the m correlated geometric Brownian motions to independent Brownian motions. First we remove state dependence in the diffusion coefficient via a log-transformation, i.e. we define a process X by $X_t := (\ln(S_t^1), \ldots, \ln(S_t^m))^{\mathrm{T}}$. Its dynamics under Q are given by

$$\mathrm{d}X_i(t) = (r - \tfrac{1}{2}\sigma_i^2)\,\mathrm{d}t + \sigma_i\,\mathrm{d}W_t^i, \quad X_i(0) = \ln(s_{i,o}) \qquad \text{for } i = 1, \ldots, m.$$

Then we decompose the variance-covariance matrix via

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \dots & \rho_{1m}\sigma_1\sigma_m \\ & \ddots & \\ \rho_{1m}\sigma_1\sigma_m & \dots & \sigma_m^2 \end{pmatrix} = GDG^T \tag{4.1}$$

with $G \in \mathbb{R}^{m\times m}$ and $D \in \mathbb{R}^{m\times m}$ diagonal. As Σ is assumed to be positive-definite, it is in particular invertible. Thus, G and D are invertible, and the diagonal elements of D are non-zero. By symmetry, (4.1) is a system of $\frac{1}{2}m(m+1)$ equations and it re-writes elementwise as

$$\sigma_i\sigma_j\rho_{ij} = \sum_{k=1}^{m} g_{ik}d_{kk}g_{jk} \qquad \text{for } i,j = 1,\dots,m. \tag{4.2}$$

As there are m^2+m free parameters, we have an infinite number of solutions and thus also an infinite number of possible decompositions. In particular, spectral decompostion as well as Cholesky decomposition can be used.

Proposition 37. *With the notation $G^{-1} := (g_{ji}^{(-1)})_{j,i=1,\dots,m}$ we introduce $Y := G^{-1}X$, i.e.*

$$Y_j(t) = \textstyle\sum_{i=1}^{m} g_{ji}^{(-1)} X_i(t) \qquad \textit{for } j = 1,\dots,m. \tag{4.3}$$

Its dynamics are given by

$$\mathrm{d}Y_j(t) = \alpha_j\,\mathrm{d}t + \sqrt{d_{jj}}\,\mathrm{d}\bar{W}_t^j, \;\; Y(0) = G^{-1}X(0) \qquad \textit{for } j = 1,\dots,m \tag{4.4}$$

where $\bar{W}_t = (\bar{W}_t^1,\dots,\bar{W}_t^m)^{\mathrm{T}}$ is a vector of independent Brownian motions. The drift vector α is given by

$$\underline{\alpha} = G^{-1}\left(r\underline{1} - \tfrac{1}{2}\underline{\sigma}^2\right)$$

with $\underline{\sigma}^2 := (\sigma_1^2,\dots,\sigma_m^2)^{\mathrm{T}}$.

Proof. Applying Itô's Formula to (4.3) yields $Y(0) = G^{-1}X(0)$ and

$$dY(t) = G^{-1}\left(r\underline{1} - \tfrac{1}{2}\underline{\sigma}^2\right)\mathrm{d}t + \begin{pmatrix} \sum_{i=1}^{m} g_{1i}^{(-1)}\sigma_i \mathrm{d}W_t^i \\ \vdots \\ \sum_{i=1}^{m} g_{mi}^{(-1)}\sigma_i \mathrm{d}W_t^i \end{pmatrix}.$$

For all $j = 1, \dots, m$ we calculate the quadratic variation process:

$$\begin{aligned}
\left\langle \sum_{i=1}^{m} \int g_{ji}^{(-1)} \sigma_i \mathrm{d}W^i \right\rangle_t &= \sum_{i=1}^{m} \sum_{r=1}^{m} \int_0^t g_{ji}^{(-1)} g_{jr}^{(-1)} \sigma_i \sigma_r \left\langle \mathrm{d}W^i, \mathrm{d}W^r \right\rangle_s \\
&= \int_0^t \sum_{i=1}^{m} \sum_{r=1}^{m} g_{ji}^{(-1)} g_{jr}^{(-1)} \sigma_i \sigma_r \rho_{ir} \mathrm{d}s \\
&\overset{(4.2)}{=} \sum_{i=1}^{m} \sum_{r=1}^{m} g_{ji}^{(-1)} g_{jr}^{(-1)} \sum_{k=1}^{m} g_{ik} d_{kk} g_{rk}\, t \\
&= \sum_{k=1}^{m} \left(\sum_{i=1}^{m} g_{ji}^{(-1)} g_{ik} \right) \left(\sum_{r=1}^{m} g_{jr}^{(-1)} g_{rk} \right) d_{kk}\, t \\
&= \sum_{k=1}^{m} \delta_{jk}^2 d_{kk}\, t = d_{jj}\, t
\end{aligned}$$

Thus, by Lévy's Characterisation Theorem for Brownian motion (compare e.g. [KS98], Theorem 3.3.16)

$$\bar{W}_t^j := \frac{1}{\sqrt{d_{jj}}} \sum_{i=1}^{m} \int_0^t g_{ji}^{(-1)} \sigma_i \mathrm{d}W_s^i$$

is a one-dimensional Brownian motion. As for $j \neq i$, we have

$$\begin{aligned}
\left\langle \bar{W}^i, \bar{W}^j \right\rangle_t &= \frac{1}{\sqrt{d_{ii} d_{jj}}} \sum_{k=1}^{m} \sum_{r=1}^{m} g_{ik}^{(-1)} g_{jr}^{(-1)} \sigma_k \sigma_r \rho_{kr}\, t \\
&\overset{(4.2)}{=} \frac{1}{\sqrt{d_{ii} d_{jj}}} \sum_{k=1}^{m} \sum_{r=1}^{m} g_{ik}^{(-1)} g_{jr}^{(-1)} \sum_{l=1}^{m} g_{kl} d_{ll} g_{rl}\, t \\
&= \frac{1}{\sqrt{d_{ii} d_{jj}}} \sum_{l=1}^{m} \left(\sum_{k=1}^{m} g_{ik}^{(-1)} g_{kl} \right) \left(\sum_{r=1}^{m} g_{jr}^{(-1)} g_{rl} \right) d_{ll}\, t \\
&= \frac{1}{\sqrt{d_{ii} d_{jj}}} \sum_{l=1}^{m} \delta_{il} \delta_{jl} d_{ll}\, t = 0,
\end{aligned}$$

Lévy's Theorem also yields independence of the components. □

Note that the Hull and White transformation can easily be embedded into our framework by

$$G^{-1} = \begin{pmatrix} \sigma_2 & \sigma_1 \\ \sigma_2 & -\sigma_1 \end{pmatrix} \qquad \Rightarrow \qquad G = \begin{pmatrix} \frac{1}{2\sigma_2} & \frac{1}{2\sigma_2} \\ \frac{1}{2\sigma_1} & -\frac{1}{2\sigma_1} \end{pmatrix}.$$

Hence, we obtain the decomposition $\Sigma = GDG^T$ with

$$D = \begin{pmatrix} 2(1+\rho)\sigma_1^2\sigma_2^2 & 0 \\ 0 & 2(1-\rho)\sigma_1^2\sigma_2^2 \end{pmatrix}.$$

Decoupling with the spectral decomposition By the Spectral Theorem there is an orthogonal matrix $G \in \mathbb{R}^{m\times m}$ with

$$\Sigma = GDG^T,$$

where $D \in \mathbb{R}^{m\times m}$ is diagonal and each element $d_{jj} =: \lambda_j$ is an eigenvalue of Σ. The j^{th} column of G is given by the corresponding normalised eigenvector. As Σ is symmetric and positive-definite, the eigenvalues are real and positive. By orthogonality we have $G^{-1} = G^T$. According to Proposition 37, the dynamics of the transformed process are

$$\begin{aligned} \mathrm{d}Y_j(t) &= \alpha_j\,\mathrm{d}t + \sqrt{\lambda_j}\,\mathrm{d}\overline{W}_t^j \qquad \text{for } j = 1,\ldots,m, \\ \alpha_j &= \sum_{i=1}^{m} g_{ij}\left(r - \tfrac{1}{2}\sigma_i^2\right) \end{aligned}$$

with $Y(0) = G^T X(0)$. Note that the new diffusion coefficients are the roots of the eigenvalues of the variance-covariance matrix.

Spectral decomposition is not unique. D can obviously be forced to be unique by arranging the eigenvalues in a certain order, e.g. in a non-increasing manner, which we assume from now on. However, G is still not unique because we can choose an arbitrary orthonormal basis of each eigenspace. Given that the eigenvalues are distinct, only the sign of each drift component is not uniquely determined.

