Technische Universität Kaiserslautern

MULTILEVEL CONSTRUCTIONS

ANTON KOSTIUK

Supervised by

PROF. DR. RALF KORN

Examination date: October 13, 2014

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

D386

Anton Kostiuk *"Multilevel Constructions"*, May 2014

ACKNOWLEDGEMENTS

I sincerely thank my scientific adviser Prof. Dr. Ralf Korn for his superb supervision and support, both scientific and personal, throughout my PhD study.

On a personal level, I am grateful to my family and my friends, and especially to my wife Anastasiia Konstantynova, for their love and support.

Further, I want to thank my colleges and friends from the Financial Mathematics, Microelectronic Systems Design Research and Computational Stochastics groups for a productive collaboration and a good working atmosphere.

Finally, I gratefully acknowledge the financial support of the state research Center for Mathematical and Computational Modeling, $(\mathrm{CM})^2$.

Anton Kostiuk, May 2014

CONTENTS

1 INTRODUCTION 1
  1.1 Option Pricing Problem 1
    1.1.1 Examples of Some Options Contracts 2
    1.1.2 Modelling 4
    1.1.3 Pricing As A Numerical Problem 7
  1.2 Monte Carlo 9
    1.2.1 Basics of Monte Carlo Method 9
    1.2.2 Monte Carlo In SDE Settings 10
  1.3 Multilevel Monte Carlo 17
  1.4 Thesis Structure 19
    1.4.1 Structure 19
    1.4.2 Notation Convention 20
2 MLMC OPTIMISATION VIEW 21
  2.1 Multilevel Construction 21
    2.1.1 Optimisation 23
    2.1.2 MLC Algorithm 30
    2.1.3 Computational Cost of the MLC estimate 31
    2.1.4 Comment on the Structural Result 33
  2.2 MLC as Variance Reduction 34
    2.2.1 Numerical Example 35
  2.3 MLC as the classic MLMC 37
3 WEAK EXTRAPOLATION AND MLMC 41
  3.1 Weak Extrapolation 43
    3.1.1 Linear Talay-Tubaro Extrapolation 43
    3.1.2 Classic High Order Weak Extrapolation 44
    3.1.3 Literature Survey 47
  3.2 Weak Extrapolation MLMC 47
    3.2.1 HOWE in MLMC settings 47
    3.2.2 Weak Extrapolation Multilevel Monte Carlo 50
  3.3 Numerical Algorithm 54
  3.4 Computational Cost Theorem 55
  3.5 Numerical Performance Analysis 62
    3.5.1 Black Scholes Model 64
    3.5.2 Heston Model 75
A OPTIMISATION 79
  A.1 Optimal Numbers of Simulated Paths 79
  A.2 Optimal Numbers of Discretisation Point for Each Scheme 80
    A.2.1 Solving System of Equations for n_i 80
    A.2.2 Value at Optima 82
  A.3 Optimal Number of Levels 83
  A.4 Computational Cost of the MLC estimate 86
B WEAK EXTRAPOLATION AND MLMC APPENDIX 95

B.1 Solving System of Equations λ^n 95
B.2 Bias Of The HOWE Scheme 96
B.3 Boundedness of The Weights κ^N 97

BIBLIOGRAPHY 99

LIST OF FIGURES

Figure 1 Optimal $\bar{n}_i$ 28
Figure 2 Optimal common ratio R 30
Figure 3 Performance of the MLC 36
Figure 4 MLC vs plain MC 37
Figure 5 Upper and Lower Bounds for κ_N^N 52
Figure 6 Bias convergence of the Euler scheme (BS model and call option) 65
Figure 7 CC vs. RMSE (BS model and call option) 67
Figure 8 Bias convergence of the Euler scheme (BS model and digital option) 68
Figure 9 CC vs. RMSE (BS model and digital option) 68
Figure 10 Bias convergence of the Euler scheme (BS model and Asian option): n bias plot 71
Figure 11 Bias convergence of the Euler scheme (BS model and Asian option) 72
Figure 12 CC vs. RMSE (BS model and Asian option) 74
Figure 13 Bias convergence of the Euler scheme (BS model and barrier option) 74
Figure 14 Bias convergence of the Euler scheme (Heston model and call option) 76

LIST OF ALGORITHMS

1 Multilevel Constructed Estimate 31
2 Weak Extrapolation Multilevel Monte Carlo 54

LIST OF TABLES

Table 1 Optimal n_i and L 36
Table 2 Common ratio 36
Table 3 MLC vs plain MC 37
Table 4 Values of a squared sum of the extrapolation weights 46
Table 5 Example of approximate values of κ^N 52
Table 6 Comparison of WEML and MLMC estimates 56
Table 7 Bias convergence of the Euler scheme (BS model and call option) 66
Table 8 Speed up of WEML compared to MLMC (BS model and call option) 66
Table 9 Bias convergence of the Euler scheme (BS model and digital option) 69
Table 10 Speed up of WEML compared to MLMC (BS model and digital option) 70
Table 11 Bias convergence of the Euler scheme (BS model and Asian option) 70
Table 12 Bias convergence of the Euler scheme (BS model and Asian option) 73
Table 13 Speed up of WEML compared to MLMC (BS model and Asian option) 73
Table 14 Bias convergence of the Euler scheme (BS model and barrier option) 75
Table 15 Bias convergence of the Euler scheme (Heston model and call option) 77

ACRONYMS

MC Monte Carlo 1
SDE stochastic differential equation 1
MLMC Multilevel Monte Carlo 1
BS Black-Scholes 5
EMM equivalent martingale measure 6
CIR Cox-Ingersoll-Ross 7
QMC Quasi Monte Carlo 9
PDE partial differential equation 9
i.i.d. independent identically distributed 9
ODE ordinary differential equation 11
w.r.t. with respect to 13
CC computational cost 15
MSE mean square error 16
RMSE root mean square error 16
SR Statistical Romberg 17
MLC multilevel constructed 22
RHS right hand side 22
HOWE high order weak extrapolation 42
WEML weak extrapolation multilevel 42
TT Talay-Tubaro 43
LHS left hand side 49

1 INTRODUCTION

One of the main problems in finance nowadays is a problem of pricing exotic derivatives, for example barrier, Asian, basket, window and other options. Typically pricing of such options involves some numerical approximation techniques, as there exists no closed form formulae for them. In this thesis we investigate Monte Carlo (MC) methods and in particular the Multilevel Monte Carlo (MLMC) approach as one of the most promising techniques. We propose an extension of the classic MLMC estimate, and also take a look at the MLMC idea from a new angle.

However, we should add that application of the forthcoming methods is not restricted to the financial area. In general they can be applied to any other problems that target estimation of an expectation of a functional of an stochastic differential equation (SDE) solution.

In the introduction we start with a brief explanation of the pricing problem [42]. Section 1.1.1 introduces several important option contract types, while examples of typical models can be found in Section 1.1.2. As the pricing is not a unique example of application of the MC methods, this will be done omitting some details. The introduction will serve as a motivation and linkage to the more general problem of the estimation of an expectation of a functional of a SDE solution. The linkage and formulation of the general problem will be given in Section 1.1.3.

In Section 1.2 we introduce basics of the MC methods and talk about the MC estimation in the SDE settings. This includes definition of the most basic discretisation schemes and their properties. An idea of the advanced MLMC technique is brought up and discussed in Section 1.3

The introduction concludes with Section 1.4.1, that explains the further structure of the thesis. Moreover the reader can find there some notational convention that is used later in the text.

1.1 OPTION PRICING PROBLEM

An increasing number of complex derivatives is traded every day on stock exchanges all over the world. This creates a huge demand for advanced mathematical methods and models to price them. As a consequence the pricing problem is one of the most important problems in applied mathematics.

In this section, we present the main ideas of stock price modelling and option pricing. Mainly we will be following the book of Korn et al. [45], where this problem is discussed in the MC estimation settings.

1.1.1 *Examples of Some Options Contracts*

Options are derivative contracts, meaning that their payoff depends on a performance of an underlying asset. There exist options on equities, bonds, currencies, commodities, such as oil, metals, etc. In this thesis we will be talking about pricing options written on a stock or commodities.

There exist two important groups of option contracts:

- plain vanilla options: European call and put options;
- exotic options.

The first group are actively traded on an organised exchange floors, e.g. the Chicago Board Options Exchange. These contracts are typically standardised and their prices are agreed via market mechanism. On a contrary the exotic options are typically traded over-the-counter and their prices are subject to individual agreement. Therefore pricing these exotic options will be a key motivation for a forthcoming discussions.

1.1.1.1 *Plain Vanilla Options*

The most basic contracts are European call and put:

EUROPEAN CALL OPTION is a contract that gives its holder the right to *buy* a certain fixed amount of an asset at a specified future time for an already agreed price;

EUROPEAN PUT OPTION is a contract that gives its holder the right to *sell* a certain fixed amount of an asset to the writer of the option at a specified future time for an already agreed price.

In both cases the agreed price is called a *strike price* and the specified future time is called a *maturity time*.

Remark 1. *There exists an American type of call and put contracts.In general, the* American type *of a contract means that its holder can execute her right of selling/buying at any time up to maturity, while in the case of a European type contract it is possible only at the maturity time.*

We are not considering this type of options here, as MC approaches to price this options are essentially different to the case of European contracts [29].

EUROPEAN CALL OPTION

A European call on one share of a stock gives its holder the right to buy this share at time $t = T$ for the strike price $K \geqslant 0$, which was fixed at time $t = 0$. Hence, if the final share price $S(T)$ exceeds K, the holder of the option buys the share for a price of K and can then sell it immediately at the market for the price of $S(T)$. This makes a gain of $S(T) - K$ for the contract holder. On the other hand, if $S(T) < K$, the rational holder does not execute his right to buy the share for the price K. As a consequence, in this case, there is no gain from holding the option. Combining the two cases leads to a final payment of

$$B(S) = (S(T) - K)^+$$

obtained by the holder at the maturity.

EUROPEAN PUT OPTION

The holder of a European put has the right to sell one share at time $t = T$ for the price $K > 0$. Similarly to the case of a European call, it can be shown that the possession of the European put leads to a payment of

$$B(S) = (K - S(T))^+$$

at the maturity.

1.1.1.2 *Examples of Exotic Options*

We proceed further by introducing some important exotic options. In particular we give an example of a path independent exotic option the basket option, and two examples of path dependent Asian and barrier options. Further details and examples can be found in e.g. [45] or [32].

An absence of closed form formulae for an option price is a typical problem of almost all exotic options. Therefore some numerical methods are required to calculate or at least approximate their price. In particular in this thesis we investigate MC methods to approximate a price of such products.

ASIAN OPTION

Asian options have the feature that their payoff depends on averaging over the time of the price of a stock. Therefore counterparts of such a contract are protected against short term market "manipulations", that might occur close to maturity. Here we will consider a continuous *fixed-strike average* call option with the following payoff

$$B(S) = \left(\frac{1}{T} \int_0^T S(t)\, dt - K \right)^+ .$$

We should add that there exist other types of Asian options [45]. Indeed in real traded contracts the continuous-time averaging is replaced by discretised versions, e.g. average daily closing prices during some period of time.

BASKET OPTION

The payment of a *basket* option depends on an average performance of a basket of stock prices. For instance a European basket call has a final payment of

$$B(S) = \left(\sum_{i=1}^{d} w_i S_i(T) - K, 0 \right)^{+} \qquad \text{with} \quad \sum_{i=1}^{d} w_i = 1 \, ,$$

for a basket of d stocks weighted by the weights w_i. This options are typically written on baskets consistent of different types of underlying assents.

BARRIER OPTION

Another example of popular path dependent option is *barrier* options. Their payoff depends on the fact of either hitting or not hitting some pre-specified barrier(s) during the life time of a contract. Basic examples of one sided barrier options are the *knock-out barrier* option given by the final payments

$$\begin{aligned} B_{doc}(S) &= (S(T) - K)^{+} \, \mathbb{I}_{\{S(t) > H, \forall t \in [0,T]\}} && \text{down-and-out call} \\ B_{uoc}(S) &= (S(T) - K)^{+} \, \mathbb{I}_{\{S(t) < H, \forall t \in [0,T]\}} && \text{up-and-out call} \end{aligned}$$

where $K > 0$ is the strike and $H \geqslant 0$ the barrier of the option. Analogously, one introduces *down-and-out* and *up-and-out put* options. Also there exist four *knock-in barrier* option (for details see e.g. [45]).

Remark 2. *The prices of aforementioned in and out barrier options are linked via an 'in-out parity", thus it is sufficient to price only out-options [45].*

Moreover in case of the BS model (see Section 1.1.2) there exxists formulae for price of the one sided barrier options [61].

In general, there are a lot of other barrier options, for instance two sided barrier options or window options, i.e. barrier options with time varying barriers. Moreover, similarity to Asian options, very often barrier options are based on discrete barrier crossing events, e.g. monitoring of the daily closing prices.

1.1.2 *Modelling*

In this section we explain the basics of a stock price modelling and define some important models. When we look at an evolution of a stock price over time we recognise certain remarkable features. They all lead to an obvious ingredient for models: the Brownian motion $W(t)$, for definition and properties consult at [37]. One of the first ideas was to model a stock price evolution using a geometric Brownian motion proposed by Samuelson in [62]. The main breakthrough in the modelling came in the 1970s when the famous Black-Scholes (BS) formulae for the price of European call options and put options were developed by Black and Scholes [10] and by Merton [57].

Nowadays, there exist arguments for considering models for stock prices or interest rates that allow for non-continuous changes, so-called jumps. However we restrict ourselves to the case where the underlying driving uncertainty is modelled by a Brownian motion. In many situations, this makes the analysis of the corresponding problems tractable and suitable for efficient numerics.

Modern modelling approaches typically assume that an underlying stock follows some SDE. The models typically have to satisfy at least the two following requirements:

- existence of closed form formulae for vanilla options or efficient technique to compute their prices;
- they should be able to capture at least some of the aspects of a real market behaviour.

The first requirement is essentially crucial as the pricing of an exotic option consists of the two steps:

1. model calibration to market prices of actively traded vanilla options (or sometimes other liquid securities), this step defines values of the model's parameters;
2. using the model with the calibrated parameters find a price of the exotic option.

We continue our introduction with the BS model. In the sequel we present examples of some modern advanced models, e.g. the Heston model [35].

1.1.2.1 *Black-Scholes Model*

We present the BS model in the most basic one-dimensional setting. It models the dynamics of a stock price as a geometric Brownian motion

$$S(t) = S(0)\ e^{\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W(t)},$$

where μ is a *drift* and σ a *volatility* of the stock. Equivalently, this dynamic can be represented in the form of the corresponding SDE

$$dS(t) = \mu S(t)\,dt + \sigma S(t)\,dW(t)\ .$$

Modern pricing theory [20] tells us that under suitable conditions a fair price of a derivative with a final payoff B can be calculated as an expectation of the discounted payoff under the equivalent martingale measure (EMM), i.e.

$$C = \mathbb{E}\left(e^{-rT} B\left(S\right)\right) \ ,$$

where r is a riskless interest rate and T is time to maturity. Moreover using the Girsanov Theorem (see e.g. [37]) we can rewrite the BS SDE under the EMM as

$$dS\left(t\right) = rS\left(t\right) dt + \sigma S\left(t\right) dW\left(t\right) \ , \tag{1}$$

and the corresponding solution is of the following form

$$S\left(t\right) = S\left(0\right) \ e^{\left(r - \frac{1}{2}\sigma^2\right)t + \sigma W\left(t\right)} \ . \tag{2}$$

It is very important that the EMM evaluations are independent from the physical measure defined by μ. This is one of the reasons that the BS theory became so popular in finance.

Another reason is the existence of the famous BS formulae that give prices of the plain vanilla options in the closed form [10]. As we already pointed out, this is important for efficient parameter calibration to the real market data (for details see e.g. [45]).

However the BS model has an essential drawback: it assumes the volatility to be constant for all maturities and strike prices, what is not the case at modern markets. Therefore there exists a serious demand for models that can capture a non-constant volatility and some other so-called stylised facts [16].

1.1.2.2 *Advanced Models*

During the 1980s the imperfections of the BS model became clear. The two main streams began to develop:

- the class of *local volatility* models, started by Dupire [23] (for more general theory see e.g. [26]);
- the class of *stochastic volatility* models, e.g. the Heston model [35] or the SABR model [31].

Both streams aim to explain the non-constant volatility, i.e. the intensity of the price fluctuations (also known as volatility clustering [16]). Recently some techniques to combine both concepts into *hybrid stochastic-local volatility* models were developed, for instance [60] or [34].

There is another trend in the modern modelling that incorporates jump behaviour into a stock modelling, for instance Bates considers extension of the Heston model with jumps [8]. However models with jumps are out of the scope of this thesis. Moreover MC methods for these models are more sophisticated as they involve simulation of jumps (for an overview see e.g. [55]).

In the upcoming paragraph we will give a brief introduction to the Heston model.

HESTON MODEL

The Heston model [35] is given through a two-dimensional SDE, that has the following form under an EMM

$$dS(t) = rS(t)\,dt + \sqrt{V(t)}S(t)\,dW^S(t) \ , \tag{3a}$$

$$dV(t) = \kappa(\theta - V(t))\,dt + \sigma\sqrt{V(t)}dW^V(t) \ , \tag{3b}$$

where W^S and W^V are correlated driving Brownian motions with correlation ρ. The model for the stock price S is similar to the BS model, but in addition the volatility of the stock V is modelled as a Cox-Ingersoll-Ross (CIR) process [17]. Both stock price S and volatility V processes are coupled via correlated Brownian drivers, what reflects the volatility clustering effect observed on markets, i.e. a big stock price leads to a smaller fluctuation and vice versa. The parameter θ is the long-term limit of the volatility V and called a *mean reversion level* or a *long-term variance*. In addition the parameter κ determines the *speed of reversion* and a *volatility of volatility* (or *vol-of-vol*) is defined by a value of the parameter σ.

Similarly to the BS model a main reason for the success of the Heston model is the existence of semi-closed formulae for the prices of the plain vanilla options [4]. Moreover these formulae are complemented by computation techniques, e.g. Fast Fourier Transform [52] or Fourier-Cosine expansion [24], that allow effective calibration.

Then under suitable conditions a price of a derivative with a final payoff B can be calculated as an expectation of the discounted payoff under EMM

$$C = \mathbb{E}\left(e^{-rT}B(S)\right) \ ,$$

where r is a riskless interest rate, T is the time to maturity and S is a solution of Equation 3.

Remark 3. *By the construction the volatility process* V *is always non-negative. Moreover if its parameters fulfil the so-called* Feller condition *, i.e.,* $\sigma^2 < 2\kappa\theta$*, the process is strictly positive [25]. Also this condition plays a crucial role for the quality of discretisation schemes (see e.g. [2] or [21] and references therein).*

One should keep in mind that this property is often violated on real markets, see e.g. [15]. On top of that the Heston model is exposed to a moment explosion [5].

1.1.3 *Pricing As A Numerical Problem*

To summarise we draw a link between the pricing model and the more general problem of an estimation of an expectation of a functional of an SDE solution.

Modern modelling approaches typically assume that an underlying stock(s) is modelled by some well defined d-dimensional SDE

$$dS(t) = a\left(t, S\left(t\right)\right) dt + \sigma\left(t, s\left(t\right)\right) dW\left(t\right) \qquad S\left(0\right) = s_0\,, \tag{4}$$

driven by an m-dimensional Brownian motion under an EMM.

Remark 4. *Note that some of "stocks" modelled by SDE* (4) *can be non-tradable, as e.g. volatility* V *in the Heston model.*

Moreover we should clearly state that we are leaving out an important question of a market completeness *[42].*

By the well definedness of SDE (4) we mean that it possesses a unique solution, what can be assured for instance by the conditions of the following theorem.

Theorem 1 (Existence and uniqueness of a solution for SDE)**.** *Let the coefficients* $a(t,x)$, $\sigma(t,x)$ *of the SDE* (4) *be continuous functions satisfying both a Lipschitz and a growth condition*

$$\|a(t,x) - a(t,y)\| + \|\sigma(t,x) - \sigma(t,y)\| \leqslant K\,\|x-y\|$$
$$\|a(t,x)\|^2 + \|\sigma(t,x)\|^2 \leqslant K^2\left(1 + \|x\|^2\right)$$

for all $t \geqslant 0$, $x, y \in \mathbb{R}^d$ *and constant* $K > 0$ *(*$\|\cdot\|$ *denotes the Euclidean norm of suitable dimension).*

Then there exists a continuous, strong solution $(S(t), \mathcal{F}_t)_{t\geqslant 0}$ *of SDE* (4) *such that*

$$\mathbb{E}\|S(t)\|^2 \leqslant Ce^{C\,T}\left(1 + \|x\|^2\right) \qquad \forall t \in [0,T]$$

for some constant $C = C(K,T)$ *and* $T > 0$*. Further,* S *is unique up to indistinguishability, i.e. if* Y *would be another solution of Equation 4 then we would have*

$$\mathbb{P}\left(S(t) = Y(t), \forall t \geqslant 0\right) = 1.$$

More results on existence and uniqueness of a solution for SDEs can be found in Section 4.5 of [39].

Under suitable conditions a price of a derivative with a final payoff B is an expectation of the discounted payoff under EMM

$$C = \mathbb{E}\left(e^{-rT} B\left(S\right)\right)\,,$$

where r is a riskless interest rate and T is time to maturity. For example the price of an Asian call option with maturity T and strike price K is equal to

$$C = \mathbb{E}\left(e^{-rT}\left(\frac{1}{T}\int_0^T S\left(t\right) dt - K\right)^+\right)\,,$$

where S is a solution of SDE (4).