In dimension two, there is a simple formula for the eigenvalues of the variance-covariance matrix and G is simply a rotation matrix. Consequently, the dynamics of the transformed process are given explicitly in terms of the variances and the correlation (compare [Nat06]). By contrast, determining the eigenvalues is more involved for higher dimensions. In particular, if the dimension is higher than four, there is in general no closed-form solution for the eigenvalues [1]; the eigenvalues have to be determined by an *iterative* algorithm instead. The QR algorithm is an efficient method to determine *all eigenvalues and eigenvectors at once* [Fra61], [Fra62].

Decoupling with the Cholesky decomposition Since the variance-covariance matrix is symmetric and positive-definite, it admits a unique Cholesky factorisation; i.e. it can

[1] Generally, the roots of polynomial equations higher than fourth degree cannot be written in terms of finite number of operations of addition, subtraction, multiplication, division and root extraction (Abel‘s Theorem (1826)).

be decomposed by

$$\Sigma = GG^T,$$

where $G \in \mathbb{R}^{m\times m}$ is a lower triangular matrix with positive diagonal entries. In contrast to the general decomposition rule (4.1), the transformation matrix is sparse, and the triangular structure will lead to further advantages. Unlike Spectral decomposition, Cholesky factorisation is a direct procedure for any dimension of the problem; i.e. it terminates after a finite number of steps.

By Proposition 37, the process Y has dynamics

$$dY_j(t) = \sum_{k=1}^{j} g_{jk}^{-1}\left(r - \tfrac{1}{2}\sigma_k^2\right) \mathrm{d}t + \mathrm{d}\bar{W}_t^j, \;\; Y_j(0) = \sum_{k=1}^{j} g_{jk}^{-1} \ln(S_j(0)) \tag{4.5}$$

for $j = 1,\dots,m$. Hence, all diffusion coefficients are equal to 1. Moreover, the sums in the drift component and in the starting value run up to the component's index only.

Observe that Cholesky factorisation is distributive and incremental in the following sense. If $A \in \mathbb{R}^{m\times m}$ is symmetric and positive-definite and $A = LL^T$ denotes its Cholesky factorisation, then

$$a_{ij} = \sum_{k=1}^{j} l_{ik} l_{jk} \qquad \text{for } i \geq j.$$

As the diagonal elements are required to be positive,

$$l_{ii} = \sqrt{a_{ii} - \sum_{k=1}^{i-1} l_{ik}^2}.$$

As presented e.g. in Hanke-Bourgeois, the originally m-dimensional problem can be split into two lower-dimensional problems: the problem of finding the Cholesky decomposition of an $(m-1)\times(m-1)$ matrix, plus that of finding the Cholesky decomposition of a positive scalar (compare [HB06], Theorem 5.4). As we see below, Cholesky factorisation is in fact distributive for any dimension $p < m$.

Lemma 6. *Let $A \in \mathbb{R}^{m\times m}$ be symmetric and positive-definite, let $1 \leq p \leq m$ and let A be split into the following blocks*

$$A \;=\; \underbrace{\begin{pmatrix} A_{11} \\ A_{21} \end{pmatrix}}_{p}\!\!\underbrace{\begin{pmatrix} A_{12} \\ A_{22} \end{pmatrix}}_{m-p} \begin{matrix} \}\, p \\ \}\, m-p \end{matrix}$$

Then A_{11} is symmetric and positive-definite. Moreover,

$$S := A_{22} - A_{21}A_{11}^{-1}A_{12} \in \mathbb{R}^{(m-p)\times(m-p)}$$

is well defined and it is also symmetric and positive-definite[2] *(compare e.g. [HB06], Lemma 5.1).*

Proposition 38. *According to Lemma 6, the matrices A_{11} and S can be decomposed by Cholesky factorisation, which will be denoted by $A_{11} = L_pL_p^T$ and $S = L_{m-p}L_{m-p}^{\mathrm{T}}$. Then for the $m \times m$ matrix*

$$M := \begin{pmatrix} L_p & 0 \\ A_{21}(L_p^T)^{-1} & L_{m-p} \end{pmatrix},$$

we have $L = M$.

Proof. M is obviously lower triangular and $MM^T = A$. By positivity of the diagonal elements in L_p and in L_{m-p}, the diagonal elements of M are positive, too. Thus, the assertion follows from uniqueness of the Cholesky decomposition. □

According to Proposition 38, determining the Cholesky decomposition of a size $m \times m$ matrix can be split into two subproblems: to determine the Cholesky decomposition of a size $(m-p) \times (m-p)$ matrix, and to determine the Cholesky decomposition of a size $p \times p$ matrix.

Corollary 13. *The matrix L_p in Proposition 38 is the leading $p \times p$ submatrix of L.*

Corollary 13 implies that the distributive structure goes hand-in-hand with incrementality of the Cholesky algorithm, i.e. the solution of the problem in dimension m already contains the solutions of the problem in dimension $p < m$. Regarding our application, we note that incrementality is inherited to the transformed process:

Proposition 39. *Consider an m-dimensional Black-Scholes model. When using Cholesky factorisation for decoupling, the first $p < m$ components of the decoupled process coincide with the decoupled process corresponding to a Black-Scholes model consisting of the first p stocks only.*

Proof. Let $\Sigma_{11} = G_pG_p^T$ denote the Cholesky factorisation of the leading $p \times p$ submatrix of the variance-covariance matrix. It follows from Proposition 38 that G_p^{-1} is the leading $p \times p$ submatrix of G^{-1}. However, G_p^{-1} is also the transformation matrix in a Black-Scholes world consisting of the first p stocks only. Combining these observations with formula (4.5) for the dynamics of the decoupled process shows that the assertion holds true. □

[2] S is called the Schur complement of A with respect to A_{11}.

Corollary 14. *Consider an m-dimensional Black-Scholes model. Each component Y_i of the transformed model involves original variances and correlations with indices up to i only.*

Note that under decoupling via spectral decomposition, the transformation matrix is not triangular. Hence, there is no analogue to Proposition 39. We analyse the resulting impact on tree methods below.

4.3 Discretisation of the Decoupled Process and Backtransformation

In the following, we suggest a discretisation scheme for the decoupled process

$$Y = G^{-1}X$$

given in equation (4.3). We demonstrate that if we apply a backtransformation to the corresponding tree, we obtain an approximation to the original stock price process, so weak convergence is ensured.

A discrete approximation of the decoupled process In principle, the decoupled process is approximated as follows: After having established independence, each 1D component process Y_i can be approximated separately by a 1D Markov chain $Y_i^{(N)}$ that matches the mean and the variance of the log-returns of Y_i. If we take the product of the measures induced by $Y_1^{(N)}, \dots, Y_m^{(N)}$, we get a law for $Y^{(N)} := (Y_1^{(N)}, \dots, Y_m^{(N)})^T$ which approximates the distribution of Y. In particular, the correlation structure need not be considered. By moment matching, there is weak convergence to each component process Y_i. Furthermore, there is also weak convergence to the m-dimensional process Y, which we can infer from the following theorem:

Theorem 7. *Let M^1, M^2 be separable metrisable topological spaces and let $(P_N^1)_N$, P^1, $(P_N^2)_N$, P^2 be probability measures on $(M^1, \mathscr{B}(M^1))$ and $(M^2, \mathscr{B}(M^2))$, respectively. Then we have*

$$P_N^1 \otimes P_N^2 \Rightarrow_w P^1 \otimes P^2$$

if and only if

$$P_N^1 \Rightarrow_w P^1 \qquad \text{and} \qquad P_N^2 \Rightarrow_w P^2$$

(compare e.g. [Bil68], Theorem 3.2.).

We will see below that as the backtransformation is continuous, we will also obtain weak convergence to the original stock price process S.

Of course, we can use any 1D discretisation scheme for the approximation of the component processes, and the schemes can even differ componentwise. In the following, we suggest approximating each component by a 1D RB scheme. In this case, all paths of the resulting m-dimensional process are equally likely. Let us anticipate that computational effort for backward induction will therefore be optimal for the rate of growth of the tree.

As in Section 3.2,

$$\mathscr{E}_k = \left\{ \underline{\omega}_k = \left(\omega_{k,1}, \ldots, \omega_{k,m}\right) \mid \omega_{k,i} \in \{-1, 1\} \quad \forall i = 1, \ldots, m \right\}$$

denotes the set of all possible up-down-scenarios for period $k \leq N$ and

$$\omega = \left(\omega_{1,1}, ..., \omega_{1,m}, ..., \omega_{N,1}, \ldots, \omega_{N,m}\right) \in \mathscr{E}_1 \times \ldots \times \mathscr{E}_N =: \mathscr{E}$$

denotes the path space. Further $Z_{k,i} : \mathscr{E} \to \{1, -1\}$ is the coordinate mapping. According to the discretisation scheme suggested by RB, we have to define a probability measure $P^{(N)}$ on $\mathscr{E}$ such that

$$Y_k^{(N)} = \begin{pmatrix} Y_{k-1,1}^{(N)} + \alpha_1 \Delta t + Z_{k,1} \sqrt{d_{11}} \sqrt{\Delta t} \\ \vdots \\ Y_{k-1,m}^{(N)} + \alpha_m \Delta t + Z_{k,m} \sqrt{d_{mm}} \sqrt{\Delta t} \end{pmatrix}, \qquad Y_0^{(N)} = Y(0) =: y_0 \tag{4.6}$$

matches the first two moments of the increment $Y(k\Delta t) - Y((k-1)\Delta t)$. Below we denote the one-period transitions of component i by

$$u_i^Y = u_i^Y(N) = \alpha_i \Delta t + \sqrt{d_{ii}} \sqrt{\Delta t} \qquad \text{and} \qquad d_i^Y = d_i^Y(N) = \alpha_i \Delta t - \sqrt{d_{ii}} \sqrt{\Delta t}.$$

Note that there is no correlation structure that needs to be matched because we have already decoupled the originally correlated Brownian motions by the transformation which defines Y. As a consequence, moment matching has to be done componentwise only! We define P_k (and thus the joint distribution of the $Z_{k,i}$, $i = 1, \ldots, m$) as the measure determined by

$$P_k\left(\{\underline{\omega}_k\}\right) := 2^{-m} \qquad \forall \underline{\omega}_k \in \mathscr{E}_k.$$

Then in accordance with the RB discretisation, we have

$$Z_{k,i} = \begin{cases} 1 & \text{with prob. } 1/2 \\ -1 & \text{with prob. } 1/2. \end{cases}$$

Finally, we define the probability measure $P^{(N)}$ on the path space $\mathscr{E}$ as the product of the measures P_k, $k = 1, \ldots, N$. Note that under $P^{(N)}$, the 1-step transition probabilities all are equal to $(1/2)^m$ and each path has probability $(1/2)^{Nm}$.

If the transformation matrix G^{-1} is obtained by Cholesky decomposition, we have the following analogue to Proposition 39 for the discrete model:

Proposition 40. *Suppose the transformation matrix G^{-1} is obtained by Cholesky decomposition. Then the first $p < m$ components of the discrete process $Y^{(N)}$ coincide with the discrete process approximating the decoupled process obtained in a Black-Scholes world consisting of the first p stocks only.*

In order to use the machinery of weak convergence of stochastic processes, we map the discrete process $Y^{(N)}$ to a continuous process $Y^{(c,N)}$ on $[0,T]$ via

$$Y_i^{(c,N)}(t) := Y_{k-1,i}^{(N)} + \tfrac{t-(k-1)\Delta t}{\Delta t} \left(Y_{k,i}^{(N)} - Y_{k-1,i}^{(N)} \right) \qquad \text{for } i = 1, \ldots, m$$

for $t \in [(k-1)\Delta t, k\Delta t]$. Note that the components $\{Y_i^{(c,N)}\}_{i=1,\ldots,m}$ remain independent. Since moment matching is ensured for each component process, we obtain the following result from Donsker's Theorem:

Proposition 41. *We have*

$$Y_i^{(c,N)} \Rightarrow_w Y_i \qquad \textit{for } i = 1, \ldots, m.$$

Finally, as the family $\{Y_i^{(c,N)}\}_{i=1,\ldots,m}$ is independent and each process $Y_i^{(c,N)}$ converges weakly to its continuous counterpart, we can apply Billingsley's Theorem on weak convergence of product measures, which yields the following result:

Corollary 15. *We have*

$$Y^{(c,N)} \Rightarrow_w Y.$$

Apparently, proving weak convergence in m dimensions is just a 1D task under decoupling.

Backtransformation Corollary 15 justifies the approximation of the transformed process Y by a binomial tree based on the discretisation scheme (4.6). Yet since the option payments are defined in terms of the original asset price process S, we have to apply the inverse of the decoupling rule to each node of the tree. More precisely, let $h : \mathbb{R}^m \to \mathbb{R}^m$ denote the inverse of the decoupling rule, i.e.

$$h(\underline{x}) := (e^{G_{1.}\underline{x}}, \dots, e^{G_{m.}\underline{x}})^T \tag{4.7}$$

where $G_{i.} \in \mathbb{R}^{1\times m}$ is the i^{th} row of G. Then $\{h(Y_t)\}_{t\in[0,T]}$ coincides with the original process S. The nodes of the tree are mapped to

$$S_k^{(N)} := h\left(Y_k^{(N)}\right) \qquad \text{for } k = 0, \dots, N, \tag{4.8}$$

and the stochastic process $S^{(c,N)}$ defined as

$$S^{(c,N)}(t) := h(Y^{(c,N)}(t)) \qquad \text{for } t \in [0,T]$$

yields an approximation to the stock price. As the backtransformation is continuous, weak convergence is preserved, i.e. the following result holds:

Proposition 42. *We have*

$$S^{(c,N)} \Rightarrow_w S.$$

Hence, according to Proposition 42, the above model can be applied for numerical option valuation.

Remark 29. *Let us stress that for the decoupling approach, the correlation structure of the continuous-time model enters the discrete model via the transformation map. It affects transition states and consequently possible payoff scenarios. By contrast, we have seen that for the standard tree procedures described in Section 3, the correlations enter transition probabilities. The possible payoff scenarios remain unaltered. The structural difference between the competing methods leads to effects in favour of decoupled trees regarding numerical performance; this is described below.*

4.4 Binomial Option Valuation via the Decoupling Approach

The following section deals with the application of the decoupling approach to numerical option valuation. In Section 4.4.1, we discuss the main aspects of the corresponding tree algorithm. Section 4.4.2 deals with numerical performance of the decoupling approach to multi-dimensional option valuation. The standard methods considered in

Chapter 3 serve as benchmarks. We consider both decoupling with the spectral decomposition (for short: orthogonal tree) and decoupling with the Cholesky decomposition (for short: Cholesky tree).

4.4.1 The Tree Algorithm

Binomial option pricing via the decoupling approach consists of the following basic steps:

Basic Steps: Decoupled Tree Option Pricing

Input: payoff function g; relevant model parameters (in particular, the variance-covariance matrix Σ); number of periods N

1. Decompose the variance-covariance matrix as $\Sigma = GDG^T$ as in (4.1).
2. Transform the stock price S into a new process Y as in (4.3). The new component processes are independent Brownian motions with drift.
3. Set up an m-dimensional RB tree with independent components using the discrete process $Y^{(N)}$ defined in (4.6).
4. Apply the backtransformation (4.7) to *each node of the tree* as in (4.8).
5. Evaluate the payoff functional along the transformed nodes using backward induction. Exploit the fact that all scenarios are equally likely.

Path-independent options For path-independent options, step 4 collapses to

4'. Apply the transformation (4.7) to the *terminal nodes of the tree* only; i.e. compute

$$S_N^{(N)}(\underline{\omega}_N) = h\left(Y_N^{(N)}(\underline{\omega}_N)\right) \qquad \forall\, \underline{\omega}_N \in \mathscr{E}_N.$$

The tree algorithm for the valuation of path-independent options with two underlyings is given below (compare Algorithm 6). Note that by storing intermediate calculation results, the number of multiplications required by backtransformation is reduced from order $m^2(N+1)^m$ to order $m^2(N+1)$.

If we use decoupling with Cholesky factorisation, we have the special form

$$S_N^{(N)} = \left(e^{g_{11}Y_{N,1}^{(N)}}, \dots, e^{\sum_{i=1}^{m-1} g_{m-1,i}Y_{N,i}^{(N)}}, e^{\sum_{i=1}^{m} g_{m,i}Y_{N,i}^{(N)}}\right)^T \tag{4.9}$$

as G is triangular. Note that in this case, $S_{N,i}^{(N)}$ depends on the random variables $Y_{N,j}^{(N)}$ with indices $1 \le j \le i$ only. Consequently, performing backtransformation is faster than in the general case (compare Algorithm 6 for details). Moreover, by Proposition 40 and Corollary 13, we can state the following result:

Proposition 43. *Consider an m-dimensional Black-Scholes model. If we use Cholesky factorisation for decoupling, the first $p < m$ components of the decoupled tree coincide with the decoupled tree corresponding to a Black-Scholes model consisting only of the first p stocks. In particular, the underlying 1D trees are such that the component i involves variances and correlations with indices up to i only.*

Remark 30. *Let us stress that according to Proposition 43, we can re-use the tree already constructed if additional assets enter the market (e.g. when the set of assets underlying a basket option is enlarged).*

As for the methods described in Section 3, the rate of growth for the decoupled tree procedure is $(N+1)^m$. However, the above tree algorithm prices path-independent options more quickly; i.e. it requires less operation counts for a fixed tree size. In fact, computational effort is optimal for the rate of growth of the tree:

Proposition 44. *In the general m-dimensional situation, Algorithm 6 requires computational effort of order $O(N^{m+1})$. The leading constant is $\frac{2^m}{m+1}$.*

Proof. Computing the arithmetic mean at each node of the tree requires 2^m operation counts per node ($2^m - 1$ additions and a *single* multiplication.) Consequently, the total effort for backward induction is of order $O(N^{m+1})$ with constant $\frac{2^m}{m+1}$. Of course, the decoupled tree requires additional operation counts for backtransformation. However for path-independent options, the additional effort does not contribute to the leading term of the total effort. □

Recall that by contrast, for the methods described in Chapter 3, computational effort is also of order $O(N^{m+1})$, but with constant $(2^{m+1}-1)/(m+1)$ for the BEG tree and with constant $(3/2 \times 2^m - 1)/(m+1)$ for the m-dimensional RB tree. Note that the difference in operation counts grows in the dimension of the problem. This is illustrated in the following example: We consider the product option

$$g(S(T)) = \left(\left(\prod_{i=1}^{m} S_i(T)\right)^{\frac{1}{m}} - K\right)^+ .$$

Algorithm 6: The decoupling approach to binomial trees for a path-independent European option with two underlyings

Input: stock price parameters, risk-neutral rate r, payoff function g
Return: price estimate $= V[0][0] \times \exp(-r \times T)$

1. Decomposition of the Variance-Covariance Matrix

Choose a decomposition of the variance-covariance matrix $\Sigma = GDG^T$ as in (4.1).