Typically in the case of exotic options this expectation does not have a closed form formula and has to be computed using some numerical techniques. Therefore the pricing problem can be seen as a problem of estimating an expectation of a functional of an SDE solution. This problem can be solved using numerical methods, e.g. MC methods, Quasi Monte Carlo (QMC) methods, using Feynman-Kac Representation Theorem [37] the problem can be represented in form of partial differential equation (PDE) that is subject to numerical solution, Binomial Trees (for overview see e.g. [43]), numerous approximation methods, etc.

In this thesis we discuss application of the MC methods to the pricing problem. The choice of the MC methods is motivated by their relative flexibility and generality. In particular they can be used to price any European option contract, including multi-asset options, which typically can be handled only by the MC methods. On the other hand the MC algorithms are known for their relatively low convergence rate and a long run time. But with the development of powerful computational tools as e.g. computer-clusters or their philosophical counterparts dedicated accelerators [56] this problem becomes less and less vivid.

1.2 MONTE CARLO

This section is dedicated to a short overview on MC methods. We begin with the most basic settings and then proceed to the MC methods for SDEs, as they correspond to the option pricing problem.

1.2.1 *Basics of Monte Carlo Method*

The MC methods are targeting a problem of the estimation of an expectation

$$\mathbb{E}\left(X\right)$$

of some random variable X. The basis of the MC approximation is the idea of simulating $(X_i)_{i=1,2,\ldots,N}$, N independent identically distributed (i.i.d.) realisations of the random variable X, and approximating $\mathbb{E}\left(X\right)$ by their average

$$\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i \,. \tag{5}$$

This approximation is justified by Kolmogorov's Strong Law of Large Numbers.

Theorem 2 (Kolmogorov's Strong Law of Large Numbers)**.** *Let* (X_n), $n \in \mathbb{N}$, *be a sequence of integrable real-valued i.i.d. random variables. Then*

$$\bar{X}_N := \frac{1}{N}\sum_{i=1}^{N} X_i \xrightarrow{a.s.} \mathbb{E}(X_1) \text{ as } N \to \infty .$$

It can be shown that the MC estimator (5) is unbiased, i.e. $\mathbb{E}(\bar{X}) = \mathbb{E}(X_1)$. However this usually will not be the case in an SDE setting in general and in forthcoming parts of this thesis in particular. Unbiasedness of an estimator ensures that it is correct in the mean, though it does not tell how good is the estimate in the absolute value. Therefore we measure a *statistical error* by a variance of an estimate. This choice of error can be justified by the Central Limit Theorem.

Theorem 3 (Central Limit Theorem)**.** *Let* (X_n) $n \in \mathbb{N}$ *be a sequence of integrable real-valued i.i.d. random varianles with finite variance* $\sigma^2 := \mathbb{V}\mathrm{ar}(X_1)$ *and expectation* $m := \mathbb{E}(X_1)$. *Then the normalised and centred sum of* X_i *convergences in distribution to a standard normal distribution, i.e.*

$$\frac{\sum_{i=1}^{N} X_i - mN}{\sigma\sqrt{N}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \text{ as } N \to \infty .$$

From the Central Limit Theorem it follows that for large values of N the MC estimator (5) is approximately $\mathcal{N}\left(\mathbb{E}(X), \mathbb{V}\mathrm{ar}\left(\bar{X}\right)\right)$ distributed. Then using the asymptotic distribution of the estimator asymptotic confidence intervals for $\bar{X}$ can be constructed [45]. Moreover their length is proportional to $\sqrt{\mathbb{V}\mathrm{ar}\left(\bar{X}\right)}$, i.e. square root of the statistical error.

1.2.2 *Monte Carlo In SDE Settings*

Very often in financial mathematics in order to price some derivative an expected value of a functional of a solution of an SDE has to be computed. For instance the price of Asian call option with maturity T and strike price K is of the form of

$$C = \mathbb{E}\left(e^{-rT}\left(\frac{1}{T}\int_0^T S(t)\,dt - K\right)^+\right) ,$$

where S is modelled as a solution of some SDE, e.g. the BS SDE under EMM

$$dS(t) = rS(t)\,dt + \sigma S(t)\,dW(t) , \quad t \in [0,T] .$$

The MC approach from Section 1.2.1 suggests to draw i.i.d. realisations of

$$e^{-rT}\left(\frac{1}{T}\int_0^T S(t)\,dt - K\right)^+ \tag{6}$$

and average them.

However we cannot directly simulate (6), as there exist no closed form representation for the integral of S over time. As a consequence it has to be approximated, e.g. using trapezoidal rule [18], what introduces a *bias* error.

Moreover in the case of the Heston model there exists no closed form solution of SDE (3). Therefore we cannot directly simulate the stock price process S even on a discrete time grid and it has to be approximated using some *discretisation scheme*, what leads to a biased estimate.

To summarise we can give at least two cases leading to a biased estimate:

- an option under consideration is path dependent, e.g. Asian options, barrier options, etc. ;
- a model SDE has no closed form solution, e.g. the Heston model.

Existence of a bias error and the necessity to approximate a solution of the SDE differentiate MC estimation in an SDE settings from the most basic case of Section 1.2.1. This particularities we will discuss in the sequel of this section.

PLAIN MONTE CARLO ESTIMATE

Discretisation schemes necessary to approximate a solution of an SDE can be regarded as a straightforward generalisation of discretisation schemes for ordinary differential equations (ODEs) [18]. Some examples of discretisation schemes will be introduced later in Section 1.2.2.1 and for the time being let us denote a numerical approximation of a process S as $\hat{S}$.

Then *plain MC* estimate of $\mathbb{E}\left(g\left(S\right)\right)$ is defined as

$$\hat{E} = \frac{1}{N}\sum_{j=1}^{N} g\left(\hat{S}_j\right), \tag{7}$$

where $\hat{S}_1, \hat{S}_2, \ldots, \hat{S}_N$ are i.i.d. realisations of the discretisation scheme $\hat{S}$.

1.2.2.1 *Discretisation Schemes*

Let us briefly introduce the two most popular discretisation schemes for SDEs. For more details we refer to the book of Kloeden and Platen [39] . Note in this thesis we consider only equidistant schemes.

EULER SCHEME

The *Euler scheme* for SDEs can be regarded as a straightforward generalisation of the Euler scheme for ODEs [18]. It is motivated by the Itô-Taylor expansion [37] in the similar way as the Euler scheme for ODEs by the Taylor expansion. In general, this approach even leads to a whole class of methods. For simplicity, we will define this scheme for a one-dimensional SDE only. A multi-dimensional version of the Euler scheme can be found in e.g. [45].

Definition 1 (Euler Discretisation Scheme)**.** *Let* n *be the number of equidistant discretisation steps and* $\Delta t = T/n$*, then for a one-dimension SDE*

$$dS(t) = a(t, S(t))dt + \sigma(t, S(t))dW(t)\ , \quad S(0) = s_0 \in \mathbb{R},$$

we approximate a realisation of S *with its discretised version* $\hat{S}^n$ *according to the iterative* Euler scheme *defined as*

$$\begin{aligned} &\hat{S}^n(t_0) = s_0\ , \\ &\hat{S}^n(t_{i+1}) = \hat{S}^n(t_i) + a\left(t_i, \hat{S}^n(t_i)\right)\Delta t + \sigma\left(t_i, \hat{S}^n(t_i)\right)\Delta W(t_i)\ , \end{aligned}$$

for $i = 0, 1, 2, ..., n-1$*, where* $t_i = i\,\Delta t$ *and* $\left(\Delta W(t_j)\right)_{j=0,1,\dots,n-1}$ *are i.i.d. normal random variables with zero mean and variance* Δt*, i.e.* $\Delta W(t_j) \sim \sqrt{\Delta t}\,\mathcal{N}(0,1)$.

Remark 5. *Defined Euler scheme is a discrete approximation of path of a solution of the corresponding SDE. To obtain a continuous approximation one typically takes linear interpolation between discrete points [39].*

However, sometimes, e.g. in the case of a barrier options pricing [30], it is beneficial to assume that in between the approximation follows a Brownian bridge (for definition and properties see [37]) construction.

Now we will present some results on a convergence of the Euler scheme.

Definition 2 (Strong Convergence of a discretisation scheme)**.** *We say that a discretisation scheme for the SDE* (4) converges strongly *on* $[0, T]$ *to the solution* S *of the SDE, if for the final time* T *we have*

$$\lim_{n\to\infty} \mathbb{E}\left|S(T) - \hat{S}^n(T)\right| = 0\ .$$

A strongly convergent scheme is said to have convergence rate α *if for asome constant* C *and* $n_0 \in \mathbb{N}$ *we have*

$$\forall n > n_0 \qquad \mathbb{E}\left|S(T) - \hat{S}^n(T)\right| \leqslant C\frac{1}{n^\alpha}\ .$$

We begin with a well-known result on a strong convergence of the Euler scheme [39].

Theorem 4 (Strong convergence of the Euler scheme)**.** *Under the assumption of Theorem 1 (i.e. Lipschitz coefficients and linear growth conditions on the coefficients of the SDE) and additionally*

$$\|a(t,x) - a(l,x)\| + \|\sigma(t,x) - \sigma(l,x)\| \leqslant K(1 + \|x\|)|t-l|^{\frac{1}{2}}$$

for some suitable constant K *, the Euler scheme converges strongly with a convergence rate of* $1/2$*, i.e. for some constant* C *and* $n_0 \in \mathbb{N}$ *we have*

$$\mathbb{E}\left|S(T) - \hat{S}^n(T)\right| \leqslant Cn^{-1/2} \quad \forall n > n_0 \, .$$

Even more, if we define the "continuous" Euler approximation $\hat{S}^n$ *as the process equal to the Euler scheme process on the discretisation grid and linearly interpolated in between, then we have*

$$\mathbb{E}\left(\sup_{0 \leqslant t \leqslant T}\left|S(t) - \hat{S}^n(t)\right|\right) < Cn^{-1/2} \, .$$

The later uniform convergence result is important for applications in finance, that involve computation of a price of a path-dependent option as for example barrier options.

As in the option pricing we are interested in estimating expectations of a function of an SDE solution, we also study a weak convergence property of the scheme.

Definition 3 (Weak Convergence of a discretisation scheme)**.** *We say that a numerical scheme for the SDE* (4) converges weakly *on* $[0, T]$ *to the solution* S *of the SDE with respect to (w.r.t.) the class of functions* $\mathcal{H}$*, if we have*

$$\forall g \in \mathcal{H} \, , \qquad \lim_{n \to \infty}\left|\mathbb{E}\left(g(\hat{S}^n(T))\right) - \mathbb{E}\left(g(S(T))\right)\right| = 0 \, .$$

A weakly convergent scheme w.r.t. the class $\mathcal{H}$ *is said to have* convergence rate α *if for some constant* C *and* $n_0 \in \mathbb{N}$ *we have*

$$\forall g \in \mathcal{H} \, , \, \forall n \geqslant n_0 \, , \qquad \left|\mathbb{E}\left(g(\hat{S}^n(T))\right) - \mathbb{E}\left(g(S(T))\right)\right| \leqslant C\frac{1}{n^\alpha} \, .$$

This type of convergence is much weaker, but allows us to consider a much wider class of numerical schemes. A weak convergence result [39] for the Euler scheme and the class of polynomial functions can be read as:

Theorem 5 (Weak Convergence of the Euler Scheme)**.** *If we have Lipschitz continuous and polynomially bounded autonomous coefficients functions* $a(x)$ *and* $\sigma(x)$*, which are in* C_P^4 *(i.e. they are four times differentiable and together with their derivatives are at most polynomially growing) then the Euler scheme is weakly convergent with a convergence rate of one with respect to the class* $\mathcal{H}$ *of all polynomials, i.e. we have*

$$\left|\mathbb{E}\left(g(\hat{S}^n(T))\right) - \mathbb{E}\left(g(S(T))\right)\right| = O\left(n^{-1}\right) \qquad \textit{as} \quad n \to \infty$$

for any polynomial g*.*

MILSTEIN SCHEME

The Milstein scheme is the next order scheme obtained via the Itô-Taylor expansion [39]. Therefore it assumes differentiability of a volatility coefficient σ. If this is the case than the Milstein scheme can be defined as:

Definition 4 (Milstein Discretisation Scheme)**.** *Let* n *be the number of equidistant discretisation steps and* $\Delta t = T/n$*, then for an one-dimension SDE*

$$dS(t) = a(t, S(t))dt + \sigma(t, S(t))dW(t)\ , \quad S(0) = s_0 \in \mathbb{R},$$

we approximate a realisation of S *with its discretised version* $\hat{S}^n$ *according to the iterative* Milstein scheme *defined as*

$$\begin{aligned} \hat{S}^n(t_0) &= s_0\ ; \\ \hat{S}^n(t_{i+1}) &= \hat{S}^n(t_i) + a\left(t_i, \hat{S}^n(t_i)\right)\Delta t + \sigma\left(t_i, \hat{S}^n(t_i)\right)\Delta W(t_i) \\ &\quad + \frac{1}{2}\sigma\left(t_i, \hat{S}^n(t_i)\right)\sigma'\left(t_i, \hat{S}^n(t_i)\right)\left((\Delta W(t_i))^2 - \Delta t\right)\ ; \end{aligned}$$

for $i = 0, 1, 2, ..., n-1$*, where* $t_i = i\,\Delta t$ *and* $\left(\Delta W(t_j)\right)_{j=0,1,\ldots,n-1}$ *are i.i.d. normal random variables with zero mean and variance* Δt*, i.e.* $\Delta W(t_j) \sim \sqrt{\Delta t}\,\mathcal{N}(0, 1)$*.*

We state a theorem on a strong convergence behaviour of the Milstein scheme given in [45].

Theorem 6 (Strong convergence of the Milstein scheme)**.** *In addition to the assumptions of Theorem 1 and Theorem 4, let* $\sigma(t, s)\frac{\partial\sigma(t,s)}{\partial x}$ *satisfy the conditions on the coefficients of Theorem 4. If further we have that* $a \in C^{1,1}$ *and* $\sigma \in C^{1,2}$ *then the Milstein scheme converges strongly with a convergence rate one, i.e. for some constant* C *and* $n_0 \in \mathbb{N}$ *we have*

$$\mathbb{E}\left|S(T) - \hat{S}^n(T)\right| \leqslant Cn^{-1} \quad \forall n > n_0\ .$$

Thus, under extra assumptions the Milstein exhibits a significantly better rate of a strong convergence compare to the Euler scheme. However it has not got any improvement w.r.t. a weak convergence order, i.e. the Milstein scheme has a weak convergence order of one [39]. On the first glance this suggests to use the Euler scheme, as it is simpler than the Milstein. Though, under appropriate assumptions the Milstein scheme outperforms the Euler, when pathwise approximations are needed. Moreover due to the bigger rate of a strong convergence it has a great advantage in case of the MLMC estimate (see Section 1.3).

Remark 6. *A generalisation of the Milstein scheme to a multi-dimensional case is not straightforward. It involves computations of double stochastic integrals, what can be done analytically only in some cases [45]. Hence this* *integrals should be approximated numerically, what makes the scheme inefficient.*

Giles and Szpruch in [27] propose to use so-called Antithetic MLMC technique to avoid computation of the aforementioned integrals.

SOME REMARKS

Very often instead of approximation of a stock process S, one uses the Itô formula to obtain SDE for $\ln(S)$ and approximate solution of the later SDE by some discretisation scheme. This transformation helps to avoid problem with negative values of $\hat{S}$ that can occur during approximation of the original SDE. Also it often leads to a smaller absolute values of a bias error [55].

If we evaluate the payoff of a barrier option based on a discrete approximation $\hat{S}^n$, we systematically overestimate or underestimate the option price. This happens due to a discrete monitoring, i.e. we check if the price has crossed the barrier only in discrete times $t_i = i\Delta t$ for $i = 0, 1, \ldots, n$. To overcome this problem we suggest to use the Brownian bridge technique [30] or the barrier shifting approach [13]. Note, both approaches do not eliminate a bias error induced by a discrete scheme approximation.

The volatility process V in the Heston model (3) is non-negative, but a straightforward application of the Euler scheme to this process leads to negative values of $\hat{V}$. Various techniques were proposed to deal with this problem. We refer to an overview article [53] and some recent advances [3].

The assumptions of the theorems on strong and weak convergences of the Euler and Milstein schemes are quite strong. In some case they can be either relaxed or substituted by some others, especially in the case of a weak convergence. We direct the reader to [39], [36], series of papers by Kurtz and Protter ([46], [48], [47]) and reference therein.

1.2.2.2 *Errors and Computational Cost*

To measure performance of MC estimate we define its computational cost (CC) and error. A comparison of different MC method will be done by comparing their CC expressed in terms of the error.

MONTE CARLO ERRORS IN SDE SETTINGS

Application of discretised or approximated solution of SDE makes MC estimate biased. Therefore in the SDE settings an error of an estimate is measured by a *mean square error (MSE)*

$$\mathbb{MSE}\left[\hat{E}\right] := \mathbb{E}\left(\hat{E} - \mathbb{E}\left(P\right)\right)^2 ,$$

where $\mathbb{E}(P)$ is value of interest or equivalently by a *root mean square error (RMSE)* defined as $\mathrm{r}\mathbb{MSE} := \sqrt{\mathbb{MSE}}$.

This error measure can be decomposed into two components:

- statistical error;
- bias error induced by inexact approximation.

For instance in the case of the plain MC estimate (7) this can be done as follows:

$$\begin{aligned}
\mathbb{MSE}\left[\hat{E}\right] &= \mathbb{E}\left(\hat{E} - \mathbb{E}\left(\hat{E}\right) + \mathbb{E}\left(\hat{E}\right) - \mathbb{E}\left(P\right)\right)^2 \\
&= \mathbb{E}\left(\hat{E} - \mathbb{E}\left(\hat{E}\right)\right)^2 + \left(\mathbb{E}\left(\hat{E}\right) - \mathbb{E}\left(P\right)\right)^2 \\
&= \underbrace{\mathbb{V}\mathrm{ar}\left(\hat{E}\right)}_{\text{stat. error}} + \Big(\underbrace{\mathbb{E}\left(\hat{E}\right) - \mathbb{E}\left(P\right)}_{\text{bias}}\Big)^2 \\
&= \frac{1}{N}\mathbb{V}\mathrm{ar}\left(g\left(\hat{S}\right)\right) + \left(\mathbb{E}\left(g\left(\hat{S}^n\right)\right) - \mathbb{E}\left(g\left(S\right)\right)\right)^2 .
\end{aligned}$$

Remark 7. *Indeed this MSE decomposition can be obtained for all estimates, that are linear combinations of some plain MC estimates.*

The MSE decomposition, in particular, means that both errors should be controlled simultaneously and consistent. The number of simulated paths N should be increased to decrease statistical error, while increasing number of discretisation steps n will lead to a smaller bias.

COMPUTATIONAL COST OF ESTIMATE

First, we define the CC of a discretisation scheme to be (asymptotically) proportional to the number of discretisation steps, i.e.

$$\mathrm{C.C.}\left[\hat{S}^n\right] \asymp n .$$

As typically $n \gg 1$, CC of evaluation of a functional g can be neglected, what immediately implies that

$$\mathrm{C.C.}\left[g\left(\hat{S}^n\right)\right] \asymp n .$$

Then in case of the plain MC estimate (7) the CC should be multiplied by the number of simulated paths N, i.e.

$$\mathrm{C.C.}\left[\hat{E}\right] \asymp N\,\mathrm{C.C.}\left[g\left(\hat{S}^n\right)\right] \asymp N\,n . \tag{8}$$

Remark 8. *There exist another definitions of a CC, e.g. proportional to a number of simulated pseudo-random variables. However they are all build in a way to reflect computer run time.*

Equation 8 implies that simultaneous control of bias and statistical errors by enlargement of n and N correspondingly leads to increasing CC of an estimate. Therefore this control should be done in an optimal way, what have been first studied by Duffie and Glynn in [22]. In particular, the optimal relation between n and N depends on an order of weak convergence of the applied scheme $\hat{S}^n$.

1.3 MULTILEVEL MONTE CARLO

The idea of the MLMC estimate was first developed by Heinrich in [33]. However it has been applied not in the SDE settings. Then half a decade later the MLMC method was redeveloped and popularised by Giles after publishing his seminal paper [28].

The MLMC method extends idea of the Statistical Romberg (SR) approach by Kebaier [38]. Instead of using one scheme with n discretisation points the MLMC method suggests to use a sequence of schemes $\hat{S}^l$ with different numbers of discretisation points.