2. Decoupling

Transform the stock price S into a new process Y as in (4.3).

3. Forward Step

$\{$initialise possible scenarios of $Y^{(N)}$ at maturity$\}$
  Set $\mathrm{YN}[0][1] := y_{0,1} + N \times d_1^Y$
  Set $\mathrm{YN}[0][2] := y_{0,2} + N \times d_2^Y$
for $k = 1$ to N **do**
    $\mathrm{YN}[k][1] := \mathrm{YN}[k-1][1] + u_1^Y - d_1^Y$
    $\mathrm{YN}[k][2] := \mathrm{YN}[k-1][2] + u_2^Y - d_2^Y$
end for

$\{$store intermediate calculation results$\}$
for $k = 0$ to N **do**
    $\mathrm{temp}[1][1][k] := g_{11} \times \mathrm{YN}[k][1]$
    $\mathrm{temp}[1][2][k] := g_{12} \times \mathrm{YN}[k][2]$
    $\mathrm{temp}[2][1][k] := g_{21} \times \mathrm{YN}[k][1]$
    $\mathrm{temp}[2][2][k] := g_{22} \times \mathrm{YN}[k][2]$

  $\{$or under Cholesky factorisation$\}$
    $\mathrm{temp}[1][1][k] := g_{11} \times \mathrm{YN}[k][1]$
    $\mathrm{temp}[2][1][k] := g_{21} \times \mathrm{YN}[k][1]$
    $\mathrm{temp}[2][2][k] := g_{22} \times \mathrm{YN}[k][2]$
end for

4. Backtransformation at Maturity and Forward Step Continued

for $k_1 = 0$ to N **do**
  for $k_2 = 0$ to N **do**
    {backtransformation of possible scenarios of $Y^{(N)}$ at maturity}
    $\mathrm{SN}[k_1][k_2][1] := \exp(\mathrm{temp}\,[1][1][k_1] + \mathrm{temp}\,[1][2][k_2])$
    $\mathrm{SN}[k_1][k_2][2] := \exp(\mathrm{temp}\,[2][1][k_1] + \mathrm{temp}\,[2][2][k_2])$

    {initialise option values at maturity}
    Set $V\,[k_1][k_2] := g\,(\mathrm{SN}\,[k_1][k_2][1], \mathrm{SN}\,[k_1][k_2][2])$
  end for
end for

{or under Cholesky factorisation}
for $k_1 = 0$ to N **do**
  $\mathrm{SN}[k_1][1] := \exp(\mathrm{temp}\,[1][1][k_1])$
  for $k_2 = 0$ to N **do**
    $\mathrm{SN}[k_1][k_2][2] := \exp(\mathrm{temp}\,[2][1][k_1] + \mathrm{temp}\,[2][2][k_2])$

    Set $V\,[k_1][k_2] := g\,(\mathrm{SN}\,[k_1][1], \mathrm{SN}\,[k_1][k_2][2])$
  end for
end for

5. Backward Induction

{step backwards through the tree}
for $k = N - 1$ to 0 **do**
  for $l_1 = 0$ to k **do**
    for $l_2 = 0$ to k **do**

$$V\,[l_1][l_2] := 0.25 \times (V\,[l_1+1][l_2+1] + V\,[l_1+1][l_2] + V\,[l_1][l_2+1] + V\,[l_1][l_2])$$

    end for
  end for
end for

Tables 4.1 and 4.2 show the price estimates obtained for the competing methods and the corresponding computing times. In each row of the tables, computing time required by the BEG tree is set to 100 % and computing time required by the other methods is given as a percentage. Note that, in fact, the option is a single-asset option whose drift rate, initial value and variance are compounded values of the model parameters. Clearly, it admits an explicit valuation formula in the Black-Scholes setting, which allows us to compare the binomial price estimates with the exact value. It is therefore considered as a simple test case.

Table 4.1: Product option ($m = 2$): Accuracy and computing time;
$T = 1$, $r = 0.1$, $\sigma_1 = 0.2$, $\sigma_2 = 0.25$, $\rho = 0.5$, $S_1(0) = 22$, $S_2(0) = 20$ and $K = 20$

N	BEG tree	2D RB tree		Orth tree		Chol tree	
10	3.26143	3.26926	-	3.25587	-	3.26747	-
30	3.2606	3.26369	-	3.26469	-	3.26323	-
50	3.26151	3.26323	-	3.26332	-	3.26241	-
100	3.26181	3.26271	80 %	3.26278	80 %	3.26256	40 %
200	3.26197	3.26243	81 %	3.26246	65 %	3.26223	65 %
300	3.26203	3.26232	79 %	3.26235	63 %	3.26231	57 %
400	3.26204	3.26227	80 %	3.26229	63 %	3.26227	62 %
500	3.26207	3.26225	80 %	3.26227	62 %	3.26221	61 %
1000	3.26210	3.26219	79 %	3.26220	63 %	3.26219	61 %
BS Value	3.26214						

Table 4.2: Product option ($m = 3$): Time/accuracy trade-off;
$T = 1,\ r = 0.1,\ \sigma_1 = 0.2,\ \sigma_2 = 0.25,\ \sigma_3 = 0.15,\ \rho_{12} = 0.5,\ \rho_{13} = -0.2,\ \rho_{23} = -0.4,\ S_1(0) = 22,$ $S_2(0) = 20$, $S_3(0) = 25$ and $K = 20$

N	BEG tree	3D RB tree		Orth tree		Chol tree	
10	3.89311	3.90280	-	3.90251	-	3.90264	-
30	3.90065	3.90379	100 %	3.90375	100 %	3.90381	50 %
50	3.90210	3.90398	86 %	3.90396	68 %	3.90400	46 %
75	3.90282	3.90406	57 %	3.90406	47 %	3.90409	36 %
100	3.90317	3.90412	68 %	3.90411	50 %	3.90413	42 %
125	3.90340	3.90414	65 %	3.90414	46 %	3.90416	39 %
150	3.90353	3.90416	67 %	3.90416	44 %	3.90418	43 %
175	3.90364	3.90418	67 %	3.90418	43 %	3.90419	43 %
200	3.90371	3.90419	67 %	3.90419	42 %	3.90420	37 %
BS Value	3.90427						

For the above example, the alternative methods do not differ significantly in the magnitude of the discretisation error. Yet, as we anticipate from previous results, there is a significant difference in computing time. Apparently, the multi-dimensional RB tree achieves a better time/accuracy trade-off than the BEG tree. However, in accordance with Proposition 44, computing time is further reduced by using decoupled trees. In dimension $m = 2$, the decoupled trees need approximately 60% - 65% of the time required by the BEG tree. In dimension $m = 3$, the computing time is reduced to approximately 40% - 45%. Furthermore, we note that the Cholesky tree is slightly faster than the orthogonal tree, which is due to the fact that it requires less operation counts for the backtransformation of terminal nodes.

While for product options convergence is quite smooth for all methods under consideration, this is not the case for multi-dimensional options with strike levels/barriers on each of the underlying assets. Let us stress that for these types of options, the methods typically differ significantly in the discretisation error. We have seen that for two-asset cash-or-nothing options, the standard tree methods suffer heavily from the sawtooth effect. However for the decoupled tree methods, oscillations in the convergence pattern are dampened. This is explained below.

Path-dependent/American options For path-dependent options, the backtransformation has to be applied to all time-layers in the tree that are relevant for the specific payoff function of interest. In particular, for American options every time-layer of the tree has to be transformed. Consequently, by contrast to the valuation of path-independent options, the computational effort required for backtransformation is typi-

Algorithm 7: The decoupling approach to binomial trees for a European knock-out option with two underlyings

Input: stock price parameters, risk-neutral rate r, payoff function g
Return: price estimate $= V[0][0] \times \exp(-r \times T)$

1. Decomposition of the Variance-covariance Matrix
- 4. Backtransformation at Maturity and Forward Step

{remains unchanged}

5. Backward Induction

{step backwards through the tree}
for $k = N - 1$ to 0 **do**

for $l = 0$ to k **do**
{adjust the YN-array to the current time step}
$\text{YN}[l][1] := \text{YN}[l][1] - d_1^Y$
$\text{YN}[l][2] := \text{YN}[l][1] - d_2^Y$

{store intermediate calculation results}
$\text{temp}[1][1][l] := g_{11} \times \text{YN}[l][1]$
$\text{temp}[1][2][l] := g_{12} \times \text{YN}[l][2]$
$\text{temp}[2][1][l] := g_{21} \times \text{YN}[l][1]$
$\text{temp}[2][2][l] := g_{22} \times \text{YN}[l][2]$
end for

for $l_1 = 0$ to k **do**
for $l_2 = 0$ to k **do**

{backtransformation of possible scenarios of $Y^{(N)}$ at the current time step}
$\text{SN}[l_1][l_2][1] := \exp(\text{temp}[1][1][l_1] + \text{temp}[1][2][l_2])$
$\text{SN}[l_1][l_2][2] := \exp(\text{temp}[2][1][l_1] + \text{temp}[2][2][l_2])$

{check whether barrier levels are crossed}
if SN$[l_1][l_2][1] < B_1$ **and** SN$[l_1][l_2][2] < B_2$ **then** {up-and-out}
{assign current option value}

$$V[l_1][l_2] := 0.25 \times (V[l_1+1][l_2+1] + V[l_1+1][l_2] + V[l_1][l_2+1] + V[l_1][l_2])$$

else
$V[l_1][l_2] := 0$
end if
end
end
end

cally not negligible. Thus, the total effort is no longer optimal for the rate of growth of the tree. However, as we see below, the fact that decoupled trees are relatively costly in the valuation of path-dependent options can be overcompensated for by advanced performance properties. In particular, for barrier options, we can exploit benefits due to monotonicity of convergence.

As an example, the tree algorithm for a European knock-out option with two underlyings can be found below (compare Algorithm 7).

High-dimensional options As discussed previously, tree methods are currently not suitable for the valuation of high-dimensional options. Yet the decoupling approach is perfectly suited to restrict these valuation problems to the ”important dimensions”, which may provide a fast first guess on the option price. This aspect of the decoupling approach is sketched in the following.

Filtering out important factors or important dimensions by a principal component analysis (for short: PCA) is a well-known method in statistics or in high-dimensional numerical integration. We anticipate that it may also be fruitful to apply such a method to numerical valuation of multi-asset options. Of course, the main motivation is a high correlation between certain stocks or submarkets. Moreover, the dynamics of stock markets (or interest rate markets) can often be explained by a relatively small number of random factors (i.e. by the dimension of the underlying Brownian motion) that is less than the number of traded stocks. In such a situation, it seems reasonable to value an option on a big basket of assets by a tree of lower dimension than the number of assets entering it. Since the orthogonal tree is based on a spectral decomposition of the variance-covariance matrix, it is especially suited to that purpose. It essentially considers the underlying independent risk factors (rather than the stocks) as the important ingredients. Moreover, it already incorporates PCA in an implicit way.

Let us recall that under decoupling with spectral decomposition, the dynamics of the transformed process Y are

$$\begin{aligned} \mathrm{d}Y_j(t) &= \alpha_j \mathrm{d}t + \sqrt{\lambda_j}\mathrm{d}\bar{W}_t^j \qquad \text{for } j = 1,\ldots,m \\ Y(0) &= G^T X(0) \end{aligned}$$

with $\lambda_j,\ j = 1,\ldots,m$, the eigenvalues of the variance-covariance matrix. Assume that the variance-covariance matrix is nearly singular. Then there are volatilities that are close to zero; say $\lambda_{m-r+1},\ldots,\lambda_m$. Hence, when we represent the original m-dimensional random object (the stock price S) with respect to its basic driving random factors that are orthogonal to each other, there remain essentially $m-r$ relevant random factors only. Assuming the other factors to be deterministic, the transformed process Y is replaced by some process $\tilde{Y}$ with dynamics

$$\begin{aligned} d\tilde{Y}_j(t) &= \alpha_j\, dt + \sqrt{\lambda_j}\, d\bar{W}_t^j \qquad && \text{for } j = 1, \dots, r, \\ d\tilde{Y}_j(t) &= \alpha_j\, dt && \text{for } j = m-r+1, \dots, m. \end{aligned}$$

Consequently, if we approximate the dynamics of the process $\tilde{Y}$ by an appropriate binomial scheme, the number of possible scenarios at maturity will be reduced from $(N+1)^m$ to $(N+1)^{m-r}$. As a result, computational effort decreases *exponentially in the number of "non-relevant" stochastic factors*. Note that this approach works independently of the particular type of option. Moreover, as PCA is already incorporated implicitly in the orthogonal tree procedure, we do not need to spend any extra effort; we just check for non-relevant stochastic factors.

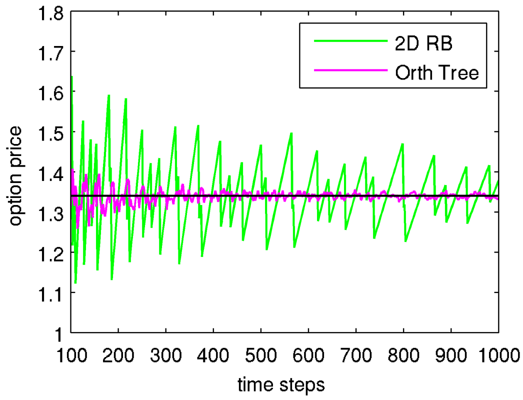
Although the reduction in computational effort is impressive, we should keep in mind that fixing some factors as deterministic leads to less accurate results; in particular, the sequence of price estimates no longer converges to the exact price. However, if the number of relevant stochastic factors is small, the approach can give a *fast first guess on high dimensional valuation problems*. We hence wish to stress the significance of this application for practical purposes. We suggest to analyse this issue for options on prominent indices; this is left for further research.

4.4.2 The Convergence Behaviour of Binomial Option Prices

This section deals with the convergence behaviour of decoupled trees. As in the previous chapter, our analysis is focused on multi-asset options that exhibit discontinuities in the underlyings. The results obtained for standard multi-dimensional trees will serve as benchmarks. We will demonstrate that the decoupling approach leads to a more regular convergence behaviour of the corresponding trees. In particular, for barrier options, convergence can be (approximately) monotone, so that extrapolation methods can be applied.

Cash-or-nothing options We first investigate the convergence behaviour of decoupling trees for cash-or-nothing options. This illustrated with our example from the previous chapter (parameters are kept unchanged). Figure 4.1 shows the corresponding price estimates obtained for the orthogonal tree for $N = 100 : 2 : 1000$. The 2D RB tree is used as a benchmark. The Cholesky tree is considered below.

Fig. 4.1: The orthogonal tree: Convergence pattern for a two-asset cash-or-nothing call
$S_1(0) = 12.0$, $S_2(0) = 12.0$, $\sigma_1 = 0.2$, $\sigma_2 = 0.25$, $\rho = 0.5$, $T = 1.0$ $r = 0.1$, $K_1 = 17.0$, $K_2 = 20.0$, $G = 100$



Apparently, the amplitude of the oscillations is significantly reduced by decoupling. Let us explain: As discussed in the above, the correlation structure of the continuous-time model affects the position of terminal nodes in the orthogonal tree (compare Remark 29). In the backtransformation h, the nodes are dislocated in such a way that the rectangular grid structure is destroyed; i.e. in contrast to standard multi-dimensional trees, there are no longer ”columns” and ”rows” that are parallel to the strike values (compare Figure 4.2). Consequently, the fraction of nodes in the in-the-money region is more stable in N than under standard methods, so that *the orthogonal tree is automatically ”in shape”*. As a result, oscillations in the convergence pattern are dampened.