To be more precise without loss of generality we can assume that $n = M^L$ discretisation points for a fixed M and some integer L are sufficient to obtain the desired bias accuracy of the plain MC estimate. Let us denote

$$n_i = M^i \quad \text{for} \quad i = 0, 1, \ldots, L\,,$$

and consider

$$\hat{S}^{n_0}, \hat{S}^{n_1}, \ldots, \hat{S}^{n_L}$$

as a sequence of approximations of the SDE solution based on different numbers of points. Then, $\mathbb{E}\left(g\left(\hat{S}^{n_L}\right)\right)$, that is in the focus of the plain MC estimate, can be represented as a telescopic sum

$$\mathbb{E}\left(g\left(\hat{S}^{n_L}\right)\right) = \mathbb{E}\underbrace{\left(g\left(\hat{S}^{n_0}\right)\right)}_{=:D_0} + \sum_{i=1}^{L} \mathbb{E}\underbrace{\left(g\left(\hat{S}^{n_i}\right) - g\left(\hat{S}^{n_{i-1}}\right)\right)}_{=:D_i}\,, \tag{9}$$

where D_i are called *levels*. The MLMC algorithm approximates each of these levels with independent plain MC estimates. This leads to the following form of the *Multilevel Monte Carlo estimate*

$$\hat{E} = \frac{1}{N_0}\sum_{j=1}^{N_0} g\left(\hat{S}_j^{n_0}\right) + \sum_{i=1}^{L}\frac{1}{N_i}\sum_{j=1}^{N_i}\left(g\left(\hat{S}_j^{n_i}\right) - g\left(\hat{S}_j^{n_{i-1}}\right)\right)\,. \tag{10}$$

It is important that within level approximations $\hat{S}_j^{n_{i-1}}$ and $\hat{S}_j^{n_i}$ are based on the same simulated Brownian paths.

The CC of simulating D_i replication is increasing. On the other hand the variance of D_i is typically decreasing for increasing i. In the classical settings this is due to the strong convergence of the discretisation scheme. To be more precise under conditions of Theorem 10.2.2 form [39] we have that

$$\mathbb{E}\left(\hat{S}^{n}(T) - S(T)\right)^{2} \asymp n^{-1} .$$

Let us consider a Lipschitz payoff g, what in particular implies that

$$\begin{aligned}\mathbb{V}\mathrm{ar}\left(g\left(\hat{S}^{n}\right) - g(S)\right) &\leqslant \mathbb{E}\left(g\left(\hat{S}^{n}\right) - g(S)\right)^{2} \\ &\leqslant c^{2}\, \mathbb{E}\left(\hat{S}^{n}(T) - S(T)\right)^{2} \asymp n^{-1} ,\end{aligned} \tag{11}$$

where c is the Lipschitz constant. Further elementary transformations yield

$$\begin{aligned}&\mathbb{V}\mathrm{ar}\left(g\left(\hat{S}^{n_i}\right) - g\left(\hat{S}^{n_{i-1}}\right)\right) \\ &\quad= \mathbb{V}\mathrm{ar}\left(g\left(\hat{S}^{n_i}\right) - g(S) + g(S) - g\left(\hat{S}^{n_{i-1}}\right)\right) \\ &\quad\leqslant \left(\sqrt{\mathbb{V}\mathrm{ar}\left(g\left(\hat{S}^{n_i}\right) - g(S)\right)} + \sqrt{\mathbb{V}\mathrm{ar}\left(g\left(\hat{S}^{n_{i-1}}\right) - g(S)\right)}\right)^{2} .\end{aligned} \tag{12}$$

Hence the variance of the levels decays w.r.t. level propagation, i.e. combining Equation 11 and Equation 12

$$\mathbb{V}\mathrm{ar}\left(g\left(\hat{S}^{n_i}\right) - g\left(\hat{S}^{n_{i-1}}\right)\right) \asymp n_{i-1}^{-1} , \qquad \text{if} \quad n_i > n_{i-1} . \tag{13}$$

Note, to obtain this behaviour it is important to use the same Brownian paths a within level.

Remark 9. *The above considerations explain the importance of a strong convergence, which seems to be useless in the plain MC setting. Moreover an order of a strong convergence defines CC of the MLMC estimate (see Theorem 7).*

This levels variance decay behaviour is exploited by the MLMC idea: Smaller variance makes less repetitions N_i necessary to reach the same level of the statistical error. Finally, the MLMC algorithm aims at reducing the overall CC by optimally distributing the workload over all levels (see [28] or Chapter 2 of this thesis).

The main advantage of the MLMC is summarised in the following so-called *Complexity Theorem* [28], that shows the improvement on CC compared to the plain MC. Below we present a simplified and adopted to our notations version of the Complexity Theorem.

Theorem 7 (Complexity Theorem)**.** *Let* $g(S)$ *denote a functional of the solution of SDE* (4) *for a given Brownian path* $W(t)$*, and let* $g\left(\hat{S}^{n_i}\right)$ *denote the corresponding approximation using a numerical discretisation with* $n_i = M^i$ *steps.*

If there exist positive constants $\alpha \geqslant 0.5$, β, c_1, c_2, c_3 *such that*

i) $\mathbb{E}\left(g\left(\hat{S}^{n}\right) - g(S)\right) \leqslant c_1\, n^{-\alpha}$, *ii)* $\mathbb{V}\mathrm{ar}(D_i) \leqslant c_2\, n_i^{-\beta}$,

iii) $\mathrm{C.C.}\,[D_i] \leqslant c_3\, n_i$,

then there exists a positive constant c_4 *such that for any* $\varepsilon < e^{-1}$ *there are values* L *and* N_i *for which the Multilevel Monte Carlo estimator* (10) *has a mean square error with bound*

$$\mathbb{MSE}\left[\hat{E}\right] = \mathbb{E}\big(\hat{E} - \mathbb{E}(P)\big)^2 \leqslant \varepsilon^2$$

with a computational cost C. C. *with bound*

$$\mathrm{C.C.}\left[\hat{E}\right] \leqslant \begin{cases} c_4\ \varepsilon^{-2} & \textit{if} \quad \beta > 1\,, \\ c_4\ \varepsilon^{-2}\ (\ln(\varepsilon))^2 & \textit{if} \quad \beta = 1\,, \\ c_4\ \varepsilon^{-2-(1-\beta)/\alpha} & \textit{if} \quad 0 < \beta < 1\,. \end{cases}$$

Remark 10. *Note, the Complexity Theorem is given in non-constructive way. To give an example of an SDE, functional and scheme we propose for instance to think about the BS SDE* (1)*, European call option and the Euler scheme. Then* assumption i) *is proven by Talay and Tubaro in [63] with* $\alpha = 1$. Assumption ii) *with* $\beta = 1$ *is an implication of the considerations leading to Equation 13 and* assumption iii) *is an implication of the Euler scheme construction.*

For a constructive version in spirit of the Central Limit Theorem see [1].

We would like to conclude by summarising the MLMC idea into two points.

- The telescopic sum property (9), that ensures correctness of result in the mean
- and the level variance decay (13)

are the main building blocks of the MLMC estimate.

1.4 THESIS STRUCTURE

Here we briefly explain what the reader can find in the imminent parts of this thesis and introduce notations we will be using in the future.

1.4.1 *Structure*

The rest of this thesis consists of the two chapters.

Chapter 2 is addressed to make a deep investigation of the MLMC method. In particular we take an optimisation view at the estimate. Rather than fixing the number of discretisation points n_i to be a geometric sequence, we are trying to find an optimal set up for n_i such that for a fixed error the estimate can be computed within a minimal time.

In Chapter 3 we propose to enhance the MLMC estimate with the weak extrapolation technique. This technique helps to improve order of a weak convergence of a scheme and as a result reduce CC of an estimate. In particular we study high order weak extrapolation approach, which is know not be inefficient in the standard settings. However, a combination of the MLMC and the weak extrapolation yields an improvement of the MLMC.

1.4.2 *Notation Convention*

In this thesis we will use the following notation convention. Very often we will be talking about an (numerical) *approximation scheme* $\hat{P}^n$, that approximates P . By this notation we usually mean

$$\hat{P}^n = g\left(\hat{S}^n\right) \qquad \text{and} \qquad P = g\left(S\right) \ ,$$

where $\hat{S}^n$ is a discretisation scheme for some process S, e.g. the Euler or Milstein scheme (see Section 1.2.2). The functional g will be typically equal to the discounted payoff of an option, e.g.

$$g^b\left(S\right) = e^{-T}\left(S\left(T\right) - K\right)^+ \mathbb{I}_{\{S(t) > H, \forall t \in [0,T]\}}$$

for the barrier down-and-out call with maturity T, strike K and barrier H. Also it might be some modification of the payoff functional, e.g. the shifting barrier technique suggests to use lifted up barrier $\tilde{H}$ in approximation (for details see [13])

$$\tilde{g}^b\left(S\right) = e^{-T}\left(S\left(T\right) - K\right)^+ \mathbb{I}_{\{S(t) > \tilde{H}, \forall t \in [0,T]\}}$$

and then

$$\hat{P}^n = \tilde{g}^b\left(\hat{S}^n\right) \qquad \text{and} \qquad P = g^b\left(S\right) \ .$$

In the rest of the thesis we will not be specifying an exact discretisation scheme. Instead we will require an approximation $\hat{P}^m$ to possess some properties, as for example $C.C.\left[\hat{P}^m\right]$ of the scheme expressed in terms of m. This approach is motivated by the formulation of the Complexity Theorem 7.

2 MLMC OPTIMISATION VIEW

In the classic MLMC settings one considers a sequence of approximations with n_i discretisation steps, where n_i follows a geometric progression. However there exists no justification for this choice. Therefore we are doing a step back and consider a multilevel construction (17) without fixing numbers of discretisation points to be a geometric sequence. Instead we are trying to find an optimal set up for n_i such that for a fixed error the estimate has a smallest possible CC.

As a result we justify the geometric progression choice. Moreover an optimal common ratio M will raise as a by-product of our considerations. Additionally, we will confirm a multiple control variate interpretation [45] of the MLMC estimate.

2.1 MULTILEVEL CONSTRUCTION

Let $\hat{P}^m$ be some approximation scheme of P based on m equidistant discretisation points (see Section 1.4.2 or Section 2.2.1 for examples). Moreover let this scheme possess the following properties:

- boundedness of variance

$$\mathbb{V}\mathrm{ar}\left(\hat{P}^n\right) < \infty \ \forall n \in \mathbb{N}; \tag{14}$$

- "strong convergence"

$$\mathbb{V}\mathrm{ar}\left(P - \hat{P}^n\right) \asymp n^{-\beta} \tag{15}$$

  for some $\beta > 0$;

- CC is asymptotically proportional to γ degree polynomial in terms of the number of discretisation points

$$\mathrm{C.C.}\left[\hat{P}_n\right] \asymp n^{\gamma} \tag{16}$$

  for some $\gamma > 0$.

The above assumptions are quite standard and, in particular, are similar to the one from the Complexity Theorem 7 for the classic MLMC. In the meanwhile there is no assumption regarding an order of weak convergence of the scheme. Also assumption (14) can be derived from assumption (15) for P with a bounded variance, but we would like to keep it explicit.

For the later we are interested in estimating

$$\mathbb{E}\left(\hat{P}^{n}\right)$$

for a fixed and given in advance n, rather then estimating $\mathbb{E}(P)$ itself, and an unbiased estimator will be constructed an (asymptotically) optimal way.

Remark 11. *Such an estimator might be needed if, for instance,*

- *the number of discretisation points n corresponding to an "appropriate" bias error of the scheme $\hat{P}^{n}$ is known in advance;*
- *$\mathbb{E}\left(\hat{P}^{n}\right)$ yields an "acceptable" upper or lower bound of a product price for known in advance n;*
- *$\mathbb{E}\left(\hat{P}^{n}\right)$ is an exact price of the product itself (e.g. discrete Asian options);*
- *n is chosen in advance due to computation latency requirements or any other computation system constrain.*

MULTILEVEL CONSTRUCTION ESTIMATE

For some sequence of schemes $\hat{P}^{n_i}$ with increasing numbers of discretisation points n_i

$$n_0 < n_1 < \ldots < n_{L-1} < n_L = n$$

we define a *multilevel constructed (MLC) estimate* $\hat{E}$ as

$$\hat{E} = \frac{1}{N_0}\sum_{j=1}^{N_0}\hat{P}_j^{n_0} + \sum_{i=1}^{L}\frac{1}{N_i}\sum_{j=1}^{N_i}\left(\hat{P}_j^{n_i} - \hat{P}_j^{n_{i-1}}\right). \tag{17}$$

The summands on the right hand side (RHS) of Equation 17

$$\hat{Y}_i = \begin{cases} \frac{1}{N_0}\sum_{j=1}^{N_0}\hat{P}_j^{n_0} & \text{for} \quad i = 0; \\ \frac{1}{N_i}\sum_{j=1}^{N_i}\left(\hat{P}_j^{n_i} - \hat{P}_j^{n_{i-1}}\right) & \text{for} \quad i = 1, 2, \ldots, L\,, \end{cases}$$

are independent and called *levels*. The telescopic sum property

$$\mathbb{E}\left(\hat{E}\right) = \mathbb{E}\left(\hat{P}^{n_0}\right) + \sum_{i=1}^{L}\left(\mathbb{E}\left(\hat{P}^{n_i}\right) - \mathbb{E}\left(\hat{P}^{n_{i-1}}\right)\right) = \mathbb{E}\left(\hat{P}^{n}\right)$$

ensures that the MLC (17) is an unbiased estimate of $\mathbb{E}\left(\hat{P}^{n}\right)$.

On the one hand the form of this estimate is quite general. On the other it is similar to the classic MLMC or can be seen as a multi-step extension of the SR estimate by Kebaier [38]. Furthermore as for the MLMC and the SR it is crucial that schemes $\hat{P}^{n_i}$ and $\hat{P}^{n_{i-1}}$ within a level are based on the same Brownian motion.

OPTIMALITY

By optimality we mean that the MLC estimate (17) has got a (asymptotically) minimal CC for a fixed statistical error. This will be achieved by an optimal choice of

- numbers of the simulated paths on each level

$$N_i \in \mathbb{N} \text{ for } i = 0, 1, \ldots, L \,,$$

- numbers of the discretisation points for each scheme

$$n_i \in \mathbb{N} \text{ for } i = 0, 1, \ldots, L-1 \,,$$

- number of the levels

$$L \,.$$

More precisely we will solve the following minimisation problem

$$\begin{aligned} \mathrm{C.C.}\left[\hat{E}\right] &\to \min \\ \text{s.t. } \mathbb{V}\mathrm{ar}\left(\hat{E}\right) &= \varepsilon^2 \,. \end{aligned} \tag{18}$$

Remark 12. *Alternatively we can consider the MLC estimate that for a fixed CC has got a minimal statistical error, i.e.*

$$\begin{aligned} \mathbb{V}\mathrm{ar}\left(\hat{E}\right) &\to \min \\ \textit{s.t. } \mathrm{C.C.}\left[\hat{E}\right] &= \mathrm{cost} \,. \end{aligned}$$

Both optimisation problems are closely related. The optimal values of numbers of the discretisation points n_i and of number of the levels L coincide for them. Moreover the optimal value of the alternative problem will be equal to ε^2 if cost is equal to the optimal value of Problem 18 and the optimal numbers of the simulated paths N_i coincide.

2.1.1 *Optimisation*

This section is dedicated to the solution of Problem (18). The problem will be solved through the sequence of three minimisation subproblems:

1. find the optimal numbers of simulated paths for each level

   $$N_i \in \mathbb{N} \text{ for } i = 0, 1, \ldots, L \,,$$

   that minimise $\mathrm{C.C.}\left[\hat{E}\right]$ under a constraint of $\mathbb{V}\mathrm{ar}\left(\hat{E}\right) = \varepsilon^2$;

2. find the optimal numbers of discretisation steps for each scheme

   $$n_i \in \mathbb{N} \text{ for } i = 0, 1, \ldots, L-1 \,,$$

   that minimise optimal value of $\mathrm{C.C.}\left[\hat{E}\right]$ from the above problem;

3. find an optimal number of levels

   $$L\,,$$

   that minimises the optimal value of $\mathrm{C.C.}\left[\hat{E}\right]$ from the above problem.

If n_i, N_i and L are treated as real numbers these sub-problems can be solved in an analytical form.

Solution of the firs sub-problem can be found in [28], but it is done rather in spirit of Remark 12. The second sub-problem (or similar) has already been studied by Marxen in [55] for a fixed in advance L. Its solution leads to the so-called multi-step SR method as an extension of the original SR method by Kebaier.

We solve minimisation Problem (18) as if n_i, N_i and L were real numbers. To emphasise this we will use overlines with n_i, N_i and L if they are assumed to be real, i.e. we write $\bar{n}_i$, $\bar{N}_i$ and $\bar{L}$ where applicable. There is an essential gap between continuous optimisation and its integer counterpart that prevents optimality arguments to be brought from one problem to the other and vice versa. Hence the following optimisation considerations are not strictly rigorous. Nevertheless the cautious reader should not be concerned, as the final result on CC of the MLC estimate will be given in a mathematically accurate way in Section 2.1.3.

In the following we first prepare the set up for our optimisation problem. Then we present solutions of the sub-problems and discuss their implications. Detailed steps of the solutions are technical and, thus, postponed to Appendix A.

SET UP

Let us quantify CC and the variance of the MLC estimate (17). From the form of the MLC estimator (17) it directly follows that

$$\begin{aligned}\mathbb{V}\mathrm{ar}\left(\hat{E}\right) &= \frac{1}{N_0}\mathbb{V}\mathrm{ar}\left(\hat{P}^{n_0}\right) + \sum_{i=1}^{L}\frac{1}{N_i}\mathbb{V}\mathrm{ar}\left(\hat{P}^{n_i} - \hat{P}^{n_{i-1}}\right) \\ &= \sum_{i=0}^{L} N_i^{-1} V_i\,, \end{aligned} \tag{19}$$

where

$$V_i = \begin{cases} \mathbb{V}\mathrm{ar}\left(\hat{P}^{n_0}\right) & \text{for} \quad i = 0\,, \\ \mathbb{V}\mathrm{ar}\left(\hat{P}^{n_i} - \hat{P}^{n_{i-1}}\right) & \text{for} \quad i = 1, 2, \ldots, L\,. \end{cases}$$

As a number of levels L is usually small (in applications it is typically less than 10) CC of summation across the levels is asymptotically negligible, thus we obtain that

$$\begin{aligned} C.C.\left[\hat{E}\right] &= N_0\, C.C.\left[\hat{P}^{n_0}\right] + \sum_{i=1}^{L} N_i\, C.C.\left[\hat{P}^{n_i} - \hat{P}^{n_{i-1}}\right] \\ &= \sum_{i=0}^{L} N_i C_i\,, \end{aligned} \tag{20}$$

where

$$C_i = \begin{cases} C.C.\left[\hat{P}^{n_0}\right] & \text{for} \quad i = 0\,, \\ C.C.\left[\hat{P}^{n_i} - \hat{P}^{n_{i-1}}\right] & \text{for} \quad i = 1, 2, \ldots, L\,. \end{cases}$$

Using assumptions (14), (15) and (16) on the scheme $\hat{P}^m$ we express C_i and V_i in terms of the numbers of discretisation points n_i. First of all it is easy to obtain the following upper and lower bounds for V_i

$$\begin{aligned} V_i &= \mathbb{V}\text{ar}\left(\hat{P}^{n_i} - \hat{P}^{n_{i-1}}\right) = \mathbb{V}\text{ar}\left(\hat{P}^{n_i} - P \;+\; P - \hat{P}^{n_{i-1}}\right) \\ &= \mathbb{V}\text{ar}\left(\hat{P}^{n_i} - P\right) + 2\text{Cov}\left(\hat{P}^{n_i} - P, P - \hat{P}^{n_{i-1}}\right) + \mathbb{V}\text{ar}\left(P - \hat{P}^{n_{i-1}}\right) \\ &\Downarrow \\ V_i &\leqslant \left(\sqrt{\mathbb{V}\text{ar}\left(P - \hat{P}^{n_i}\right)} + \sqrt{\mathbb{V}\text{ar}\left(P - \hat{P}^{n_{i-1}}\right)}\right)^2 \\ V_i &\geqslant \left(\sqrt{\mathbb{V}\text{ar}\left(P - \hat{P}^{n_i}\right)} - \sqrt{\mathbb{V}\text{ar}\left(P - \hat{P}^{n_{i-1}}\right)}\right)^2 . \end{aligned}$$

As $n_i > n_{i-1}$, under the assumption (15) $\mathbb{V}\text{ar}\left(P - \hat{P}^{n_{i-1}}\right)$ asymptotically dominates $\mathbb{V}\text{ar}\left(P - \hat{P}^{n_i}\right)$ and, thus,

$$V_i = \mathbb{V}\text{ar}\left(P^{n_i} - P^{n_{i-1}}\right) \asymp n_{i-1}^{-\beta} \quad \text{for } i = 1, 2, \ldots, L\,.$$

Morover we assume that $\mathbb{V}\text{ar}\left(\hat{P}^{n_0}\right) \asymp 1$, as it converges to the $\mathbb{V}\text{ar}\,[P]$, e.g. due to assumption (15) and $\left(\mathbb{V}\text{ar}\left(\hat{P}^{n_0}\right)\right)^2 \leqslant \left(\mathbb{V}\text{ar}\left(\hat{P}^{n_0} - P\right)\right)^2 + \left(\mathbb{V}\text{ar}\,(P)\right)^2$. Hence

$$V_i \asymp \begin{cases} 1 & \text{for} \quad i = 0\,, \\ n_{i-1}^{-\beta} & \text{for} \quad i = 1, 2, \ldots, L\,. \end{cases} \tag{21}$$

Assumption (16) on CC of the scheme directly gives

$$C_i \asymp \begin{cases} n_0^{\gamma} & \text{for} \quad i = 0\,, \\ n_i^{\gamma} + n_{i-1}^{\gamma} & \text{for} \quad i = 1, 2, \ldots, L\,. \end{cases} \tag{22}$$

Furthermore n_i^{γ} asymptotically dominates n_{i-1}^{γ} implying that

$$C_i \asymp n_i^{\gamma}\,. \tag{23}$$

OPTIMAL NUMBERS OF SIMULATED PATHS

On the first step we determine the numbers of simulated paths on each level $\bar{N}_i$ that minimise computational effort to obtain the estimate with the statistical error equal to ε^2.