Fig. 4.2: The orthogonal tree: Realisations of $(S_1^{(N)}, S_2^{(N)})$;
$S_1(0) = 12.0$, $S_2(0) = 12.0$, $\sigma_1 = 0.2$, $\sigma_2 = 0.25$, $\rho = 0.5$, $T = 1.0$ $r = 0.1$, $K_1 = 17.0$, $K_2 = 20.0$, $G = 100$

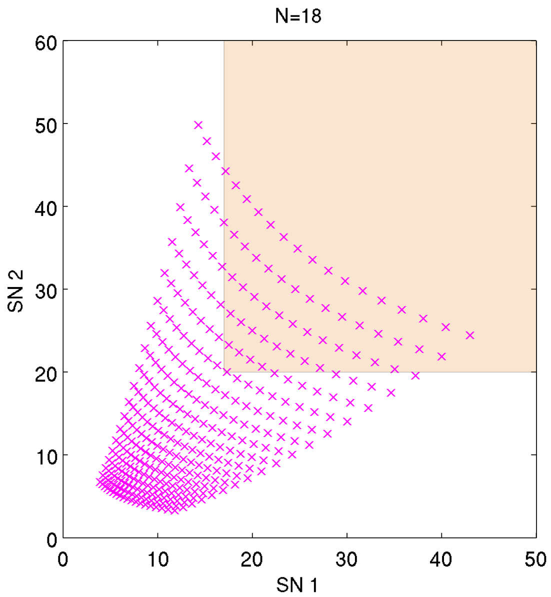
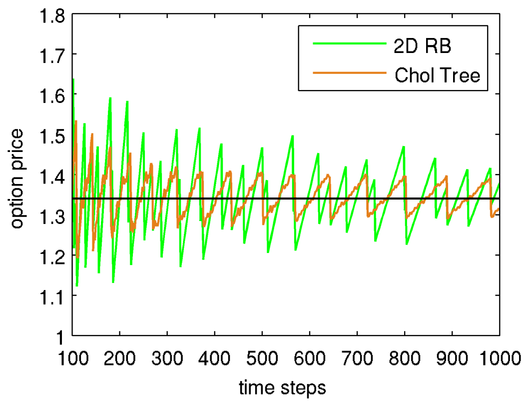


Figure 4.3 shows the price estimates obtained for the Cholesky tree for $N = 100 : 2 : 1000$. Apparently, the performance of the Cholesky tree is better than that of the 2D RB tree, but it is not as good as that of the orthogonal tree. In particular, the sawtooth pattern is still present; yet with a lower amplitude of the oscillations.

Fig. 4.3: The Cholesky tree: Convergence pattern for a two-asset cash-or-nothing call $S_0 = (12,12)^T$, $\sigma_1 = 0.2$, $\sigma_2 = 0.25$, $\rho = 0.5$, $r = 0.1$, $K = (17,20)^T$ and $G = 100$



For the Cholesky tree, the backtransformation h dislocates the terminal nodes in such a way that the rectangular grid structure is only partially destroyed: By backtransformation, the second component of the terminal value $S_{N,2}^{(N)}$ depends on both $Y_{N,1}^{(N)}$ and $Y_{N,2}^{(N)}$, while the first component of the terminal value $S_{N,1}^{(N)}$ depends exclusively on $Y_{N,1}^{(N)}$ (compare equation (4.9)). Consequently, when we fix a possible realisation of $S_{N,1}^{(N)}$, there are $N+1$ possible realisations of $S_{N,2}^{(N)}$, but not the other way around. As a result, we obtain a ”columnwise grid”, i.e. the probability mass is smeared relative to the strike on stock S_2, but it is concentrated in bunches of $N+1$ nodes relative to the strike on stock S_1 (compare Figure 4.4). Hence, the structure of the Cholesky tree has features of both the multi-dimensional RB tree and the orthogonal tree. Clearly, for a cash-or-nothing option with a single strike on stock S_1, the convergence pattern of the Cholesky tree will be similar to that of the 2D RB tree, while it will be similar to that of the orthogonal tree for an option with a single strike on stock S_2.

Fig. 4.4: The Cholesky tree: Realisations of $(S_1^{(N)}, S_2^{(N)})$;
$S_1(0) = 12.0$, $S_2(0) = 12.0$, $\sigma_1 = 0.2$, $\sigma_2 = 0.25$, $\rho = 0.5$, $T = 1.0$ $r = 0.1$, $K_1 = 17.0$, $K_2 = 20.0$, $G = 100$

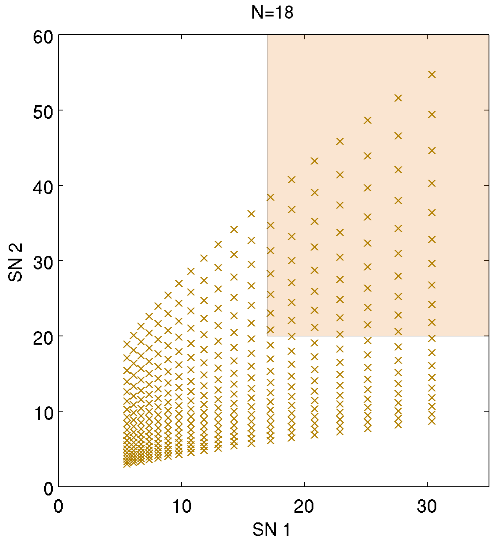


Table 4.3 illustrates accuracy and computing time for the valuation of the two-asset cash-or-nothing option with the alternative methods. We see that the decoupling approach leads to better results in less time. In accordance with the results above, the orthogonal tree performs best.

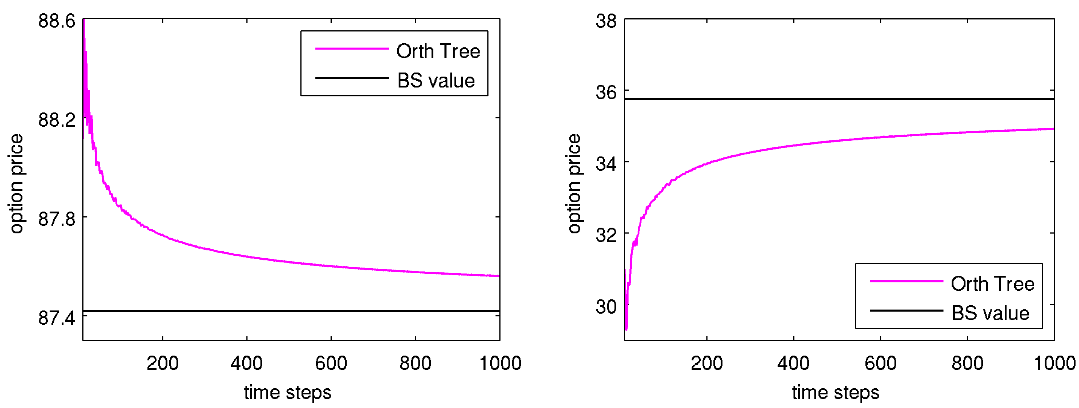
Table 4.3: Two-asset cash-or-nothing option: Accuracy and computing time;
$S_1(0) = 12.0$, $S_2(0) = 12.0$, $\sigma_1 = 0.2$, $\sigma_2 = 0.25$, $\rho = 0.5$, $T = 1.0$ $r = 0.1$, $K_1 = 17.0$, $K_2 = 20.0$, $G = 100$

N	BEG tree	2D RB tree		Orth tree		Chol tree	
50	1.02019	1.27041	-	1.38673	-	1.41077	-
100	1.25755	1.57154	75 %	1.31009	75 %	1.35354	50 %
200	1.27876	1.33817	83 %	1.31146	66 %	1.33912	62 %
300	1.31285	1.31208	82 %	1.34208	64 %	1.35734	63 %
400	1.31009	1.36935	78 %	1.33433	63 %	1.32219	62 %
500	1.40518	1.46723	81 %	1.34317	63 %	1.39960	62 %
700	1.21216	1.35047	82 %	1.34315	63 %	1.36423	62 %
1000	1.31603	1.37889	82 %	1.33373	62 %	1.31235	62 %
BS Value	1.34087						

Barrier options This paragraph deals with the convergence behaviour of decoupled trees for barrier options. This is again illustrated with our examples from the previous chapter. As seen in Chapter 3, the corresponding prices obtained for standard multi-dimensional tree methods suffer heavily from an irregular convergence behaviour. For the decoupling approach, the valuation of these options requires applying the back-transformation (4.7) to every time-layer of the tree. This affects, on the one hand, that computing time is increased, but, on the other hand, the probability mass is smeared for every period of the discrete-time model. This induces an *averaging effect* on the effective barriers. As a result, the decoupling approach can lead to monotone convergence. As we explain next, the additional computational effort is overcompensated for by the benefits due to monotonicity of convergence.

Figure 4.5 shows the price estimates obtained for the orthogonal tree for $N = 10:2:1000$. Apparently, convergence is (approximately) monotone, so that extrapolation methods can be applied. Let us remark that as discussed for cash-or-nothing options, the smoothing effect will be weaker for the Cholesky tree.