First we solve this problem in the general settings, i.e. involving only C_i and V_i. Using representations (19) and (20) for $\mathbb{V}\mathrm{ar}\left(\hat{E}\right)$ and $C.C.\left[\hat{E}\right]$ correspondingly it reads as

$$\begin{aligned} &\sum_{i=0}^{L} N_i C_i \to \min \\ \text{s.t. } &\sum_{i=0}^{L} N_i^{-1} V_i = \varepsilon^2 \, . \end{aligned}$$

If the N_i are treated as a real numbers, this problem can be solved using the Lagrange multiplier optimisation technique (see e.g. [14]). Detailed steps of the solution may be found in Section A.1.

This yields the optimal numbers of simulated paths

$$\bar{N}_i = \frac{\sum_{i=0}^{L} \sqrt{C_i \cdot V_i}}{\varepsilon^2} \sqrt{\frac{V_i}{C_i}} \, . \tag{24}$$

and the minimal CC of the MLC estimate is

$$C.C.\left[\hat{E}\right] = \left(\sum_{i=0}^{L} \sqrt{C_i V_i} \right)^2 \varepsilon^{-2} \, . \tag{25}$$

In particular, under our assumptions combining (21) and (23) in Equation 24 and Equation 25 we obtain

$$\bar{N}_i \asymp \begin{cases} \frac{n_0^{\gamma/2} + \sum_{i=1}^{L} n_i^{\gamma/2} n_{i-1}^{-\beta/2}}{\varepsilon^2} \, n_0^{-\gamma/2} & \text{for} \quad i = 0 \, , \\ \frac{n_0^{\gamma/2} + \sum_{i=1}^{L} n_i^{\gamma/2} n_{i-1}^{-\beta/2}}{\varepsilon^2} \left(n_i^{-\gamma/2} n_{i-1}^{-\beta/2} \right) & \text{for} \quad i = 1, 2, \dots, L \, . \end{cases}$$

and the corresponding CC

$$C.C.\left[\hat{E}\right] \asymp \left(n_0^{\gamma/2} + \sum_{i=1}^{L} n_i^{\gamma/2} n_{i-1}^{-\beta/2} \right)^2 \varepsilon^{-2} \, . \tag{26}$$

It is worth to comment that a spread of the computational effort across levels is proportional to

$$C.C.\left[\hat{Y}_i\right] \sim \sqrt{C_i V_i} \, .$$

Hence it depends on a mutual relation between the rate of V_i's decay and the rate of C_i's growth. Moreover it will turn out that this relation has an essential impact on CC of the MLC.

OPTIMAL NUMBERS OF DISCRETISATION POINTS

Further we are minimising the CC of the estimate w.r.t. the numbers of discretisation points $n_0, n_1, \ldots, n_{L-1}$. To do so it suffices to solve

$$\left(n_0^{\gamma/2} + \sum_{i=1}^{L} n_i^{\gamma/2} n_{i-1}^{-\beta/2}\right)^2 \to \min ,$$

as the ε^{-2} term on the RHS of Equation 26 is independent from n_i. Note, the number of points for the finest discretisation n_L is fixed to be n and not subject to minimisation. While we agreed to treat n_i as real numbers, the continuous optimisation technique can be applied. Here we present the solution and comment on it (more details can be found in Section A.2).

For the sake of notational simplicity we define

$$a := \frac{\beta}{\gamma} \qquad \text{and} \qquad b := \left(\frac{\beta}{\gamma}\right)^{\frac{2}{\gamma}} .$$

Then the optimal numbers of discretisation points

$$\bar{n}_i = \begin{cases} n_L^{\frac{i+1}{L+1}} & \text{if} \quad a = 1 , \\ b^{\frac{(i+1)\left(a^{L+1}-1\right)-(L+1)\left(a^{i+1}-1\right)}{\left(a^{L+1}-1\right)\ (a-1)}} \quad n_L^{\frac{a^{i+1}-1}{a^{L+1}-1}} & \text{if} \quad a \neq 1 , \end{cases} \tag{27}$$

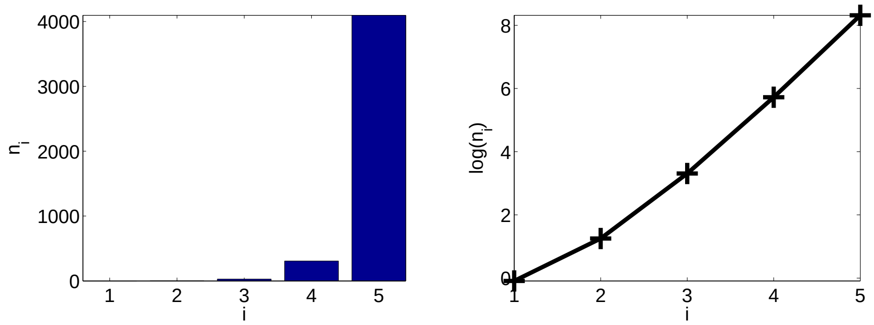
and the corresponding minimised CC reads as

$$\mathrm{C.C.}\left[\hat{E}\right] \asymp \begin{cases} \left((L+1)\, n_L^{\frac{1}{L+1}\frac{\gamma}{2}}\right)^2 \varepsilon^{-2} & \text{if} \quad a = 1 , \\ \left(\frac{a^{L+1}-1}{a^{L+1}-a^L}\ a^{\left(\frac{1}{(a-1)} - \frac{(L+1)}{a^{L+1}-1}\right)} n_L^{\frac{a-1}{a^{L+1}-1}\frac{\gamma}{2}}\right)^2 \varepsilon^{-2} & \text{if} \quad a \neq 1 . \end{cases} \tag{28}$$

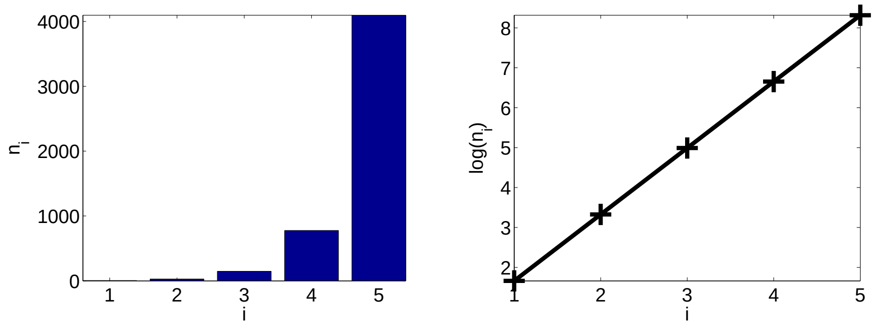
Also it can be shown that for the optimal $\bar{n}_i$ we have

$$\bar{n}_i = \begin{cases} \bar{n}_0 \bar{n}_{i-1} & \text{if} \quad a = 1 , \\ b^{-i} \bar{n}_{i-1}^{a} & \text{if} \quad a \neq 1 . \end{cases}$$

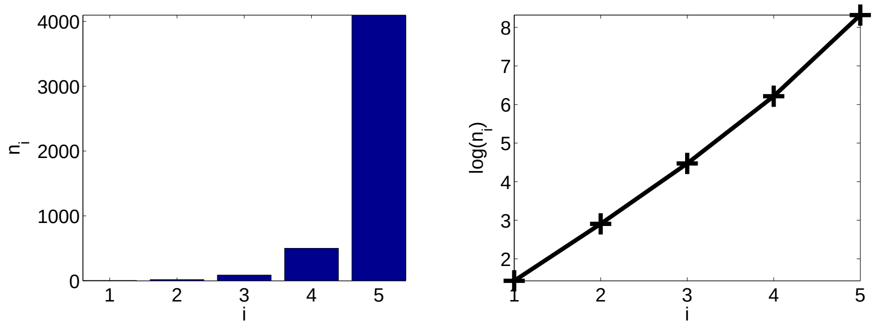
This recursive relation implies that in general the optimal sequence of $\bar{n}_i$ does not obey a geometric progression rule (see Figure 1), while this is the case for the MLMC.



(a) $\beta = 2$



(b) $\beta = 1$



(c) $\beta = 0.5$

Figure 1: Optimal $\bar{n}_i$ for $n_L = 4096$, $L = 4$, $\gamma = 1$ and different values of β

OPTIMAL NUMBER OF LEVELS

During the last step we minimise CC w.r.t. the number of levels L, i.e.

$$\mathrm{C.C.}\left[\hat{E}\right] \to \min ,$$

where $\mathrm{C.C.}\left[\hat{E}\right]$ as defined in Equation 28. Thus to find an optimal L we have to minimise the function G

$$G(L) \asymp \begin{cases} \left((L+1)\, n_L^{\frac{1}{L+1}\frac{\gamma}{2}} \right)^2 & \text{if} \quad a = 1 , \\ \left(\frac{a^{L+1}-1}{a^{L+1}-a^L}\, a^{\left(\frac{1}{(a-1)} - \frac{(L+1)}{a^{L+1}-1}\right)} n_L^{\frac{a-1}{a^{L+1}-1}\frac{\gamma}{2}} \right)^2 & \text{if} \quad a \neq 1 . \end{cases} \tag{29}$$

This problem can be solved analytically for L being treated as real number (for details see Section A.3).

Defining

$$R := \begin{cases} e^{\frac{2}{\gamma}} & \text{if} \quad \gamma = \beta , \\ \left(\frac{\gamma}{\beta}\right)^{\frac{2}{\gamma-\beta}} & \text{if} \quad \gamma \neq \beta , \end{cases} \tag{30}$$

the optimal number of levels is given as

$$\bar{L} = \log_{[R]}(n_L) - 1 . \tag{31}$$

For this choice of $\bar{L}$ the optimal numbers of discretisation points $\bar{n}_i$ take an elegant form of the geometric progression with the common ratio R

$$\bar{n}_i = R^{i+1} . \tag{32}$$

Moreover they obey the following recursive relation

$$\bar{n}_i = R^{-1}\, \bar{n}_{i+1} \tag{33}$$

and the corresponding CC is asymptotically equal to

$$\mathrm{C.C.}\left[\hat{E}\right] \asymp \begin{cases} \left(e \ln\left(n_L^{\frac{\gamma}{2}} \right)\right)^2 \varepsilon^{-2} & \text{if} \quad \gamma = \beta , \\ \left(\frac{a^{\frac{a}{a-1}}}{a-1} \right)^2 \left(n_L^{\frac{\gamma-\beta}{2}} - 1 \right)^2 \varepsilon^{-2} & \text{if} \quad \gamma \neq \beta . \end{cases}$$

Note, for a fixed γ the optimal common ration R is a decreasing function of β implying that for schemes with a slow strong convergence of an order $\beta < 1$ the optimal value of R takes significantly bigger values compared to the case of a fast convergent schemes (see Figure 2).

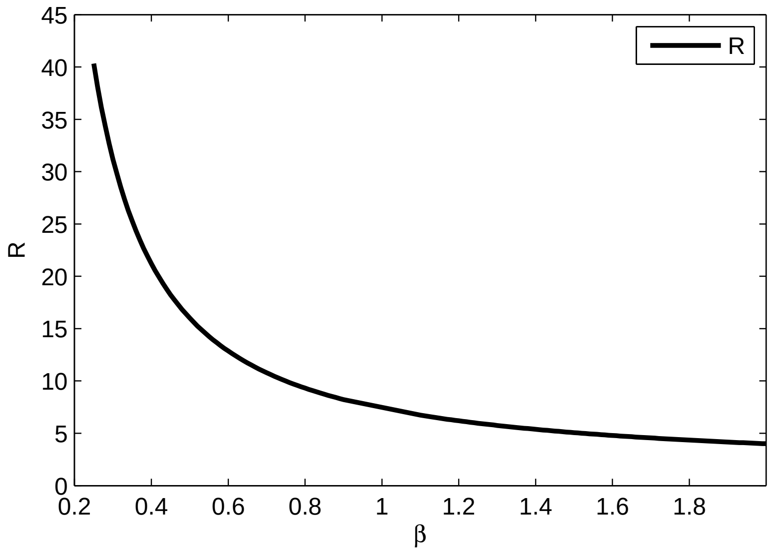


Figure 2: Optimal common ratio R for $\gamma = 1$ and different values of β.

Moreover it is very important that R is independent of n_L and ε, but depends on the properties of an approximation scheme.

The optimal number of levels L is probably the most sensitive to a rounding issue, since it typically takes small values. However the function $G(L)$ is strictly decreasing for $L \in (0, \bar{L})$ and strictly increasing for $L \in (\bar{L}, +\infty)$. Therefore an optimal solution of the corresponding integer minimisation problem will be either $\lfloor \bar{L} \rfloor$ or $\lceil \bar{L} \rceil$.

2.1.2 *MLC Algorithm*

As the values of $\bar{n}_i$, $\bar{N}_i$ and $\bar{L}$ are not necessarily integer, in this section we present their integer counterparts, that should be used in the MLC algorithm.

In the first place the number of levels L should be chosen as

$$L = \operatorname{argmin}\left(G(L) \mid L \in \left\{ \lfloor \bar{L} \rfloor, \lceil \bar{L} \rceil \right\} \right), \tag{34}$$

where the function G is as defined in Equation 29. In general, the L is no longer equal to $\bar{L}$, thus the numbers of discretisation points $\bar{n}_i$ should be chosen according to Equation 27. Besides we propose to round them to obtain integer n_i

$$n_i = \text{round}\{\bar{n}_i\} . \tag{35}$$

Remark 13. *Although the above choice of n_i is the most intuitive one, there exist other reasonable options. For instance, being inspired by the recurrent representation (33), the numbers of discretisation points can be chosen to follow a decreasing geometric progression (up to rounding). Namely*

$$n_i = \text{round}\left\{ R^{-(L-i)} n_L \right\} ,$$

where R *is the optimal common ratio given by Equation 30.*

According to Equation 24 the optimal numbers of simulated paths depends on V_i and C_i. As V_i are not known in advance , they have to be estimated on the fly. Then combining the expression (22) for C_i with Equation 24 gives

$$N_i = \begin{cases} \left\lceil \frac{\sqrt{n_0^\gamma V_0}+\sum_{i=1}^{L}\sqrt{(n_i^\gamma+n_{i-1}^\gamma)V_i}}{\varepsilon^2}\sqrt{\frac{V_0}{n_0^\gamma}} \right\rceil & \text{for} \quad i=0\,, \\ \left\lceil \frac{\sqrt{n_0^\gamma V_0}+\sum_{i=1}^{L}\sqrt{(n_i^\gamma+n_{i-1}^\gamma)V_i}}{\varepsilon^2}\sqrt{\frac{V_i}{n_i^\gamma+n_{i-1}^\gamma}} \right\rceil & \text{for} \quad i=1,\dots,L\,. \end{cases} \tag{36}$$

To estimate V_i on the fly we propose to use a semi-adaptive algorithm in style of the adaptive MLMC algorithm [28].

Then we summarise our suggestions in Algorithm 1.

Algorithm 1 Multilevel Constructed Estimate

Input: ε and n_L
Output: $\hat{E}$

1: define optimal L using Equation 34
2: define optimal $n_0,\dots,n_{L-1}$ using Equation 35
3: estimate $V_0,\dots,V_L$ using an initial $N_l = 10^4$ samples
4: define optimal $N_0,\dots,N_L$ using Equation 36
5: evaluate extra samples at each level as needed for new N_i
6: calculate $\hat{E}$ according to Equation 17

Remark 14. *In general there exit sets of* n_L*,* γ *and* β *for which the above procedure gives degenerate optimal* n_i*, e.g.* $n_i = n_{i-1}$ *for some* i *or* $n_0 = 0$*. However in applications this procedure is usually well-defined.*

2.1.3 *Computational Cost of the MLC estimate*

In this section we give a main result on CC of the MLC Algorithm 1.

Proposition 1. *Let* P *denote a functional of the solution of some stochastic differential equation driven by Brownian motion, and let* $\hat{P}^m$ *be some approximating discretisation scheme with* m *equidistant discretisation points.*

Moreover let for some positive constants β, γ, c_c *and* c_v *the discretisation scheme* $\hat{P}^m$ *is such that* $\forall m \in \mathbb{N}$

i) $\mathrm{C.C.}\left[\hat{P}^m\right] \leqslant c_c m^\gamma$,

ii) $\mathbb{V}\mathrm{ar}\left(P - \hat{P}^m\right) \leqslant c_v m^{-\beta}$,

iii) $\mathbb{V}\mathrm{ar}\left(\hat{P}^m\right) \leqslant c_v$.

Then for $n_L \geqslant R^2$ *and sufficiently small* ε *there exists a choice of* N_i, n_i *and* L *such that the unbiased MLC estimate of* $\mathbb{E}\left(\hat{P}^{n_L}\right)$

$$\hat{E} = \frac{1}{N_0}\sum_{j=1}^{N_0}\hat{P}_j^{n_0} + \sum_{i=1}^{L}\frac{1}{N_i}\sum_{j=1}^{N_i}\left(\hat{P}_j^{n_i} - \hat{P}_j^{n_{i-1}}\right),$$

has a statistical error

$$\mathbb{V}\mathrm{ar}\left(\hat{E}\right) \leqslant \varepsilon^2$$

and a bounded computational cost

$$\mathrm{C.C.}\left[\hat{E}\right] \leqslant \begin{cases} C \quad \varepsilon^{-2} & \text{if} \quad \gamma < \beta\,, \\ C\left(\ln\left(n_L\right)\right)^2 \quad \varepsilon^{-2} & \text{if} \quad \gamma = \beta\,, \\ C n_L^{\gamma-\beta} \quad \varepsilon^{-2} & \text{if} \quad \gamma > \beta\,, \end{cases}$$

for some constant C *independent of* n_L *and* ε.

The proof relies on the choice of N_i, n_i and L from Section 2.1.2. However it is rather technical, thus, given in the appendix (see Section A.4). Note, the requirement of ε to be sufficiently small is rather technical and usually satisfied in applications. It is imposed to ensure that for all levels the numbers of simulated paths N_i computed according to Equation 36 are at least greater or equal than 1. Moreover we consider n_L such that the optimal number of levels (34) is at least greater or equal than 1, otherwise it is advisable to use the plain MC estimate instead.

The result of Proposition 1 is in line with the Complexity Theorem 7 for the MLMC. For instance, CC of the estimate is bounded regardless of n_L if $\gamma < \beta$, what corresponds to an asymptotically negligible CC of an extra level of the MLMC in this case. Moreover the spread of computational effort across levels of the MLC estimate is defined by γ and β in the similar to the MLMC way:

- if $\gamma < \beta$, more effort is assigned to starting levels;
- if $\gamma = \beta$, effort is almost uniformly spread;
- if $\gamma > \beta$, more effort is assigned to last levels.

We will continue the discussion on the relation of the MLC to the MLMC in Section 2.3.

Remark 15. *The constant* C *obtained in the proof of Proposition 1 is very conservative and by far not tight. Moreover for a big enough* n_L *and a small enough* ε *it is expected to be about*

$$C \approx \begin{cases} c_c c_v \left(R^{\gamma/2} + \frac{\sqrt{(1+R^{-\gamma})(R^{\beta}-1)}}{a^2-a} \right)^2 & \textit{if} \quad \gamma < \beta \,, \\ c_c c_v \left(e^2 - e^{-2}\right) \frac{\gamma^2}{4} & \textit{if} \quad \gamma = \beta \,, \\ c_c c_v \frac{(1+R^{-\gamma})\left(R^{\beta}-1\right)}{(a-1)^2} & \textit{if} \quad \gamma > \beta \,, \end{cases}$$

for the choice of N_i, n_i *and* L *according to* (36), (35), (34) *correspondingly (for a justification see Section A.4).*

2.1.4 *Comment on the Structural Result*

In general to obtain a multilevel estimator with CC that is bounded as in Proposition 1 it is sufficient that

- $\bar{n}_i$ are elements of a geometric progression with some common ratio $r > 1$;
- and $\bar{L}$ is the logarithm of n_L w.r.t. the base r.

However the constant C will be bigger, but still independent of n_L and ε. Also it can be proven for the choice of n_i as given in Remark 13.

An important by-product of the optimisation is

- the optimal common ratio R,

that minimises the constant C. This can be seen as a generalisation of the optimal M for the MLMC derived in Section 4.1 of [28]. We would like to stress it again: the optimal common ratio is independent from errors and depends only on the properties of an approximation scheme.

Selection of an appropriate starting level is another important issue that has an impact on CC of the multilevel methods [55]. Our optimisation approach also gives notion of the optimal starting level. Though it has got practical drawbacks (for details see following remark).

Remark 16 (On The Optimal Starting Level)**.** *According to our findings* $\bar{n}_0$ *is equal to* $R^1 > 1$*, meanwhile in the MLMC method it is chosen to be* $M^0 = 1$*. This is explained by the simplification of our assumptions on values of* V_0 *and* V_1*. To be more precise for* $\bar{n}_0 = 1$

$$V_0 = c_v \qquad \textit{and} \qquad V_1 = c_v \bar{n}_0^{-\beta} = c_v \,,$$

while

$$C_0 = c_c \bar{n}_0^{\gamma} = c_c \qquad \textit{and} \qquad C_1 = c_c \bar{n}_1^{\gamma} \gg c_c \,.$$

Then, as $V_0 = V_1$ *while* $C_0 \ll C_1$*, this choice a priory cannot be optimal under our assumptions. However in applications it often happens that* $V_0 \gg V_1$ *and, thus, the optimal choice of* $\bar{n}_0$ *could be less than* R*.*

This effect can be incorporated in the solution of the optimisation Problem (18) *and will lead to the notion of an* optimal starting level. *Under the assumption of* $\mathbb{V}\mathrm{ar}\left(\hat{P}^m - P\right) \approx c_v m^{-\beta}$ *the optimal* $\bar{n}_i$ *and* $\bar{L}$ *will be*

$$\bar{L} = \log_{[R]}\left(\hat{c}^{\frac{1}{\beta}}\, n_L\right) - 1\,,$$
$$\bar{n}_i = \hat{c}^{-\frac{1}{\beta}}\, R^{i+1}\,,$$

where

$$\hat{c} = \frac{\mathbb{V}\mathrm{ar}\left(\hat{P}^{n_0}\right)}{c_v}\,.$$

Note, the optimal common ratio R *is the same and still is independent of* n_L*,* ε *and the aforementioned constants.*

However this finding has got serious practical drawbacks for a practical application. In particular

- *the constant* $\hat{c}$ *depends on the unknown in advance* c_v *and* $\mathbb{V}\mathrm{ar}\left(\hat{P}^{n_0}\right)$*, and furthermore the last constant depends on the solution itself;*
- *optimality essentially relies on the fact that* $\mathbb{V}\mathrm{ar}\left(\hat{P}^m - P\right) \approx c_v\, m^{-\beta}$*, what is sometimes not the case (e.g. for barrier options under the Heston model approximated by the Euler scheme [55]).*

As the recursive relation $\bar{n}_{i-1} = R^{-1}\bar{n}_i$ *still holds and* R *depends only on* γ *and* β*, the first drawback can be eliminated by computing an estimator backwards, i.e. starting from the last levels and estimating unknown constants on the fly.*

2.2 MLC AS VARIANCE REDUCTION

The first application on the MLC is a direct estimation of

$$\mathbb{E}\left(\hat{P}^n\right)\,.$$

In such a setting the MLC estimate can be seen as an optimal multistep control variate variance reduction technique. The reason is that n_i, N_i and L are chosen to minimise the variance of the estimate for a fixed CC (see Remark 12). To measure the improvement we introduce a *speed up* of the MLC compared to the plain MC algorithm yielding an estimate with the same statistical error ε^2

$$\text{speed up} = \frac{\text{C. C. [“plain MC”]}}{\text{C. C. [“MLC”]}}\,.$$

As CC of the plain MC estimator is equal to $c_c n_L \mathbb{V}\text{ar}\left(\hat{P}^{n_L}\right)\varepsilon^{-2}$, under the assumptions of Proposition 1 it holds that

$$\text{speed up} \geqslant \begin{cases} \widetilde{C} \quad n_L^{\gamma} & \text{if} \quad \gamma < \beta\,, \\ \widetilde{C} \quad n_L^{\gamma} \ \left(\ln\left(n_L^{\gamma}\right)\right)^{-2} & \text{if} \quad \gamma = \beta\,, \\ \widetilde{C} \quad n_L^{\beta} & \text{if} \quad \gamma > \beta\,, \end{cases}$$

for some positive constant $\widetilde{C}$ independent of n_L and ε. Hence the speed up increases (upto log-terms) polynomially in n_L with a degree of $\min\{\gamma, \beta\}$.

Remark 17. *Note, the above constant $\widetilde{C}$ is equal to*

$$\widetilde{C} = C^{-1} c_c \mathbb{V}\text{ar}\left(\hat{P}^{n_L}\right)\,,$$

where C is as in Remark 15. Then under the assumptions of Proposition 1 the constant $\widetilde{C}$ is expected to be about

$$\widetilde{C} \approx \begin{cases} \frac{\mathbb{V}\text{ar}\left(\hat{P}^{n_L}\right)}{c_v}\left(R^{\gamma/2} + \frac{\sqrt{(1+R^{-\gamma})(R^{\beta}-1)}}{a^2-a}\right)^{-2} & \textit{if} \quad \gamma < \beta\,, \\ \frac{\mathbb{V}\text{ar}\left(\hat{P}^{n_L}\right)}{c_v}\frac{1}{e^2-e^{-2}}\frac{4}{\gamma^2} & \textit{if} \quad \gamma = \beta\,, \\ \frac{\mathbb{V}\text{ar}\left(\hat{P}^{n_L}\right)}{c_v}\frac{(a-1)^2}{(1+R^{-\gamma})(R^{\beta}-1)} & \textit{if} \quad \gamma > \beta\,, \end{cases}$$

for a big enough n_L and a small enough ε.

2.2.1 *Numerical Example*

We illustrate the speed up effect by analysing the numerical performance of MLC Algorithm 1. As a show case we price arithmetic discrete Asian call option averaged over daily closings (see Section 1.1.1.2) in the one dimensional BS model (see Section 1.1.2.1) with

today's price of asset	$S_0 = 1\,,$
riskless interest rate	$r = 0.02\,,$
volatility	$\sigma = 0.25\,.$

In general the MC methods might be not the best numerical technique in this setting (for alternatives see e.g. [44] and references therein).

We make use of the closed form (2) of a solution S to the BS SDE (1). Then the numerical approximation $\hat{P}^m$ is defined as

$$\hat{P}^m := e^{-rT}\left(\frac{1}{m}\sum_{i=1}^{m} S\left(\frac{T}{m}\,i\right) - K\right)^{+}\,,$$

L	n_0	n_1	n_2	n_3	n_4
4	4	16	67	283	1260

Table 1: Optimal n_i and L

γ	β	R	n_1/n_0	n_2/n_1	n_3/n_2	n_4/n_3
1	2	4	4.0	≈ 4.2	≈ 4.2	≈ 4.5

Table 2: Commom ratio

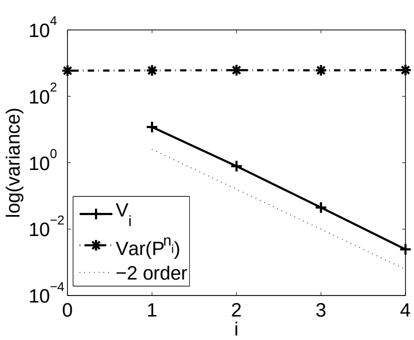
where S is a solution of the BS SDE. Note, that for $n_L = 252T$ the approximation $\hat{P}^n$ is an exact discounted payoff of the arithmetic discrete Asian call option averaged over daily closings with maturity T and strike K. Therefore we are interested in estimation of $\mathbb{E}\left(\hat{P}^{n_L}\right)$.

In this case CC of the approximation $\hat{P}^m$ is proportional to m, i.e. $\gamma = 1$ in the notations of Proposition 1, and the rate of variance decay β is equal to 2 (see Figure 3a).

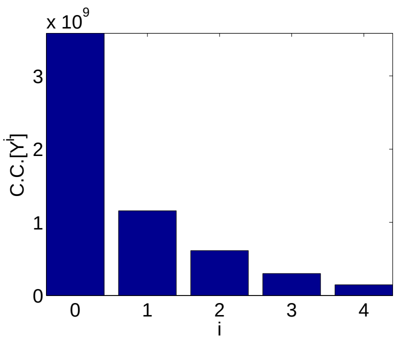
First we analyse pricing of the option with time to maturity $T = 5y$ and strike $K = 100$. In particular $n_L = 1260$ and the price of such a product is approximately 13.96.

The optimal configuration of n_i and L as chosen by the MLC Algorithm 1 is presented in Table 1. Note it is infeasible to chose n_i to be equal to the optimal continuous solutions $\bar{n}_i$. However Table 2 suggests that n_i are close to a geometric progression with the common ratio $R = 4$.

On Figure 3a we plot variances V_i and $\mathbb{V}\text{ar}\left(\hat{P}^{n_i}\right)$. The spread of



(a) variances



(b) computation spread

Figure 3: Performance of the MLC

computational effort across the levels, i.e. $C.C.\left[\hat{Y}_i\right]$, is showed on Figure 3b. As expected more computational effort is assigned to the starting levels because $\beta > \gamma$. Moreover the MLC is about 120-times

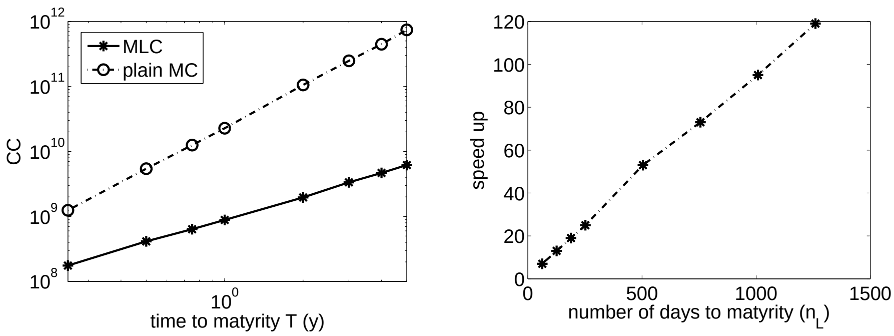
T (y)	0.25	0.5	0.75	1	2	3	4	5
# of days (n_L)	63	126	189	252	504	756	1008	1260
speed up	7	13	19	25	53	73	95	119

Table 3: Speed up of the MLC compare to the plain MC

faster than the plain MC method, while Remark 17 predicts 180-times speed up.

In general, the speed up depends on a value of n_L. In Table 3 we compare speed ups for different times to maturity T given.

As the approximation $\hat{P}^m$ depends on T, what we are doing is not a comparison of the MLC computing $\mathbb{E}\left(\hat{P}^m\right)$ for a different m. However this still confirms an expected linear increase of the speed up with increasing n_L (see Figure 4).



(a) CC (b) speed up

Figure 4: Speed up of the MLC compare to the plain MC

2.3 MLC AS THE CLASSIC MLMC

Here we are showing the way the MLC and the classic MLMC are related. First we will explain how Proposition 1 leads to the Complexity Theorem 7 for the MLMC. Then we present the way to construct the adaptive MLMC algorithm [28] from the MLC algorithm. As a by-product we obtain a new interpretation of the adaptive MLMC algorithm in the character of an unbounded search problem.

To take into account a bias error we assume that the approximation scheme $\hat{P}^m$ has an order of week convergence $\alpha > 0$, i.e.

$$\left|\mathbb{E}\left(P\right) - \mathbb{E}\left(\hat{P}^m\right)\right| \leqslant c_b m^{-\alpha} \tag{37}$$

for some positive constant $c_b >$ that is independent from m. Then it directly follows that the choice of

$$n_L = \left\lceil \left(c_b^{-1}\ \varepsilon\right)^{-\frac{1}{\alpha}} \right\rceil \tag{38}$$

ensures that the bias error of $\mathbb{E}\left(\hat{P}^{n_L}\right)$ is less than ε. Moreover under the assumptions of Proposition 1, there exists an MLC estimate $\hat{E}$ with the statistical error less than ε^2 such that its CC is bounded by

$$\text{C.C.}\left[\hat{E}\right] \leqslant \begin{cases} C \quad \varepsilon^{-2} & \text{if} \quad \gamma < \beta\,, \\ C\,(\,\ln(n_L))^2 \quad \varepsilon^{-2} & \text{if} \quad \gamma = \beta\,, \\ C n_L^{\gamma-\beta} \quad \varepsilon^{-2} & \text{if} \quad \gamma > \beta\,. \end{cases}$$

Then if n_L is chosen according to Equation 38, the MSE decomposition into the sum of a statistical error and squared bias (see Remark 7) implies that

$$\mathbb{MSE}\left[\hat{E}\right] = \mathbb{V}\text{ar}\left(\hat{E}\right) + \left(\mathbb{E}\left(P\right) - \mathbb{E}\left(\hat{P}^{n_L}\right)\right)^2 \leqslant \varepsilon^2 + \varepsilon^2 = 2\varepsilon^2$$

and the corresponding CC of the estimate $\hat{E}$ is bounded by

$$\text{C.C.}\left[\hat{E}\right] \leqslant \begin{cases} \bar{C} \quad \varepsilon^{-2} & \text{if} \quad \gamma < \beta\,, \\ \bar{C} \quad \varepsilon^{-2}\ (\ln(\varepsilon))^2 & \text{if} \quad \gamma = \beta\,, \\ \bar{C} \quad \varepsilon^{-2-\frac{\gamma-\beta}{\alpha}} & \text{if} \quad \gamma > \beta\,. \end{cases}$$

This CC bound is in agreement with the generalised Complexity Theorem for the MLMC [55] and in particular for $\gamma = 1$ is identical to the one from the Complexity Theorem 7.

BIAS ERROR AND UNBOUNDED SEARCH

According to Equation 38 to chose the appropriate n_L we should know the constant c_b, which is a priori unknown . Hence, the following search problem has to be solved

$$n_L = \min\left\{m \in \mathbb{N}\,, \quad \text{s.t.}\ \left|\mathbb{E}\left(P\right) - \mathbb{E}\left(\hat{P}^m\right)\right| \leqslant \varepsilon\right\}\ . \tag{39}$$

This problem can be regarded as the well-known *unbounded search* problem. It was first introduced and studied by Bentley and Yao [9] (for a general introduction to the class of the search problems see [40]). The problem can be seen as searching for a key in an ordered table of infinite size. To give an exact formulation of the problem we define the key function $F : \mathbb{N} \to \{0, 1\}$

$$F(i) = \begin{cases} 0 & \text{for} \quad i < n\,, \\ 1 & \text{for} \quad i \geqslant n\,. \end{cases}$$

The aim of the unbounded search is to determine n using primitive operations of evaluation F and comparing its value to 0 or 1.

In our setting the key function F is defined as

$$F(i) = \mathbb{I}\left(\left|\mathbb{E}\left(P\right) - \mathbb{E}\left(\hat{P}^i\right)\right| \leqslant \varepsilon\right)\ .$$

It is well-defined under the assumption of a monotonic decay of the bias error of the scheme $\hat{P}^m$. This link explains the unbounded search nature of the problem (39).

Bentley and Yao in [9] gave a lower bound for the cost, measured by the number of comparisons, of unbounded search algorithms and proposed an almost optimal algorithm for this problem. The almost optimal algorithm consists of a sequence of the *binary search* algorithms. The first binary search algorithm began with determining m, s.t. $2^{m-1} \leqslant n \leqslant 2^m - 1$ by successively evaluating $F\left(2^i - 1\right)$ and then it is complemented by the classic binary search within 2^{m-1} elements.

This motivates us to apply this idea to the problem 39. Namely, by successively evaluating $\mathbb{I}\left(\left|\mathbb{E}(P) - \mathbb{E}\left(\hat{P}^{R^l}\right)\right| \leqslant \varepsilon\right)$ determine L, s.t. $R^{L-1} \leqslant n_L \leqslant R^L - 1$ and then set n_L to be R^L. In the case of a scheme with a smooth first order weak convergence the bias entering the aforementioned indicator can be approximated (see [28] Section 4.2) by

$$\left|\mathbb{E}(P) - \mathbb{E}\left(\hat{P}^{R^l}\right)\right| \approx \frac{1}{R^2 - 1}\left|\mathbb{E}\left(\hat{P}^{R^l}\right) - R^{-1}\mathbb{E}\left(\hat{P}^{R^{l-1}}\right)\right| .$$

Hence, $\mathbb{I}\left(\left|\mathbb{E}(P) - \mathbb{E}\left(\hat{P}^{R^l}\right)\right| \leqslant \varepsilon\right)$ can be heuristically substituted by

$$\mathbb{I}\left(\left|\mathbb{E}\left(\hat{P}^{R^l}\right) - R^{-1}\mathbb{E}\left(\hat{P}^{R^{l-1}}\right)\right| \leqslant \left(R^2 - 1\right)\varepsilon\right) .$$

Further $\mathbb{E}\left(\hat{P}^{R^l}\right)$ is approximated using the MLC estimate. Also in this setting

- $\mathbb{E}\left(\hat{P}^{R^{l-1}}\right)$ is approximated by a sample mean of the level before last one;
- the approximation of $\mathbb{E}\left(\hat{P}^{R^{l+1}}\right)$ adds one extra level to the computation of $\mathbb{E}\left(\hat{P}^{R^l}\right)$ and reuses its samples.

Then by omitting binary search for n_L with in $R^{L-1}, \ldots, R^L - 1$ we are getting rid of any extra CC of the MLC estimate that has the MSE less than $2\varepsilon^2$.

This approach leads to the adaptive MLMC algorithms (see [28] Section 5) proposed by Giles. Therefore this algorithms can be interpreted as a combination of the (asymptotically) optimal multi-step control variate estimate MLC with the (upto constants) optimal unbounded search algorithm.

3 WEAK EXTRAPOLATION AND MLMC

We continue to consider a problem of estimation of an expected value of a functional of the solution of an SDE using the MC methods. In particular this formulation reflects the problem of numerical pricing of exotic options (see Section 1.1.1.2).

We further stick to the assumption that some approximation scheme should be used (see Section 1.2.2 for examples). As a consequence an overall MSE of an estimate consist of two components:

- statistical error;
- bias error induced by an inexact discretisation scheme.

Moreover MSE can be decomposed into a sum of the aforementioned components (see Section 1.2.2.2), namely

$$\mathbb{MSE}\left[\hat{E}\right] = \underbrace{\mathbb{V}\mathrm{ar}\left(\hat{E}\right)}_{\text{stat. error}} + \Big(\underbrace{\mathbb{E}\left(\hat{E}\right) - \mathbb{E}\left(P\right)}_{\text{bias}}\Big)^2 .$$

This in particular means that both errors should be controlled simultaneously and consistent. For instance in the case of the plain MC estimate the number of simulated paths should be increased to decrease the statistical error, while increasing the number of discretisation steps leads to a smaller bias. However these both straightforward controls lead to substantial increase of CC of the estimate, thus they should be done in an optimal way (see e.g. [22]). On the contrary there exist error reduction techniques, that reduce errors with a minor impact on CC of the estimate, e.g. variance reduction techniques dealing with statistical error and weak extrapolation methods improving bias error convergence. General surveys on existing variance and bias reduction techniques can be found in e.g. [45] or [29].

In Chapter 2 we argued that the MLMC approach can be seen as a multi-step control variate variance reduction. While it is proven to be an effective variance reduction, it does not target reduction of a bias error. Therefore the MLMC estimator will be taken as a base and enhanced with a weak extrapolation bias reduction technique.

Weak extrapolation methods are well known bias reduction techniques for the MC method in the SDE setting.

Remark 18. *There exist alternatives for speeding up bias error convergence. For instance*

- *schemes with faster bias convergence;*

- *schemes with fast strong convergence in the MLMC settings, as in the case of* $\beta > 1$ *a rate of bias convergence has no influence on the CC of the estimate (see the Complexity Theorem 7).*

However schemes with higher orders of bias or strong convergence might be unavailable or, which is often the case, are more sophisticated from the implementation point of view [53]. Moreover they might be unsuitable for an efficient hardware implementation (see [56] and references therein), what seems to be a very recent trend [49].

In the SDE setting weak extrapolation was first proposed and justified by Talay and Tubaro in [63]. The main idea is to linearly extrapolate the bias error using two approximation schemes with n and $2n$ discretisation points. Under appropriate assumptions on the bias convergence it is possible to construct high order weak extrapolation (HOWE) approximations, e.g. quadratic, cubic and so forth. However a straightforward extension based on schemes with $n, 2n, 3n, \ldots$ discretisation points has essential drawbacks that prevent it from being used in applications [29]. In particular they lead to explosion of variance of the estimate and as a consequence enormous CC [59].

In this chapter the HOWE methods are applied to enhance the MLMC estimate proposed in [28]. This merge of the HOWE and the MLMC constructions circumvents the aforementioned drawbacks of the HOWE. We will show that if the bias error admits a power series expansion in terms of the number of discretisation points, CC to achieve an RMSE of ε for the extended estimate can be reduced to at least

$$O\left(\varepsilon^{-(2+o(1))}\right)$$

regardless of the speed of variance decay with increasing levels. This result might be seen as a special case of the Complexity Theorem 7 for a scheme with "infinite" order of weak convergence. However, in contrast to the plain MLMC the extended estimate relies on the weighted sum for the HOWE rather than on the telescopic sum. The reduction of the CC can be explained by a faster bias decay with increasing maximal number of levels.

Debrabant and Rößler in [19] also consider HOWE technique in the MLMC settings. However they applied it to directly to an discrete approximation of a solution of the SDE. In contrast, we extrapolate a functional of the solution. The linear extrapolation in the case of the plain MLMC has been already studied in the seminal paper [28], but results on HOWE applied on the level of functionals were missing in the literature until the very resent and independent result by Lemaire and Pagès [51].

This chapter is build in the following way. We first give a brief recall of the weak extrapolation methods, and afterwards adapt the HOWE to the MLMC settings. In the sequel we define the weak extrapolation multilevel (WEML) estimate, describe the corresponding algorithm and prove the CC theorem. Numerical results presented in Section 3.5 support our findings and illustrate the saving of the computational effort. On the other hand they also show some limitations of the WEML application.

3.1 WEAK EXTRAPOLATION

In this section we briefly recall the idea of the classic Talay-Tubaro (TT) weak extrapolation [63]. The second part of the section is dedicated to the discussion of the HOWE estimate and its drawbacks. In Section 3.1.3 we collect some theoretical results justifying the application of the weak extrapolation technique.