Fig. 4.5: Convergence pattern for a barrier option with an with an up-and-out barrier B_1 on stock 1 and a down-and-out barrier B_2 on stock 2 (left) / up-and-in barrier B_1 on stock 1 and a down-and-out barrier B_2 on stock 2 (right)
$S_1(0) = 20.0$, $S_2(0) = 30.0$, $\sigma_1 = 0.2$, $\sigma_2 = 0.25$, $T = 1.0$ $r = 0.1$, $B_1 = 33.0$ / $B_1 = 25.0$, $B_2 = 15.0$, $G = 100$ and correlation $\rho = 0.5$

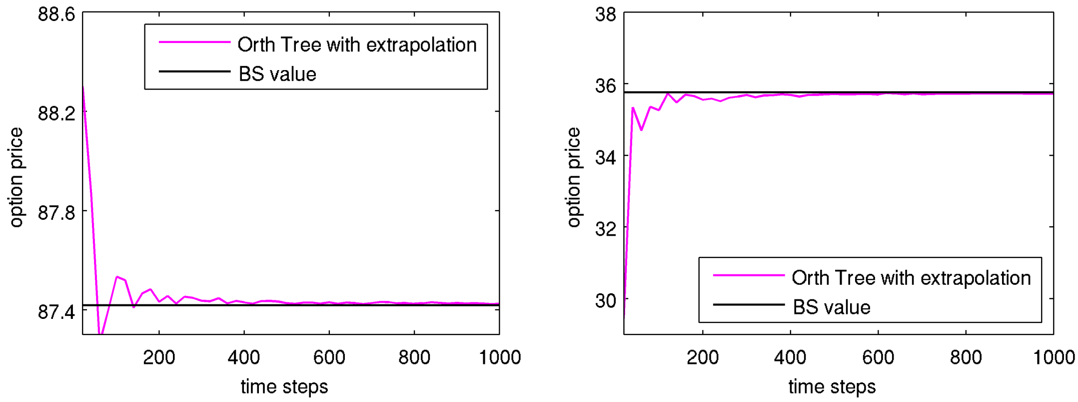


In accordance with the Berry-Esséen inequality, Figure 4.5 suggests that the sequence of price estimates obtained by the orthogonal tree converges in order $1/\sqrt{N}$. Then Richardson extrapolation leads to a sequence of aggregated price estimates given by

$$\hat{C}(2N) = \frac{\sqrt{2}C(2N) - C(N)}{\sqrt{2}-1}.$$

Figure 4.6 shows the sequence of aggregated price estimates for $2N = 20:20:1000$.

Fig. 4.6: Convergence pattern for a barrier option with an with an up-and-out barrier B_1 on stock 1 and a down-and-out barrier B_2 on stock 2 (left) / up-and-in barrier B_1 on stock 1 and a down-and-out barrier B_2 on stock 2 (right): Aggregated Estimates



We see that the improvement in accuracy is enormous. Yet, the total effort has multiplied compared to standard procedures (compare Tables 4.4 and 4.5). However, the results indicate that the aggregated estimates achieve a superior time/accuracy trade-off: The estimates obtained by the 2D RB tree are consistently far above/below the exact price. For $N = 2500$, the prices obtained for 2D RB tree are 87.497 £ for the first example (exact price=87.4192 £) and 35.3717 £ for the second example (exact price=35.75 £). This corresponds to a relative error[3] of 0.09% and 1.06%, respectively. By contrast, the aggregated estimate obtained for the orthogonal tree already achieves a relative error of 0.009% and 0.11%, respectively, for $N = 500$. Note that computing the aggregated estimate for $N = 500$ requires less than 20 s, while it takes approximately 5 min to run the 2D RB tree with $N = 2500$. Hence, the orthogonal tree clearly outperforms the 2D RB tree.

As discussed in the previous chapter, the BEG tree leads to highly accurate results if the grid size is optimally located in relation to the barriers. In the first example, $N = 500$ is a preferred choice for the number of periods in the discrete model. However, the preferred grid size always depends both on the payoff structure and on the model parameters. By contrast, decoupling *consistently leads to small relative errors without assuming knowledge of the problem under consideration.*

Remark 31.

1. *The results observed rely on the assumption that the barriers are constant in the underlying stocks. In principle, we could think of a payoff structure, for which dislocating nodes by the transformation h leads to more oscillations. However, this is rather a theoretical objection as for traded options the barriers are typically constant in the underlyings.*

[3]The relative error computes as $|\text{Price Estimate} - \text{BS Value}| / \text{BS Value} \times 100$.

Table 4.4: Barrier option with an with an up-and-out barrier B_1 on stock 1 and a down-and-out barrier B_2 on stock 2: Time [4] /accuracy trade-off

N	BEG tree		2D RB tree			Orth tree		Orth tree with extrapolation		
100	88.0402		87.8670		100 %	87.8486	650 %	87.5334		750 %
200	87.7317		87.5988		86 %	87.7267	476 %	87.4326		538 %
300	87.7021		87.6300		91 %	87.6718	500 %	87.4365		564 %
400	87.7647		87.6450		90 %	87.6399	505 %	87.4304		571 %
500	87.4211		87.6133		90 %	87.6167	518 %	87.4273	(16 s)	583 %
700	87.6407		87.5315		90 %	87.5874	522 %	87.4280		592 %
1000	87.6007		87.5292		89 %	87.5605	529 %	87.4247		597 %
1500	87.4416		87.5043		90 %	87.5353	547 %	87.4247		612 %
2000	87.4225		87.5161		92 %	87.5202	554 %			623 %
2500	87.5344	(336 s)	87.4970	(305 s)	91 %	87.5098	552 %	87.4217		621 %
BS Value	87.4192									

[4] platform=Toshiba Satellite notebook; machine=Intel Centrino Duo processor, 1.6 GHz, 1.0 GB RAM; operating system=Linux; source=C++; compiler=g++-4.0.1.

Table 4.5: Barrier option with an with an up-and-in barrier B_1 on stock 1 and a down-and-out barrier B_2 on stock 2: Time [4] /accuracy trade-off

N	BEG tree		2D RB tree			Orth tree		Orth tree with extrapolation		
100	32.4044		33.7137		100 %	33.2753	700 %	35.2636		800 %
200	35.2165		34.1574		95 %	33.9422	595 %	35.5523		653 %
300	34.1478		34.5878		89 %	34.2597	583 %	35.6879		673 %
400	34.2484		34.7124		90 %	34.4525	607 %	35.6845		684 %
500	35.6530		34.6848		91 %	34.5847	631 %	35.7098	(19 s)	696 %
700	34.9990		34.9840		89 %	34.7586	640 %	35.7047		700 %
1000	34.7874		35.1509		90 %	34.9177	632 %	35.7219		710 %
1500	34.8805		35.2693		90 %	35.0680	637 %	35.7312		718 %
2000	35.6708		35.2334		91 %	35.1600	647 %	35.7450		729 %
2500	35.5809	(329 s)	35.3717	(298 s)	91 %	35.2225	646 %	35.7455		728 %
BS Value	35.7500									

2. *In contrast to the optimal drift model suggested for a single asset option, the decoupling approach does not require adjusting the location of nodes to the parameter setting or to the option type of interest. We therefore claim that decoupling is an easy and universal recipe to cope with the sawtooth effect in multiple dimensions. Of course, we do not claim that the corresponding trees perform best for every particular type of (exotic) option; but in contrast to more complicated multi-dimensional models as that suggested by Kamrad and Ritchken, decoupled trees show superior performance compared to standard methods without increasing the complexity of the model.*

4.5 Conclusion

To conclude, let us summarise the major advantages of the decoupling approach:

- Non-negativity of transition probabilities can be ensured independently of the correlation structure. Hence, there is no restriction on the applicability of the method regarding the parameter setting. Thus, decoupled trees have a *broader range of application*.
- Decoupling can be used to construct a tree in which *every path has the same probability*.
- If the payoff functional exhibits discontinuities in the underlyings, oscillations in the sequence of price estimates can be dampened significantly by decoupling. Thus, decoupled trees often exhibit *a more regular convergence behaviour*.
- Due to the decoupling of the components, one can use different underlying 1D trees for individual components of a decoupled tree. As an extreme case, one can use a constant (!) for those components that show nearly no variation. Thus, decoupled trees offer *the possibility to give a fast first guess on high-dimensional valuation problems*.
- When applying decoupling with the Cholesky decomposition, it is possible to re-use the original tree if additional assets enter the market. Thus, decoupled trees are *more flexible*.

Bibliography

[AIL04] F. AitSahlia, L. Imhof, and T.L. Lai. Pricing and Hedging of American Knock-in Options. *Journal of Derivatives*, 11:44–50, 2004.

[AK94] K. Amin and A. Khanna. Convergence of American Option Values from Discrete- to Continuous-Time Financial Models. *Mathematical Finance*, 4:289–304, 1994.

[Ami91] K.I. Amin. On the Computation of Continuous Time Option Prices Using Discrete Approximations. *Journal of Financial and Quantitative Analysis*, 26:477–495, 1991.

[AS72] M. Abramowitz and I.A. Stegun, editors. *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*. Dover, New York, US, 10th edition, 1972.

[Bau92] H. Bauer. *Maß- und Integrationstheorie*. de Gruyter, Berlin, Germany, 2nd edition, 1992.

[BAW87] G. Barone-Adesi and R. Whaley. Efficient Analytic Approximation of American Option Values. *Journal of Finance*, 42:301–320, 1987.