3.1.1 *Linear Talay-Tubaro Extrapolation*

The TT approach is a Richardson type extrapolation (see e.g. [18]). It uses linear extrapolation based on two estimates based on the same scheme but with n and $2n$ discretisation points to obtain an estimate with higher order of a weak convergence.

To be more precise let $\hat{P}^m$ be some approximation scheme based on m equidistant discretisation points (see Section 1.2.2.1) that approximates P. Moreover let for this scheme hold

$$\mathbb{E}\left(P\right) - \mathbb{E}\left(\hat{P}^n\right) = C\, n^{-1} + O\left(n^{-2}\right) , \tag{40}$$

for some constant C independent from n. In particular this means that the scheme has first order weak convergence. Also this implies that the convergence is smooth in the number of discretisation points. Talay and Tubaro first derived necessary conditions on the underlying SDE and functional, such that representation (40) holds for the Euler discretisation scheme. In Section 3.1.3 we recall the main result from [63] and present a brief survey on similar results.

Due to the weak convergence of the scheme our ultimate goal is to approximate $\mathbb{E}\left(P\right)$ by $\mathbb{E}\left(\hat{P}^\infty\right)$. Relation (40) implies that $\mathbb{E}\left(\hat{P}^n\right)$ as a function of n^{-1} is affine upto a second term

$$\mathbb{E}\left(\hat{P}^n\right) \approx \mathbb{E}\left(P\right) - C\, n^{-1} .$$

This motivates the application of linear extrapolation based on approximations with n and $2n$ discretisation points to extrapolate the $\mathbb{E}\left(\hat{P}^\infty\right)$ value. Linear extrapolation directly gives the so-called *Talay-Tubaro scheme*

$$\mathbb{E}\left(\hat{P}^\infty\right) \approx \mathbb{E}\left(2\, \hat{P}^{2n}\right) - \mathbb{E}\left(\hat{P}^n\right) ,$$

that has second order weak convergence, i.e.

$$\mathbb{E}\left(2\hat{P}^{2n} - \hat{P}^n\right) - \mathbb{E}\left(P\right) = O\left(n^{-2}\right) .$$

Consequentially the *Talay-Tubaro estimate* based on this scheme is defined as

$$\bar{E} = \frac{1}{N}\sum_{j=1}^{N}\left(2\hat{P}_j^{2n} - \hat{P}_j^{n}\right) ,$$

and has a bias error of the order two in contrast to first order of the corresponding plain MC estimator.

Remark 19. *An alternative construction of the TT scheme can be seen as an attempt to eliminate the leading term in the bias error by combining approximations with n and $2n$ discretisation points. To be more precise writing Equation 40 for approximations with n and $2n$ discretisation points we obtain*

$$\mathbb{E}\left(P\right) - \mathbb{E}\left(\hat{P}^{n}\right) = Cn^{-1} + O\left(n^{-2}\right) , \tag{41}$$

$$\mathbb{E}\left(P\right) - \mathbb{E}\left(\hat{P}^{2n}\right) = C\left(2n\right)^{-1} + O\left(n^{-2}\right) . \tag{42}$$

Then by doubling Equation 42 and subtracting it from Equation 41 we derive the same TT scheme

$$\mathbb{E}\left(P\right) - \mathbb{E}\left(2\hat{P}^{2n} - \hat{P}^{n}\right) = O\left(n^{-2}\right) ,$$

that has second order of weak convergence.

By the result of Duffie and Glynn [22] on an optimal choice of the number of simulated paths, N should be proportional to n^4 for the TT estimate. Therefore CC of the estimate with RMSE of ε is reduced from $O\left(\varepsilon^{-3}\right)$ for the plain MC to $O\left(\varepsilon^{-2.5}\right)$ for the TT estimate due to the faster order of bias convergence.

Remark 20. *It is advisable to uses consistent Brownian increments to simulate $\hat{P}^{n}$ and $\hat{P}^{2n}$ as this typically reduces variance of the TT scheme. However the effect depends on a correlation between approximations [29].*

3.1.2 *Classic High Order Weak Extrapolation*

Under appropriate assumptions on bias error expansion the weak extrapolation idea can be extended to obtain schemes with higher orders of weak convergence. However we will see that a straightforward generalisation has essential drawbacks.

Let for an approximating scheme $\hat{P}^{n}$ expansion (40) hold for higher orders of n^{-1}, i.e.

$$\mathbb{E}\left(P\right) - \mathbb{E}\left(\hat{P}^{n}\right) = \sum_{i=1}^{N-1} c_i \frac{1}{n^i} + O\left(\frac{1}{n^N}\right) \tag{43}$$

for some positive integer N and real constants c_i, that are independent of n. Typically power-like expansion (43) is proven in full details for $N = 2$. However, under additional technical assumptions it can be shown for larger values of N or even for all positive integers. Also in some applications it still holds despite unavailable theoretical justification.

In what follows we present the straightforward extension of the weak extrapolation approach and recall some results form [63] and [59]. We consider an approximation scheme for N different numbers of discretisation points, i.e.

$$\hat{P}^{n}, \hat{P}^{2n}, \hat{P}^{3n}, \dots, \hat{P}^{Nn}$$

for some positive integer n. The N^{th} *order HOWE scheme* takes the form of a weighted sum of these base approximations

$$\sum_{i=1}^{N} \lambda_i^N \hat{P}^{in},$$

where λ_i^N are real extrapolation weights independent of n. The weights are obtained as solution of the system of equations established in the spirit of Remark 19 (details can be found in [59]) and take the following form

$$\lambda_i^N = (-1)^{N-i} \frac{i^N}{i!\,(N-i)!}.$$

This choice of the weights leads to the HOWE scheme with N^{th} order of weak convergence, i.e.

$$\mathbb{E}(P) - \mathbb{E}\left(\sum_{i=1}^{N} \lambda_i^N \hat{P}^{in}\right) = O\left(n^{-N}\right).$$

As a consequence the *high order weak extrapolation estimate*, defined as

$$\bar{E} = \frac{1}{M} \sum_{j=1}^{M} \lambda_i^N \hat{P}_j^{in},$$

has the bias error of an order N.

Analogously to the TT estimator case, the errors and CC trade off [22] yields an optimal choice of the number of simulated paths $M \sim n^{2N}$. Hence, the HOWE estimate with RMSE of ε has CC of $O\left(\varepsilon^{-(2+1/N)}\right)$. As a result if expansion (43) holds for any positive integer N, it is theoretically possible to construct a HOWE estimate with CC that has an order arbitrary close to 2.

DRAWBACKS OF THE CLASSIC HOWE

However the HOWE approach in the above form has essential drawbacks that prevent it from being used in practice. Here we present and discuss at least some of them.

N	2	3	4	5
$\sum \left(\lambda_i^N\right)^2$	5	≈ 37	≈ 312	≈ 2916

Table 4: Values of a squared sum of the extrapolation weights

The first problem is a rapid increase of CC of simulating a replication of $\sum_{i=1}^{N} \lambda_i^N \hat{P}^{in}$ with increasing order N. To illustrate this let us assume that CC of $\hat{P}^m$ is linear in m, i.e.

$$\text{C.C.}\left[\hat{P}^m\right] \asymp m\,.$$

This leads to

$$\text{C.C.}\left[\sum_{i=1}^{N} \lambda_i^N \hat{P}^{in}\right] \asymp n\frac{N\,(N+1)}{2}\,.$$

In particular, already for a quadratic extrapolation, i.e. $N = 3$, CC is 6-times larger compared to the plain MC simulations.

The second reason and probably the most crucial one is an "explosion" of the scheme's variance $\mathbb{V}\text{ar}\left(\sum_{i=1}^{N} \lambda_i^N \hat{P}^{in}\right)$. In the case of an independent implementation of $\hat{P}^{in}$ and under some regularity requirements this variance grows asymptotically proportionally to a squared sum of the extrapolation weights

$$\mathbb{V}\text{ar}\left(\sum_{i=1}^{N} \lambda_i^N \hat{P}^{in}\right) \to \mathbb{V}\text{ar}\,(P) \sum_{i=1}^{N} \left(\lambda_i^N\right)^2 .$$

Then due to a rapid growth of the weights

$$\sum_{i=1}^{N} \left(\lambda_i^N\right)^2 \geqslant \left(\lambda_N^N\right)^2 = \left(\frac{N^N}{N!}\right)^2 \to \infty\,,$$

the variance of the scheme tends to infinity with increasing N. Even for a relatively small N the squared sum takes quite tremendous values, see Table 4. As a consequence the number of simulations M needed to keep up with a pre-fixed statistical error

$$\mathbb{V}\text{ar}\left(\bar{E}\right) = \frac{\mathbb{V}\text{ar}\left(\sum_{i=1}^{N} \lambda_i^N \hat{P}^{in}\right)}{M} \leqslant \text{stat. error}$$

will proportionally "explode" leading to an enormous growth of CC of the estimate.

In [59] Pagès partially eliminates this problems using consistent Brownian increments for schemes $\hat{P}^{j\cdot n}$. However our forthcoming approach circumvents this problem in a more efficient way.

3.1.3 *Literature Survey*

We recall the seminal result of Talay and Tubaro on weak convergence of the Euler scheme.

Proposition 2. *Let the coefficient functions* a *and* σ *of the SDE*

$$dS(t) = a(t, S(t))\,dt + \sigma(t, S(t))\,dW(t)\ , \quad t \in (0, T]$$

driven by a Brownian motion W *and the function* g *be in* C^∞*. Further, let* g *satisfy the following growth condition:*

$$\exists s, c > 0 \quad \text{s.t. } \forall x \in \mathbb{R}^d \ \ |g(x)| \leqslant c(1 + |x|^s)\ .$$

Then, with $\hat{S}^n$ *being the discretised version of* S *obtained by the Euler scheme, there exists a constant* C *independent of* n*, such that we have*

$$\mathbb{E}\left(g\left(\hat{S}^n(T)\right)\right) - \mathbb{E}\left(g\left(S(T)\right)\right) = C\,\frac{1}{n} + O\left(\frac{1}{n^2}\right)\ .$$

The requirements on the functions g, a and σ can be substantially relaxed, e.g. to g being a bounded measurable function and a and σ satisfying a condition of an elliptic type [7].

In [6] Bally and Talay established convergence of a density of the Euler scheme at the terminal time toward $S(T)$. Kurtz and Protter studied a weak convergence in series of papers [46], [48] and [47]) under different assumptions on the SDE coefficients.

Sharp rates of a weak convergence were also established in case of some path dependent options and the Euler scheme, e.g. Asian options [50] or barrier options [30]. However the assumptions therein are typically not satisfied, for instance in [30] the barrier domain $\mathcal{D}$ is assumed to have a smooth enough boundary, what is often not the case for barrier options. Though these results tell what behaviour we might expect.

3.2 WEAK EXTRAPOLATION MLMC

In this section we adopt and apply the HOWE into the MLMC setting. This combination not only eliminates the aforementioned drawbacks, but also reduces CC of the MLMC estimate. It turns out that the extrapolation weights are bounded for schemes with $n, n^2, n^3, \ldots$ discretisation points. Further, under appropriate assumptions, the variance of the estimate is controlled due to the level's variance decay.

3.2.1 *HOWE in MLMC settings*

Let us consider an approximation scheme for $N+1$ different numbers of discretisation points

$$\hat{P}^{n_0}, \hat{P}^{n_1}, \hat{P}^{n_2}, \ldots, \hat{P}^{n_N}\ ,$$

where $n_i = nM^i$ for some positive integers n and $M > 1$. In contrast to the considered HOWE the numbers of discretisation points n_i follow a geometric progression, as they do the MLMC settings. Based on these approximations we construct a $(N+1)^{\text{th}}$ order weak extrapolation scheme

$$\sum_{i=0}^{N} \lambda_i^N \, \hat{P}^{n_i} \,, \tag{44}$$

where λ_i^N are the real extrapolation weights independent of n.

FINDING EXTRAPOLATION WEIGHTS

Following the idea of Remark 19 we by eliminate the first N terms in the bias error. This will lead to a system of equation for λ_i^N. The idea is similar to the one proposed in [59], but the solution of the system is slightly more sophisticated.

To be more general we assume that the bias has a power series expansion in terms of n^α, i.e.

$$\mathbb{E}(P) - \mathbb{E}\left(\hat{P}^n\right) = \sum_{j=1}^{N+1} c_j \frac{1}{n^{j\alpha}} + o\left(\frac{1}{n^{(N+1)\alpha}}\right) \tag{45}$$

for some positive integer N, real constants c_i and rate $\alpha > 0$, that are all independent of the number of discretisation points n. It directly follows that for $i = 0, 1, \ldots, N$

$$\begin{aligned}\mathbb{E}(P) = \mathbb{E}\left(\hat{P}^{n_i}\right) &+ \sum_{j=1}^{N} \left(c_j \frac{1}{n^{j\alpha} M^{ij\alpha}}\right) \\ &+ c_{N+1} \frac{1}{(nM^i)^{(N+1)\alpha}} + o\left(\frac{1}{(nM^i)^{(N+1)\alpha}}\right) .\end{aligned}$$

This can be equivalently represented in the following matrix form

$$\begin{bmatrix} \mathbb{E}(P) \\ \vdots \\ \mathbb{E}(P) \\ \vdots \\ \mathbb{E}(P) \end{bmatrix} = \begin{bmatrix} \mathbb{E}\left(\hat{P}^{n_0}\right) \\ \vdots \\ \mathbb{E}\left(\hat{P}^{n_i}\right) \\ \vdots \\ \mathbb{E}\left(\hat{P}^{n_N}\right) \end{bmatrix} + A \cdot \begin{bmatrix} c_1 \frac{1}{n^{\alpha}} \\ \vdots \\ c_i \frac{1}{n^{j\alpha}} \\ \vdots \\ c_N \frac{1}{n^{N\alpha}} \end{bmatrix}$$

$$+ \frac{c_{N+1}}{n^{(N+1)\alpha}} \cdot \begin{bmatrix} 1 + o(1) \\ \vdots \\ \frac{1}{M^{i(N+1)\alpha}} + o\left(\frac{1}{M^{i(N+1)\alpha}}\right) \\ \vdots \\ \frac{1}{M^{N(N+1)\alpha}} + o\left(\frac{1}{M^{N(N+1)\alpha}}\right) \end{bmatrix}, \tag{46}$$

where A is $(N+1)$-by-N matrix defined as

$$A = \begin{bmatrix} 1 & \dots & 1 & \dots & 1 \\ \frac{1}{M^{1\alpha}} & \dots & \frac{1}{M^{1j\alpha}} & \dots & \frac{1}{M^{1N\alpha}} \\ \vdots & \dots & \vdots & \dots & \vdots \\ \frac{1}{M^{i\alpha}} & \dots & \frac{1}{M^{ij\alpha}} & \dots & \frac{1}{M^{iN\alpha}} \\ \vdots & \dots & \vdots & \dots & \vdots \\ \frac{1}{M^{N\alpha}} & \dots & \frac{1}{M^{Nj\alpha}} & \dots & \frac{1}{M^{NN\alpha}} \end{bmatrix} = \left[\frac{1}{M^{ij\alpha}}\right]_{0 \leqslant i \leqslant N,\ 1 \leqslant j \leqslant N}.$$

Let us define a co-vector of the weights $\lambda^N \in \left(\mathbb{R}^{N+1}\right)^*$

$\lambda^N = \left(\lambda_0^N, \lambda_1^N, \dots, \lambda_N^N\right)$

Then we aim to find weights, such that by left multiplication of Equation 46 with λ^N

1. all leading $n^{-j\alpha}$ terms for $j \leqslant N$ are eliminated from the RHS of the equation, i.e.

$$\lambda^N \cdot A = 0\,; \tag{47}$$

2. and the left hand side (LHS) is equal to $\mathbb{E}[P]$, i.e.

$$\lambda^N \cdot \mathbf{1} = 1\,, \tag{48}$$

where $\mathbf{1} = (1, 1, \dots, 1)^* \in \mathbb{R}^{N+1}$.

Identities (47) and (48) ensure that

$$\mathbb{E}(P) = \sum_{i=0}^{N} \lambda_i^N \mathbb{E}\left(\hat{P}^{n_i}\right) + \frac{c_{N+1}}{n^{(N+1)\alpha}} \sum_{i=0}^{N} \lambda_i^N \left(\frac{1}{M^{iN\alpha}} + o\left(\frac{1}{M^{iN\alpha}}\right)\right).$$

Moreover they can be combined into a system of linear equations

$$\lambda^N \cdot \tilde{A} = e_1^* \,, \tag{49}$$

where $e_1^* = (1, 0, \ldots, 0) \in \left(\mathbb{R}^{N+1}\right)^*$, and $\tilde{A}$ is a matrix concatenation

$$\tilde{A} = [\mathbf{1} \;\; A] = \left[\frac{1}{M^{ij\alpha}}\right]_{0 \leqslant i \leqslant N,\; 0 \leqslant j \leqslant N} .$$

The system of linear equations (49) admits a closed form solution

$$\lambda^N_{N-j} = (-1)^j \left(\prod_{i=1}^{j} \frac{1}{M^{i\alpha}-1}\right) \left(\prod_{i=1}^{N-j} \frac{M^{i\alpha}}{M^{i\alpha}-1}\right) . \tag{50}$$

The detailed derivation of the solution can be found in appendix Section B.1.

BIAS OF HOWE

We summarise our result on the residual bias error of the newly proposed HOWE scheme in the following proposition.

Proposition 3. *Let* $\hat{P}^m$ *be an approximation scheme, such that its bias error admits power series expansion* (45)*.*
Then for $n_i = nM^i$ *and* λ_i^N *as in Equation 50 we have*

$$\left|\mathbb{E}\,(P) - \mathbb{E}\left(\sum_{i=0}^{N} \lambda_i^N \; \hat{P}^{n_i}\right)\right| = c_{N+1} \frac{1}{n^{(N+1)\alpha} M^{\frac{(N+1)N}{2}\alpha}} + o\left(\frac{1}{n^{(N+1)\alpha} M^{\frac{(N+1)N}{2}\alpha}}\right) ,$$

for $n \to \infty$*.*

The proof of the proposition follows from the construction of the weights and its details can be found in appendix Section B.2. Note the residual bias term does not "explode" and even turns out to be smaller than expected $n^{-(N+1)\alpha}$.

3.2.2 *Weak Extrapolation Multilevel Monte Carlo*

To combine the HOWE and the MLMC we will rewrite the weighted sum (44) as a sum of levels with appropriate weights. Let us define

$$\kappa_i^N = \sum_{j=i}^{N} \lambda_j^N \,, \tag{51}$$

in particular, $\kappa_0^N = 1$ and $\kappa_N^N = \lambda_N^N$. Then the HOWE sum (44) can be represented as

$$\sum_{i=0}^{N} \lambda_i^N \hat{P}^{n_i} = \hat{P}^{n_0} + \sum_{i=1}^{N} \kappa_i^N \left(\hat{P}^{n_i} - \hat{P}^{n_{i-1}}\right),$$

what gives an alternative formulation of the HOWE estimate

$$\tilde{E} = \frac{1}{N_0} \sum_{j=1}^{N_0} \hat{P}_j^{n_0} + \sum_{i=1}^{N} \kappa_i^N \frac{1}{N_i} \sum_{j=1}^{N_i} \left(\hat{P}_j^{n_i} - \hat{P}_j^{n_{i-1}}\right). \tag{52}$$

This estimator is similar to the MLMC sum of levels, but enhanced with the weights κ_i^N. Analogously to the MLMC it is crucial to use consistent Brownian increments inside each level, and the levels should be simulated independently.

On the contrary to the straightforward HOWE (see Section 3.1.2) the weights κ_j^N do not "explode" under mild assumption. They are indeed bounded for all N and we formalise this in the following proposition.

Proposition 4. *Let κ_j^N be defined as in Equation 51 , where λ_j^N are given by Equation 50.*

Then, if $M^\alpha \geqslant 2$, we have that $\forall N \in \mathbb{N}$

$$0 < \kappa_i^N < \kappa_N^N \qquad \text{for } i = 0, 1, \dots, N-1\,.$$

Moreover κ_N^N is monotonously increasing in N and

$$\kappa_N^N \to C < \infty \qquad \text{as } N \to \infty\,.$$

The proof is rather technical and is given in appendix Section B.3. As the common ratio M is of our choice, we tacitly assume that for the rest of this chapter it is chosen such that $M^\alpha \geqslant 2$.