[BEG89] P.P. Boyle, J. Evnine, and S. Gibbs. Numerical Evaluation of Multivariate Contingent Claims. *Review of Financial Studies*, 2:241–250, 1989.

[Bil68] P. Billingsley. *Convergence of Probability Measures*. John Wiley & Sons, New York, US, 1968.

[Bjö04] T. Björk. *Arbitrage Theory in Continuous Time*. Oxford University Press, Oxford, UK, 2nd edition, 2004.

[BL94] P.P. Boyle and S.H. Lau. Bumping up Against the Barrier with the Binomial Method. *Journal of Derivatives*, 1:6–14, 1994.

[Boy88] P.P. Boyle. A Lattice Framework for Option Pricing with Two State Variables. *Journal of Financial and Quantitative Analysis*, 23:1–12, 1988.

[BS73] F. Black and M.S. Scholes. The Pricing of Options and Corporate Liabilities. *Journal of Political Economy*, 81:637–654, 1973.

[BS78] M.J. Brennan and E.S. Schwartz. Finite Difference Methods and Jump Processes Arising in the Pricing of Contingent Claims: A Synthesis. *Journal of Financial and Quantitative Analysis*, 13:461–474, 1978.

[CP07] L.-B. Chang and K. Palmer. Smooth Convergence in the Binomial Model. *Finance and Stochastics*, 11:91–105, 2007.

[CRR79] J.C. Cox, S.A. Ross, and M. Rubinstein. Option Pricing: A Simplified Approach. *Journal of Financial Economics*, 7:229–263, 1979.

[CRS71] Y.S. Chow, H. Robbins, and D. Siegmund. *Great Expectations: The Theory of Optimal Stopping*. Houghton Mifflin, Boston, US, 1971.

[CS98] L. Clewlow and C. Strickland. *Implementing Derivatives Models*. J. Wiley & Sons, Chichester, England, 1998.

[DD04] F. Diener and M. Diener. Asymptotics of the Price Oscillations of a European Call Option in a Tree Model. *Mathematical Finance*, 14:271–293, 2004.

[DKEB95] E. Derman, I. Kani, D. Ergener, and I. Bardhan. Enhanced Numerical Methods for Options with Barriers. Quantitative Strategies Research Notes, Goldman Sachs & Co., May 1995.

[Dur05] R. Durrett. *Probability: Theory and Examples*. Duxburry Advanced Series, Belmont, US, 3 edition, 2005.

[EK86] S.N. Ethier and T.G. Kurtz. *Markov Processes - Characterization and Convergence*. John Wiley & Sons, New York, US, 1986.

[Fra61] J.G.F. Francis. The QR Transformation, Part I. *Computer Journal*, 4:265–271, 1961.

[Fra62] J.G.F. Francis. The QR Transformation, Part II. *Computer Journal*, 4:332–345, 1962.

[HB06] M. Hanke-Bourgeois. *Grundlagen der Numerischen Mathematik*. B.G. Teubner Verlag, Wiesbaden, Germany, 2$^{\text{nd}}$ edition, 2006.

[He90] H. He. Convergence from Discrete- to Continuous-Time Contingent Claim Prices. *Review of Financial Studies*, 3:523–546, 1990.

[HK79] J.M. Harrison and D.M. Kreps. Martingales and Arbitrage in Multiperiod Securities Markets. *Journal of Economic Theory*, 20:381–408, 1979.

[HKR98] H. He, W.P. Keirstead, and J. Rebholz. Double Lookbacks. *Mathematical Finance*, 8:201–228, 1998.

[HP81] J.M. Harrison and S.R. Pliska. Martingales and Stochastic Integrals in the Theory of Continuous Trading. *Stochastic Processes and Their Applications*, 11:215–260, 1981.

[Hul06] J. Hull. *Options, Futures, and Other Derivatives*. Pearson/Prentice Hall, Englewood Cliffs, US, 6th edition, 2006.

[HW90] J. Hull and A. White. Valuing Derivative Securities Using the Explicit Finite Difference Method. *Journal of Financial and Quantitative Analysis*, 25(1):87–100, 1990.

[HW94] J. Hull and A. White. Numerical Procedures for Implementing Term Structure Models II: Two-Factor Models. *Journal of Derivatives*, 2:37–48, 1994.

[HZ00] S. Heston and G. Zhou. On the Rate of Convergence of Discrete-Time Contingent Claims. *Mathematical Finance*, 10:53–75, 2000.

[Kal01] O. Kallenberg. *Foundations of Modern Probability*. Springer, New York, US, 2nd edition, 2001.

[KK01] R. Korn and E. Korn. *Optionsbewertung und Portfoliooptimierung*. Vieweg, Wiesbaden, Germany, 2nd edition, 2001.

[KM09a] R. Korn and S. Müller. The Decoupling Approach to Binomial Pricing of Multi-Asset Options. *Journal of Computational Finance*, 12:1–30, 2009.

[KM09b] R. Korn and S. Müller. Getting Multi-Dimensional Trees into a New Shape. *Wilmott Journal*, 1:145–153, 2009.

[KR91] B. Kamrad and P. Ritchken. Multinomial Approximating Models for Options with k State Variables. *Management Science*, 37:1640–1652, 1991.

[KS98] I. Karatzas and S.E. Shreve. *Brownian Motion and Stochastic Calculus*. Springer, New York, US, 2nd edition, 1998.

[Kus77] H.J. Kushner. *Probability Methods for Approximations in Stochastic Control and for Elliptic Equations*. Academic Press, New York, US, 1977.

[Lei98] D.P.J. Leisen. Pricing the American Put Option: A Detailed Convergence Analysis for Binomial Models. *Journal of Economic Dynamics and Control*, 22:1419–1444, 1998.

[LR96] D.P.J. Leisen and M. Reimer. Binomial Models for Option Valuation - Examining and Improving Convergence. *Applied Mathematical Finance*, 3:319–346, 1996.

[Mac86] L.W. MacMillan. Analytic Approximation for the American Put Price. *Advances in Futures and Options Research*, 1:119–139, 1986.

[Nat06] K.P. Natcheva. *On Numerical Pricing Methods of Innovative Financial Products*. PhD thesis, TU Kaiserslautern, 2006.

[Pet75] V.V. Petrov. *Sums of Independent Random Variables*. Springer, Berlin, Germany, 1975.

[RB79] R.J. Rendleman and B.J. Bartter. Two-State Option Pricing. *Journal of Finance*, 34:1093–1110, 1979.

[RS98] L.C.G. Rogers and E.J. Stapleton. Fast Accurate Binomial Pricing. *Finance and Stochastics*, 2:3–17, 1998.

[Rub92] M. Rubinstein. Guiding Force, in *From Black-Scholes to Black Holes: New Frontiers in Options*. *Risk Magazine*, November 1992.

[Ŝ88] A.N. Ŝirjaev. *Wahrscheinlichkeit*. VEB Deutscher Verlag der Wissenschaften, Berlin, Germany, 1988.

[Sha79] W.F. Sharpe. *Investments*. Prentice-Hall, Englewood Cliffs, US, 1979.

[Sta04] M. Staunton. From Top Dog to Straw Man. *Wilmott Magazine*, January/February 2004.

[Sta05] M. Staunton. *Efficient Estimates for Valuing American Options: in The Best of Wilmott 2*. John Wiley & Sons, Chichester, UK, 2005.

[Tia99] Y.S. Tian. A Flexible Binomial Option Pricing Model. *Journal of Future Markets*, 19:817–843, 1999.

[Usp37] J.V. Uspensky. *Introduction to Mathematical Probability*. McGraw-Hill, New York, 1937.

[Zha98] P. G. Zhang. *Exotic Options: A Guide to Second Generation Options*. World Scientific, Singapore, 2^{nd} edition, 1998.

Education

03/12/1981	born in Mayen
22/06/2001	Abitur at Megina Gymnasium Mayen, with distinction
07/2001–09/2001	internship, Rasselstein GmbH, ThyssenKrupp Steel Group
23/09/2003	Vordiplom in Management Mathematics
	(”Wirtschaftsmathematik”) at the University of Kaiserslautern
09/2003–06/2004	Erasmus student at the University of Warwick, UK
08/2004–09/2004	internship, WestLB AG
04/2005–07/2005	teaching assistant in mathematical finance
07/2005–08/2005	AARMS summer student at Dalhousie University, Canada
11/2005–07/2008	member of the Rheinland-Pfalz cluster of excellence DASMOD
08/03/2006	Diplom in Management Mathematics (Dipl.-Math. oec.)
	at the University of Kaiserslautern,
	specialisation in mathematical finance, top grade
since 04/2006	PhD student of Prof. Dr. Ralf Korn,
	teaching assistant in mathematical finance
02/2008–03/2008	
and 01/2009	visiting researcher at the University of Cambridge, UK
since 07/2008	member of the Rheinland-Pfalz Center for Computational
	and Mathematical Modelling