Remark 21. *As κ_N^N is the upper bound for all weights we would like to discus its asymptotic behaviour for N going to infinity.*

From the proof of Proposition 4 it follows that

$$\kappa_N^N \nearrow f(M^\alpha) < \infty \quad \text{as } N \to \infty\,,$$

where

$$f(x) = \prod_{i=1}^{\infty} \frac{x^i}{x^i - 1}\,.$$

Note that f is well-defined and is a particular case of the q-Pochhammer Symbol (see e.g. [41]). As a consequence the leading weight κ_N^N is bounded by

$$\frac{M^\alpha}{M^\alpha - 1} \leqslant \kappa_N^N \leqslant f(M^\alpha)\,,$$

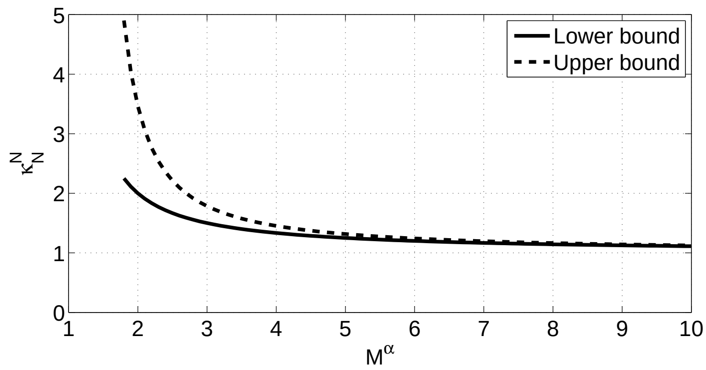


Figure 5: Upper and Lower Bounds for κ_N^N

N	κ_N^N	κ_{N-1}^N	κ_{N-2}^N	κ_{N-3}^N	κ_{N-4}^N	κ_{N-5}^N	κ_{N-6}^N
1	1.3333	1					
2	1.4222	0.9778	1				
3	1.4448	0.9707	1.0004	1			
4	1.4505	0.9689	1.0005	1.0000	1		
5	1.4519	0.9684	1.0005	1.0000	1.0000	1	
6	1.4522	0.9683	1.0005	1.0000	1.0000	1.0000	1

Table 5: Example of approximate values of κ^N for $M = 4$, $\alpha = 1$ and different values of N.

where the lower bound is equal to κ_1^1. For the sake of notational simplicity we denote the upper bound by κ_∞^∞, i.e. $\kappa_\infty^\infty = f(M^\alpha)$.

On Figure 5 we plot the upper and lower bounds of κ_N^N as a function of M^α. It can be seen that the leading weight κ_N^N takes moderate values. As the weights have an impact on the variance of an estimate, we illustrate an example of their possible values in Table 5 to give the reader feeling about their impact.

Further we propose to substitute the plain MC estimator of $\mathbb{E}\left(\hat{P}^{n_0}\right)$ in (52) by the MLMC. This is motivate motivated by the finding of Chapter 2 and in particular Proposition 1.

Without loss of generality let n be some power of M. Then by redefining $n_i = M^i$ for $i = 0, 1, \ldots, L$, we rewrite the HOWE estimate (52) as a combination of the MLMC estimator with HOWE based on the N finest discretisation schemes

$$\tilde{E} = \underbrace{\frac{1}{N_0}\sum_{j=1}^{N_0}\hat{P}^{n_0} + \sum_{i=1}^{L-N}\frac{1}{N_i}\sum_{j=1}^{N_i}\left(\hat{P}^{n_i} - \hat{P}^{n_{i-1}}\right)}_{\text{plain MLMC}}$$

$$+ \underbrace{\sum_{i=L-N+1}^{L}\kappa^{N}_{i-(L-N)}\frac{1}{N_i}\sum_{j=1}^{N_i}\left(\hat{P}^{n_i} - \hat{P}^{n_{i-1}}\right)}_{\text{HOWE based on } N \text{ finest schemes}} . \quad (53)$$

As the coarsest scheme used for the HOWE has $n = M^{L-N}$ discretisation points Proposition 3 implies that

$$\mathbb{E}(P) - \mathbb{E}\left(\tilde{E}\right) = O\left(\frac{1}{M^{(L-N)(N+1)\alpha}M^{\frac{(N+1)N}{2}\alpha}}\right) . \quad (54)$$

There still remains the question:"How many levels should be used for the HOWE within estimate (53) ?". From Table 5 we can notice that the values of κ^{N}_{N-i} do not differ much for different N. Therefore for a fixed L, M and N_i we reasonably assume that the choice of N has a minor influence on the variance of the estimate and hence on CC of the estimate for a fixed statistical error. Meanwhile Equation 54 implies that N has a great impact on an order of bias error convergence. This suggests to choose N such that it maximises the order of the bias convergence. Solving

$$\alpha\left((L-N)(N+1) + \frac{(N+1)N}{2}\right) \to \max$$

for positive integer N we obtain the optimal values of

$$N^* = L \quad \text{or} \quad N^* = L-1 .$$

For the sake of consistence in the later we assume that $N^* = L$, but the similar consideration can be replicated for $N^* = L-1$.

Thus we should build the HOWE based on the all levels. This leads to the *weak extrapolation multilevel MC estimate*

$$\hat{E} = \frac{1}{N_0}\sum_{j=1}^{N_0}\hat{P}^{n_0} + \sum_{i=1}^{L}\kappa^{L}_{i}\frac{1}{N_i}\sum_{j=1}^{N_i}\left(\hat{P}^{n_i} - \hat{P}^{n_{i-1}}\right) . \quad (55)$$

From Equation 54 it follows that the bias error of this estimate is equal to

$$\mathbb{E}(P) - \mathbb{E}\left(\tilde{E}\right) = O\left(\frac{1}{M^{\frac{(L+1)L}{2}\alpha}}\right) .$$

Note, it turns out that the bias error is controlled by increasing the order of weak extrapolation rather then increasing the number of discretisation points for the coarsest grid.

Remark 22. *Sometimes in applications the bias expansion assumption* (45) *might hold only starting from some* $n > 1$*. For instance this is the case for the Euler scheme and barrier options (see numerical example in Section 3.5.1.5). Then the HOWE estimator could be applied omitting some of the starting levels, e.g. the modified WEML estimator could look like*

$$\hat{E} = \frac{1}{N_0}\sum_{j=1}^{N_0}\hat{P}^{n_0} + \frac{1}{N_1}\sum_{j=1}^{N_1}\left(\hat{P}^{n_1} - \hat{P}^{n_0}\right) + \\ + \sum_{i=2}^{L}\left(\kappa_{i-1}^{L-1}\frac{1}{N_i}\sum_{j=1}^{N_i}\left(\hat{P}^{n_i} - \hat{P}^{n_{i-1}}\right)\right).$$

3.3 NUMERICAL ALGORITHM

To give a final algorithm we have to specify the numbers of paths N_i for each level, such that the WEML estimate (55) has the targeted statistical error ε^2. Using the optimality results of Section 2.1.1, in particular Equation 24, we obtain

$$N_i = \left\lceil \frac{\sum_{j=0}^{L}\sqrt{\left(\kappa_j^L\right)^2 V_j n_j}}{\varepsilon^2}\sqrt{\frac{\left(\kappa_i^L\right)^2 V_i}{n_i}} \right\rceil, \tag{56}$$

where V_l defined as

$$V_i = \begin{cases} \mathbb{V}\mathrm{ar}\left(\hat{P}^{n_0}\right) & \text{for} \quad i = 0\,, \\ \mathbb{V}\mathrm{ar}\left(\hat{P}^{n_i} - \hat{P}^{n_{i-1}}\right) & \text{for} \quad i = 1, 2, \ldots, L\,. \end{cases}$$

This choice ensures that the WEML estimate has a statistical error less or equal to the targeted one, while CC of the estimate is (modulo rounding) as small as possible.

We summarise the proposed approach in Algorithm 2. Note, in contrast to the classic MLMC there exists no efficient stopping heuristic. As a consequence the number of levels L should be given by some oracle. In particular it should be chosen such that the bias error matches the statistical one.

Algorithm 2 Weak Extrapolation Multilevel Monte Carlo

Input: ε and L
Output: $\hat{E}$
1: estimate $V_0, \ldots, V_L$ using an initial $N_l = 10^4$ samples
2: define optimal $N_0, \ldots, N_L$ using Equation 56
3: evaluate extra samples at each level as needed for new N_i
4: calculate $\hat{E}$ according to Equation 55

3.4 COMPUTATIONAL COST THEOREM

In this section we formulate and prove a CC theorem for the WEML estimate (55). This will be done in the spirit of the Complexity Theorem 7 for the MLMC estimate.

Theorem 8 (Computational Cost Theorem)**.** *Let* P *be a functional of the solution of some SDE driven by Brownian motion. Further let* $\hat{P}^m$ *be a numerical approximation based on* m *equidistant discretisation points, such that for some positive constants* β, γ, c_c *and* c_v *the following holds* $\forall m \in \mathbb{N}$

i) $\mathrm{C.C.}\left[\hat{P}^m\right] \leqslant c_c m$,

ii) $\mathbb{V}\mathrm{ar}\left(P - \hat{P}^m\right) \leqslant c_v m^{-\beta}$,

iii) $\mathbb{V}\mathrm{ar}\left(\hat{P}^m\right) \leqslant c_v$.

Moreover let the bias error of the approximation $\hat{P}^n$ *admit a power series expansion to an arbitrary order with bounded coefficients. I.e. there exist real coefficients* c_i *such that for* $\forall N, n \in \mathbb{N}$

$$\mathbb{E}(P) - \mathbb{E}\left(\hat{P}^n\right) = \sum_{i=1}^{N} c_i \frac{1}{n^{i\alpha}} + O\left(\frac{1}{n^{(N+1)\alpha}}\right) ,$$

and $\exists c > 0$ *such that for* $\forall i \in \mathbb{N}$, $|c_i| < c$.

Then there exists a choice of N_i *and* L *for which the WEML estimate*

$$\hat{E} = \frac{1}{N_0} \sum_{j=1}^{N_0} \hat{P}_j^{n_0} + \sum_{i=1}^{L} \kappa_i^L \frac{1}{N_i} \sum_{j=1}^{N_i} \left(\hat{P}_j^{n_i} - \hat{P}_j^{n_{i-1}}\right) ,$$

where $n_i = M^i$ *for* M *such that* $M^\alpha \geqslant 2$*, has a mean square error with bound*

$$\mathbb{MSE}\left[\hat{E}\right] = \mathbb{E}\left(\hat{E} - \mathbb{E}(P)\right)^2 \leqslant \varepsilon^2 ,$$

and for a sufficiently small ε *computational cost of the estimate is bounded by*

$$\mathrm{C.C.}\left[\hat{E}\right] \leqslant \begin{cases} \quad C\varepsilon^{-2} & \textit{if} \quad \beta > 1 , \\ \quad C\varepsilon^{-2}\ln\left(\varepsilon^{-1}\right) & \textit{if} \quad \beta = 1 , \\ \forall \delta > 0 \quad C\varepsilon^{-2-\delta} & \textit{if} \quad \beta < 1 , \end{cases}$$

where the constant C *is independent of* ε*.*

Note, the requirement of ε to be sufficiently small is rather technical and is usually satisfied in applications. It is imposed to ensure that at each level at least one replication should be simulated to meet the targeted statistical error.

	MLMC	WEML
C.C. $[\hat{E}]$	*identical*	*identical*
$\mathbb{V}\mathrm{ar}(\hat{E})$	ε^2	$\leqslant (\kappa_\infty^\infty)^2 \varepsilon^2$
bias $(\hat{E})$	$\asymp M^{-\alpha L}$	$\asymp M^{-\alpha \frac{L(L+1)}{2}}$

Table 6: Comparison of WEML and MLMC estimates

Before proceeding to the proof we would like to compare this result to the Complexity Theorem 7 for the MLMC. The most essential improvement corresponds to the case of $\beta < 1$. In particular the CC rate can be as close as possible to 2, what is the lower bound for such problems, see [58]. Moreover it is independent of α, what is not the case for the MLMC. This could be explained as application of a scheme with the "infinite" order of weak convergence in the setting of the Complexity Theorem 7. In the case of $\beta = 1$ CC of the WEML drops in log-terms compared to the MLMC. In the renaming case of $\beta > 1$ CC of both estimates are asymptotically equal. However we still might be better off in terms of a constant, for numerical comparisons see Section 3.5.

To explain where the benefit comes from, let us have a brief look at the bias and statistical errors for WEML and MLMC estimates. We use the same number of levels L and the numbers of paths N_i for both WEML and MLMC estimates, what ensures that both estimates have identical CC. Moreover let N_i be chosen such that the MLMC estimate has statistical error of ε^2. Then while the variances of both estimates differ by a constant independent for targeted accuracy, the bias error converges essentially faster for the WEML estimate, see Table 6. As a consequence we need less levels to obtain the required bias error. The improved bias convergence explains the reduction of CC of the WEML estimate proven in Theorem 8.

Nevertheless the WEML estimate has clear advantages in terms of CC, its application is still limited due to some drawbacks. First of all the WEML approach strongly relies on the bias expansion assumption, that is not always satisfied. For example in the case of the Euler scheme it is satisfied for call options under the BS model and breached for call options under the Heston model (see Section 3.5). Furthermore the WEML lacks an effective stopping criterion, while MLMC has one. As a consequence the number of levels should be chosen using some meta information. However this can be done on the fly and should not be specified in advance.

Proof of the Computational Cost Theorem. According to the MSE decomposition (see Section 1.2.2.2)

$$\mathbb{MSE}\left[\hat{E}\right] = \mathbb{V}\mathrm{ar}\left(\hat{E}\right) + \left(\mathbb{E}\left(\hat{E}\right) - \mathbb{E}\left(P\right)\right)^2$$

it is sufficient to show that both squared bias error and statistical error are less or equal to half of ε^2, i.e.

$$\left|\mathbb{E}\left(\hat{E}\right) - \mathbb{E}\left(P\right)\right| \leqslant \frac{1}{\sqrt{2}}\,\varepsilon \qquad \text{and} \qquad \mathbb{V}\mathrm{ar}\left(\hat{E}\right) \leqslant \frac{1}{2}\,\varepsilon^2\,,$$

to prove that

$$\mathbb{MSE}\left[\hat{E}\right] \leqslant \varepsilon^2\,.$$

In the following the bias error will be bounded by an appropriate choice of the number of levels L, while the variance will be controlled by a choice of the numbers of simulated paths N_i.

We first explain how to chose L, such that the squared bias error is less or equal to half of ε^2. Proposition 3 compounded by the boundedness of the bias expansion coefficients implies that

$$\left|\mathbb{E}\left(\hat{E}\right) - \mathbb{E}\left(P\right)\right| \leqslant c_b M^{-\alpha \frac{L^2}{2}}$$

for some consonant $c_b >$ independent form ε. Then, by choosing L

$$L = \left\lceil \sqrt{\frac{2\ln\left(\sqrt{2}c_b\varepsilon^{-1}\right)}{\alpha\ln\left(M\right)}} \right\rceil$$

we guarantee that

$$\left|\mathbb{E}\left(\hat{E}\right) - \mathbb{E}\left(P\right)\right| \leqslant \frac{\varepsilon}{\sqrt{2}}\,.$$

In the later we will use the following elementary bounds of L

$$\sqrt{\frac{2\ln\left(\sqrt{2}c_b\varepsilon^{-1}\right)}{\alpha\ln\left(M\right)}} \leqslant L < \sqrt{\frac{2\ln\left(\sqrt{2}c_b\varepsilon^{-1}\right)}{\alpha\ln\left(M\right)}} + 1\,, \tag{57}$$

and the proven below Corollary 1.

Corollary 1. $\forall\delta > 0\,, \quad \exists 0 < \varepsilon^* < 1$ *such that* $\forall\varepsilon < \varepsilon^*$ *holds*

$$M^{\sqrt{\frac{2\ln\left(\sqrt{2}c_b\varepsilon^{-1}\right)}{\alpha\ln\left(M\right)}}} \leqslant \varepsilon^{-\delta}\,.$$

Proof of corollary. Let us consider the quadratic function q

$$q\left(x\right) := \delta^2 x^2 - \frac{2\ln\left(M\right)}{\alpha}x - \frac{2\ln\left(M\right)\ln\left(\sqrt{2}c_b\right)}{\alpha}$$

for positive x. As the main coefficient of the quadratic function q is positive there exists x^* such that

$$\forall x > x^*, \quad q\left(x\right) \geqslant 0\,.$$

Then substituting x by $\ln\left(\varepsilon^{-1}\right)$ we obtain that

$$\forall \varepsilon < e^{-x^*}, \quad \delta^2 \left(\ln\left(\varepsilon^{-1}\right)\right)^2 \geqslant \left(\ln\left(M\right)\right)^2 \frac{2\ln\left(\sqrt{2}c_b \varepsilon^{-1}\right)}{\alpha \ln\left(M\right)}\,.$$

For $\varepsilon < \varepsilon^* := \min\left\{e^{-x^*}, 1\right\}$ by taking the square root and then the exponent from both sides of the inequality we prove the assertion

$$\forall \varepsilon < \varepsilon^*, \quad \varepsilon^{-\delta} \geqslant M^{\sqrt{\frac{2\ln\left(\sqrt{2}c_b \varepsilon^{-1}\right)}{\alpha \ln(M)}}}\,.$$

□

Combining inequality (57) and Corollary 1 we derive that for any $\delta > 0$ there exists an $\varepsilon_1^* > 0$ such that

$$\forall \varepsilon < \varepsilon_1^*, \quad M^L < M^{\sqrt{\frac{2\ln\left(\sqrt{2}c_b \varepsilon^{-1}\right)}{\alpha \ln(M)}}+1} \leqslant M\varepsilon^{-\delta}\,. \tag{58}$$

Furthermore, for $\delta = 2$ this implies that

$$\forall \varepsilon < \varepsilon_2^*, \quad \sum_{i=0}^{L} M^i = \frac{M^{L+1}-1}{M-1} \leqslant \frac{M^2}{M-1}\,\varepsilon^{-2}\,. \tag{59}$$

The variance of the estimate can be bounded as

$$\begin{aligned}\mathbb{V}\mathrm{ar}\left(\hat{E}\right) &= \frac{1}{N_0}\mathbb{V}\mathrm{ar}\left(\hat{P}^{n_0}\right) + \sum_{i=1}^{L}\left(\kappa_i^L\right)^2 \frac{1}{N_i}\mathbb{V}\mathrm{ar}\left(\hat{P}^{n_i} - \hat{P}^{n_{i-1}}\right)\\ &\leqslant \left(\kappa_\infty^\infty\right)^2 \left(\frac{1}{N_0}\mathbb{V}\mathrm{ar}\left(\hat{P}^{n_0}\right) + \sum_{i=1}^{L}\frac{1}{N_i}\mathbb{V}\mathrm{ar}\left(\hat{P}^{n_i} - \hat{P}^{n_{i-1}}\right)\right)\,,\end{aligned}$$

where the last inequality is an implication of Proposition 4 and Remark 21. Also

$$\begin{aligned}\mathbb{V}\mathrm{ar}\left(\hat{P}^{n_i} - \hat{P}^{n_{i-1}}\right) =& \mathbb{V}\mathrm{ar}\left(P - \hat{P}^{n_i}\right) + \mathbb{V}\mathrm{ar}\left(P - \hat{P}^{n_{i-1}}\right) + \\ &+ 2\mathrm{Cov}\left(\hat{P}^{n_i} - P, P - \hat{P}^{n_{i-1}}\right)\\ \leqslant& \left(\sqrt{\mathbb{V}\mathrm{ar}\left(P - \hat{P}^{n_i}\right)} + \sqrt{\mathbb{V}\mathrm{ar}\left(P - \hat{P}^{n_{i-1}}\right)}\right)^2\,.\end{aligned}$$

Then combined with the assumption of the theorem this yields

$$V_i = \mathbb{V}\mathrm{ar}\left(P^{n_i} - P^{n_{i-1}}\right) \leqslant c_v \left(1 + M^{\beta/2}\right)^2 M^{-\beta i}\,.$$

Hence

$$\mathbb{V}\mathrm{ar}\left(\hat{E}\right) \leqslant C_v \sum_{i=0}^{L} N_i^{-1} M^{-\beta i} ,$$

where the constant

$$C_v = (\kappa_\infty^\infty)^2 c_v \left(1 + M^{\beta/2}\right)^2$$

is independent of ε, N_i and L. At the same time CC of the estimate can be bounded as

$$\mathrm{C.C.}\left[\hat{E}\right] \leqslant c_c \left(N_0 n_0 + \sum_{i=0}^{L} N_i \left(n_i + n_{i-1}\right)\right) \leqslant 2c_c \sum_{i=0}^{L} N_i\, M^i .$$

In the following we will specify numbers of simulations N_i on each level, such that the statistical error of the estimate is less or equal to half of ε^2. For this choice of N_i we will derive an upper bound for CC of the estimate. To proceed further we have to distinguish between three cases depending on the speed of the strong convergence:

I) $\beta = 1$;

II) $\beta > 1$;

III) $\beta < 1$.

I) CASE $\beta = 1$

We set $N_i = \left\lceil 2C_v \varepsilon^{-2} \left(L+1\right) M^{-i} \right\rceil$, what directly implies that

$$\mathbb{V}\mathrm{ar}\left(\hat{E}\right) \leqslant \frac{1}{2}\varepsilon^2 .$$

Hence the MSE of the estimate is less or equal to ε^2. As

$$N_i \leqslant 2C_v \varepsilon^{-2} \left(L+1\right) M^{-i} + 1 ,$$

for $\varepsilon < \varepsilon_2^*$ Equation 59 gives an upper bound for CC

$$\begin{aligned}\mathrm{C.C.}\left[\hat{E}\right] &\leqslant 4C_v c_c \varepsilon^{-2} \sum_{i=0}^{L} \left(L+1\right) + 2c_c \sum_{i=0}^{L} M^i \\ &\leqslant 4C_v c_c \varepsilon^{-2} \left(L+1\right)^2 + 2c_c \varepsilon^{-2} M^2 \left(M-1\right)^{-1} .\end{aligned} \tag{60}$$

Using inequality (57) we derive that

$$\begin{aligned}\left(L+1\right)^2 &< \left(\sqrt{\frac{2\ln\left(\sqrt{2}c_b \varepsilon^{-1}\right)}{\alpha \ln\left(M\right)}} + 2\right)^2 \leqslant 2\left(\frac{2\ln\left(\sqrt{2}c_b \varepsilon^{-1}\right)}{\alpha \ln\left(M\right)} + 4\right) \\ &= \frac{4\ln\left(\varepsilon^{-1}\right)}{\alpha \ln\left(M\right)} + \frac{4\ln\left(\sqrt{2}c\right)}{\alpha \ln\left(M\right)} + 8 .\end{aligned} \tag{61}$$

Then if we define

$$C := \max \left\{ \frac{16 C_v c_c}{\alpha \ln(M)}, \; 4 C_v c_c \left(\frac{4 \ln\left(\sqrt{2}\, c\right)}{\alpha \ln(M)} + 8 \right) + \frac{2 c_c M^2}{M-1} \right\} ,$$

Equation 60 and Equation 61 yield the assertion, i.e. let $\varepsilon^* = \min\left\{\varepsilon_2^*, e^{-1}\right\}$ then

$$\forall \varepsilon < \varepsilon^* , \quad \mathrm{C.C.}\left[\hat{E}\right] \leqslant C \varepsilon^{-2} \ln\left(\varepsilon^{-1}\right) .$$

II) CASE $\beta > 1$

The choice of

$$N_l = \left\lceil 2 C_v \varepsilon^{-2} \left(1 - M^{-\frac{\beta-1}{2}}\right)^{-1} M^{-i\frac{\beta+1}{2}} \right\rceil$$

together with a formula for an infinite sum of a decaying geometric progression, imply that

$$\begin{aligned} \mathbb{V}\mathrm{ar}\left(\hat{E}\right) &\leqslant \frac{1}{2}\varepsilon^2 \left(1 - M^{-\frac{\beta-1}{2}}\right) \sum_{i=0}^{L} M^{-i\frac{\beta-1}{2}} \\ &\leqslant \frac{1}{2}\varepsilon^2 \left(1 - M^{-\frac{\beta-1}{2}}\right) \left(1 - M^{-\frac{\beta-1}{2}}\right)^{-1} = \frac{1}{2}\varepsilon^2 . \end{aligned}$$

Therefore the MSE of the estimate is less or equal than ε^2. On the other hand as

$$N_i \leqslant 2 C_v \varepsilon^{-2} \left(1 - M^{-\frac{\beta-1}{2}}\right)^{-1} M^{-i\frac{\beta+1}{2}} + 1 ,$$

the CC can be bounded as

$$\mathrm{C.C.}\left[\hat{E}\right] \leqslant \frac{2 C_v c_c \varepsilon^{-2}}{1 - M^{-\frac{\beta-1}{2}}} \sum_{i=0}^{L} M^{-i\frac{\beta-1}{2}} + 2 c_c \sum_{i=0}^{L} M^i . \tag{62}$$

The formula for an infinite sum of decaying geometric progression implies that

$$\sum_{i=0}^{L} M^{-i\frac{\beta-1}{2}} \leqslant \left(1 - M^{-\frac{\beta-1}{2}}\right)^{-1} ,$$

while according to Equation 59

$$\exists \varepsilon^*, \quad \text{s.t. } \forall \varepsilon < \varepsilon^* \quad \sum_{i=0}^{L} M^i \leqslant \frac{M^2}{M-1} \varepsilon^{-2} .$$

Then putting these inequalities into Equation 62 we derive the assertion, i.e.

$$\forall \varepsilon < \varepsilon^*, \quad \mathrm{C.C.}\left[\hat{E}\right] \leqslant C\, \varepsilon^{-2} \ln\left(\varepsilon^{-1}\right) ,$$

for the constant C

$$C = \max \left\{ \frac{2 C_v c_c}{\left(1 - M^{-\frac{\beta-1}{2}}\right)^2}, \frac{2 c_c M^2}{M-1} \right\} .$$

III) CASE $\beta < 1$

Choose N_i to be

$$N_l = \left\lceil 2C_v \varepsilon^{-2} M^{L\frac{1-\beta}{2}} \left(1 - M^{-\frac{1-\beta}{2}}\right)^{-1} M^{-i\frac{\beta+1}{2}} \right\rceil .$$

Then the variance of the estimate can be bounded by

$$\mathbb{V}\mathrm{ar}\left(\hat{E}\right) \leqslant \frac{1}{2}\varepsilon^2 M^{-L\frac{1-\beta}{2}} \left(1 - M^{-\frac{1-\beta}{2}}\right) \sum_{i=0}^{L} M^{i\frac{1-\beta}{2}} .$$

Using the formula for an infinite sum of a geometric progression it can be shown that

$$\sum_{i=0}^{L} M^{i\frac{1-\beta}{2}} = M^{L\frac{1-\beta}{2}} \sum_{i=0}^{L} M^{-i\frac{1-\beta}{2}} \leqslant M^{L\frac{1-\beta}{2}} \left(1 - M^{-\frac{1-\beta}{2}}\right)^{-1} . \tag{63}$$

Hence

$$\mathbb{V}\mathrm{ar}\left(\hat{E}\right) \leqslant \frac{1}{2}\varepsilon^2 ,$$

what implies that the MSE of the estimate is less or equal to ε^2. Meanwhile the CC of the estimate in such a case can be bounded as

$$\begin{aligned} \mathrm{C.C.}\left[\hat{E}\right] &\leqslant 2C_v c_c \varepsilon^{-2} \frac{M^{L\frac{1-\beta}{2}}}{1 - M^{-\frac{1-\beta}{2}}} \sum_{i=0}^{L} M^{i\frac{1-\beta}{2}} + 2c_c \sum_{i=0}^{L} M^i \\ &\leqslant 2C_v c_c \varepsilon^{-2} \frac{M^{L(1-\beta)}}{\left(1 - M^{-\frac{1-\beta}{2}}\right)^2} + 2c_c \sum_{i=0}^{L} M^i , \end{aligned}$$

where the later step is justified by inequality (63). Making use of (58) and (59) we obtain that for any $\delta > 0$ there exists an $\varepsilon^* = \min\left\{\varepsilon_1^*, \varepsilon_2^*\right\} > 0$ such that $\forall \varepsilon < \varepsilon^*$

$$M^{L(1-\beta)} \leqslant M^{1-\beta}\varepsilon^{-\delta} \qquad \text{and} \qquad \sum_{i=0}^{L} M^i \leqslant \frac{M^2}{M-1}\varepsilon^{-2} .$$

Hence

$$\forall \delta > 0 \; \exists \varepsilon^* > 0 \quad \text{s.t.} \; \forall \varepsilon < \varepsilon^*, \quad \mathrm{C.C.}\left[\hat{E}\right] \leqslant C\varepsilon^{-(2+\delta)} ,$$

where the constant C is equal to

$$C := \max\left\{ \frac{2C_v c_c M^{1-\beta}}{\left(1 - M^{-\frac{\beta-1}{2}}\right)^2}, \frac{2c_c M^2}{M-1} \right\} .$$

□

3.5 NUMERICAL PERFORMANCE ANALYSIS

In this section we discuss the numerical performance of the WEML estimate for different models and option types. In all case we first numerically test bias expansion assumption (45) and then compare performance of the WEML algorithms to the MLMC algorithm.

In this analysis we consider two models

- the BS model (see Section 1.1.2.1);
- the Heston model (see Section 1.1.2.2);

and focus on the following European options (see Section 1.1.1 for definitions):

- call options;
- digital options (see e.g. [45] for a definition);
- one-sided barrier options;
- Asian options.

We will see below that a full analysis will only be given in the cases of call, digital and Asian options under the BS model. Note that these cases cover the whole range of different rates of strong convergences. In particular, Asian options exhibit fast strong convergence with $\beta = 2$, call options − moderate with $\beta = 1$ and digital options − slow with $\beta = 1/2$. In the rest cases assumption (45) is either not fulfilled (e.g. the Heston model with the Euler full truncation scheme [53]) or it is not clear if it is fulfilled or not (e.g. barrier options in the BS model).

POWER SERIES BIAS EXPANSION

To verify correctness of a power series bias expansion (45) we study bias convergence of the HOWE schemes (44) of different orders. Let us briefly explain how it will be done and introduce some notations.

We begin with the plain approximation scheme for n discretisation points, which we denote by

$$E_n^0 = \hat{P}^n .$$

An order of the bias convergence of this scheme should indicate the first term in the expansion (45). To be more precise, if assumption (45) holds, the bias error

$$\text{bias}_n^0 = \mathbb{E}\left(E_n^0\right) - \mathbb{E}\left(P\right) = \mathbb{E}\left(\hat{P}^n\right) - \mathbb{E}\left(P\right)$$

should behave as

$$\text{bias}_n^0 \asymp n^{-\alpha} \, .$$

In particular, such a behaviour would imply that

$$\frac{\text{bias}_n^0}{\text{bias}_{2n}^0} \approx 2^{\alpha} \, .$$

Thus, we are checking if the bias of the scheme exhibits this converging behaviour.

As the values of bias_n^0 cannot be obtained in an analytical form, they have to be approximated numerically. This will be done via the MC methods, i.e.

$$\overline{\text{bias}}_n^0 = \bar{E}_n^0 - \mathbb{E}\left(P\right) \, ,$$

where $\bar{E}_n^0$ is an MC estimate of E_n^0 and $\mathbb{E}\left(P\right)$ is assumed to be know either due to a closed form formula (e.g. European call option in the BS model) or via a good quality approximation (e.g. Asian option in the BS model).

We continue by studying the bias error of the first order extrapolation schemes, i.e. the HOWE scheme (44) for $M = 2$ and $N = 1$

$$E_n^1 = \lambda_0^1 \hat{P}^n + \lambda_1^1 \hat{P}^{2n} \, .$$

Then if assumption (45) holds, according to Proposition 3 the bias error $\text{bias}_n^1 = \mathbb{E}\left(E_n^1\right) - \mathbb{E}\left(P\right)$ should behave as

$$\text{bias}_n^1 \asymp n^{-2\alpha} \, .$$

Thus the order of the bias convergence of this scheme indicates the second term in the bias expansion. Morover such a behaviour would imply that

$$\frac{\text{bias}_n^1}{\text{bias}_{2n}^1} \approx 2^{2\alpha} \, .$$

Analogously to bias_n^0 we will use an MC estimate of E_n^1 to approximate

$$\overline{\text{bias}}_n^1 = \bar{E}_n^1 - \mathbb{E}\left(P\right) \, .$$

Using these approximation we check if the bias has the expected behaviour.

We repeat this procedure with the second order extrapolation scheme, i.e. the HOWE scheme (44) for $M = 2$ and $N = 2$

$$E_n^2 = \lambda_0^2 \hat{P}^n + \lambda_1^2 \hat{P}^{2n} + \lambda_2^2 \hat{P}^{4n}$$

to find the next term in the bias expansion. Analogously, under assumption (45) Proposition 3 yields that

$$\text{bias}_n^2 = \mathbb{E}\left(E_n^2\right) - \mathbb{E}\left(P\right) \asymp n^{-3\alpha}$$

and thus

$$\frac{\text{bias}_n^2}{\text{bias}_{2n}^2} \approx 2^{3\alpha} .$$

Then approximating bias_n^2 by $\overline{\text{bias}}_n^2 = \bar{E}_n^2 - \mathbb{E}\left(P\right)$ where $\bar{E}_n^2$ is an MC estimate of E_n^2, we check this convergence.

The procedure can be continued with the further HOWE scheme, but we are typically stopping at this point, as on the one hand this is usually sufficient to make a conclusion about the assumption (45) and on the other hand any further test take enormous run time.

Remark 23.

- *The MC estimates of* E_n^1 *and* E_n^2 *have the form of the combination of the HOWE and the classic MLMC estimator* (53)*.*
- *To carry out the bias analysis estimates of* bias_n^i *should be computed with an error*

$$\text{err}_n = \sqrt{\mathbb{V}\text{ar}\left(\overline{\text{bias}}_n^i\right)} ,$$

that is at least a magnitude less then the bias in order to have relabel confidence intervals for $\overline{\text{bias}}_n^i$*.*

- *Moreover in the case of the second order HOWE estimates the bias takes very small values. For instance in the case of a call option under the BS model already* $\overline{\text{bias}}_8^2 = 0.0009$*, while the corresponding error* $\text{err}_8 = 0.0003$ *and computation time is equal to* 10 *hours. Then to obtain* $\overline{\text{bias}}_8^2$ *with an error of a magnitude less we will need at least* 41 *day. This explains the lack of data in our tests and their long run time.*

3.5.1 *Black Scholes Model*

We start with the BS model for the following set of parameters

$$S_0 = 100 , \qquad r = 0.02 , \qquad \sigma = 0.2 .$$

In the most cases we use the Euler discretisation scheme applied to the stock SDE (1). Note that this indeed not the bast approach, as a solution of the BS SDE is known in a closed form (see Section 1.1.2.1). We are doing this on purpose to study biased estimates.

3.5.1.1 *Call Option*

We consider a European call option with maturity $T = 1$ and strike $K = 120$, which has a true value of

$$C = 2.5469 .$$

POWER SERIES BIAS EXPANSION

First we check if the power series bias expansion (45) holds. The results of this studies are summarised in Table 7. Additionally the reliable estimations are visualised on Figure 6, where we plot the bias errors of the Euler scheme, the first order and the second order extrapolation schemes against the number of discretisation points n in a log-log scale.

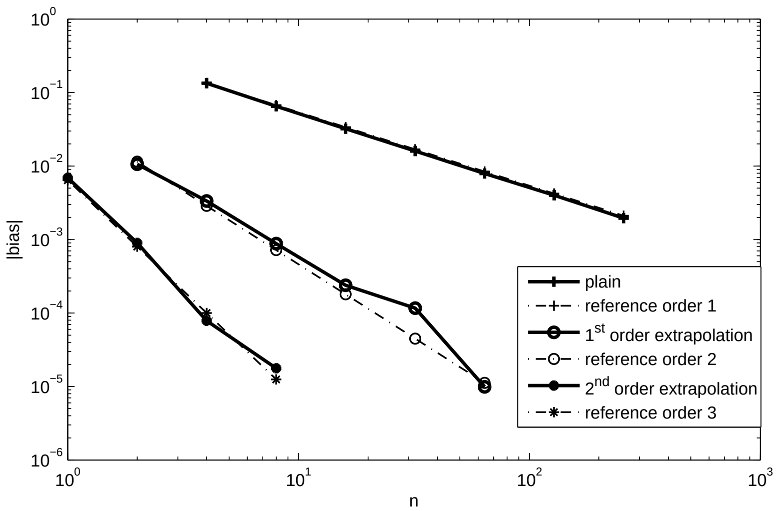


Figure 6: Bias convergence of the Euler scheme (BS model and call option)

According to Talay and Tubaro [63] we are expecting the bias power expansion (45) to be fulfilled for $\alpha = 1$. The "$\overline{\text{bias}}_n^0/\overline{\text{bias}}_{2n}^0$" row of Table 7a indicates that the bias error of the Euler scheme has an order of convergence one, i.e. $\alpha = 1$. However the "$\overline{\text{bias}}_n^2/\overline{\text{bias}}_{2n}^2$" and "$\overline{\text{bias}}_n^2/\overline{\text{bias}}_{2n}^2$" rows of Table 7b and Table 7c correspondingly are not so convincing, as for a big n the values therein are not necessary close to 4 and 8 correspondingly. This is explained by relatively small accuracies of the $\overline{\text{bias}}_n^1$ and $\overline{\text{bias}}_n^2$ estimates for a big n, see the "err_n" rows. Nevertheless the first three values in the "$\overline{\text{bias}}_n^1/\overline{\text{bias}}_{2n}^1$" row of Table 7b suggest that the bias has an order of convergence 2. Moreover the first value in the "$\overline{\text{bias}}_n^2/\overline{\text{bias}}_{2n}^2$" row of Table 7c seems to indicate that the bias has an order of convergence 3. Thus we conclude by assuming that assumption (45) is fulfilled for $\alpha = 1$.

n	4	8	16	32	64	128	256
$\bar{E}_n^0$	2.4131	2.4817	2.5147	2.5310	2.5390	2.5429	2.5450
$\overline{\text{bias}}_n^0$	−0.133	−0.065	−0.032	−0.016	−0.007	−0.004	−0.002
$\frac{\overline{\text{bias}}_n^0}{\overline{\text{bias}}_{2n}^0}$	2.0523	2.0253	2.0163	2.0201	1.9743	2.0452	
err_n	0.0003	0.0002	0.0002	0.0003	0.0003	0.0003	0.0003

(a) Bias convergence of the Euler scheme

n	4	8	16	32	64	128	256
$\bar{E}_n^1$	2.5575	2.5503	2.5478	2.5472	2.5470	2.5469	2.5469
$\overline{\text{bias}}_n^1$	0.0106	0.0033	0.0009	0.0002	0.0001	0.0000	0.0000
$\frac{\overline{\text{bias}}_n^1}{\overline{\text{bias}}_{2n}^1}$	3.1649	3.8068	3.6810	2.0498	11.7703	0.4482	
err_n	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003

(b) Bias convergence of the first order HOWE

n	4	8	16	32	64	128	256
$\bar{E}_n^2$	2.5538	2.5478	2.5470	2.5469	2.5470	2.5469	2.5469
$\overline{\text{bias}}_n^2$	0.0069	0.0009	0.0001	0.0000	0.0001	−0.0000	0.0000
$\frac{\overline{\text{bias}}_n^2}{\overline{\text{bias}}_{2n}^2}$	7.6828	11.4332	4.4280	0.2750	−3.6496	−1.7070	
err_n	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003

(c) Bias convergence of the second order HOWE

Table 7: Bias convergence for the Euler scheme (BS model and call option)

WEML AND MLMC

Now we compare performance of the WEML to the MLMC estimate. On Figure 7 we plot the RMSE against CC of the WEML and the MLMC estimates in a log-log scale. The speed up of the WEML compared to the MLMC is presented in Table 8 .

rMSE	0.1	0.05	0.01	0.005	0.001	0.0005	0.0001
speed up	1.6	1.8	2.3	2.6	3.3	3.7	4.7

Table 8: Speed up of WEML compared to MLMC (BS model and call option)

The fact that the CC graphs on Figure 7 have almost the same slope is in agreement with the results of Theorem 8 and the Complexity Theorem 7 for $\beta = 1$. In particular, the theory predicts the CC of the

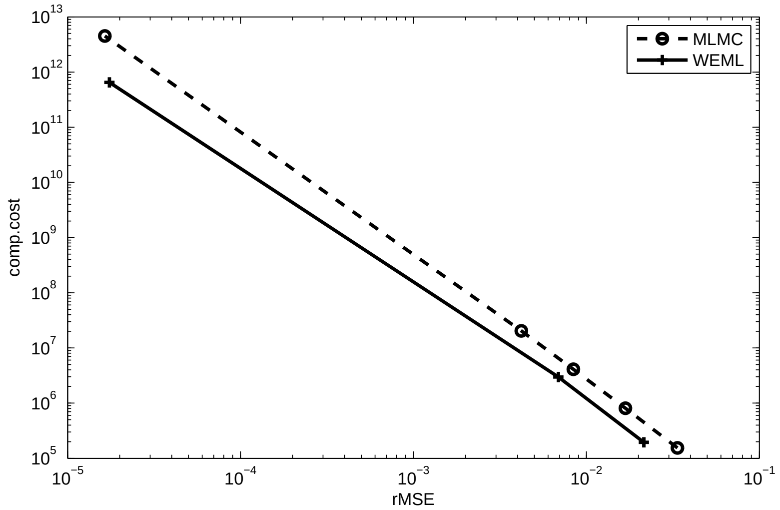


Figure 7: CC vs. RMSE (BS model and call option)

WEML and the MLMC estimates to differ in a log-term only, what leads to the moderately increasing speed up in Table 8.

3.5.1.2 *Digital Call Option*

We consider a European digital call option with maturity $T = 1$ and strike $K = 80$ scaled up by 100, which has a true value of

$$C = 85.0546 .$$

POWER SERIES BIAS EXPANSION

The results of checking the assumption (45) are summarised in Table 9. The reliable estimations therein are visualised on Figure 8 by plotting the bias errors of the Euler scheme, the first order and the second order extrapolation schemes against the number of discretisation points n in a log-log scale.

Due to Bally and Talay [7] it is expected that the bias power expansion (45) is fulfilled for $\alpha = 1$. The "$\overline{\text{bias}}_n^0/\overline{\text{bias}}_{2n}^0$" row of Table 9a indicates that the bias error of the Euler scheme has an order of convergence one, i.e. $\alpha = 1$. Though the "$\overline{\text{bias}}_n^2/\overline{\text{bias}}_{2n}^2$" and "$\overline{\text{bias}}_n^2/\overline{\text{bias}}_{2n}^2$" rows of Table 9b and especially Table 9c are not so convincing, as for a big n the values therein are not necessary close to 4 and 8 correspondingly, what is happening due to small relative accuracies of the $\overline{\text{bias}}_n^1$ and $\overline{\text{bias}}_n^2$ estimates. However the first two values in the "$\overline{\text{bias}}_n^1/\overline{\text{bias}}_{2n}^1$" row of Table 9b suggest that the bias has an order of convergence 2. The first value in the "$\overline{\text{bias}}_n^2/\overline{\text{bias}}_{2n}^2$" row of Table 9c might be seen as a weak evidence for the bias to have an order of convergence 3. As a result we say that assumption (45) is fulfilled for $\alpha = 1$.

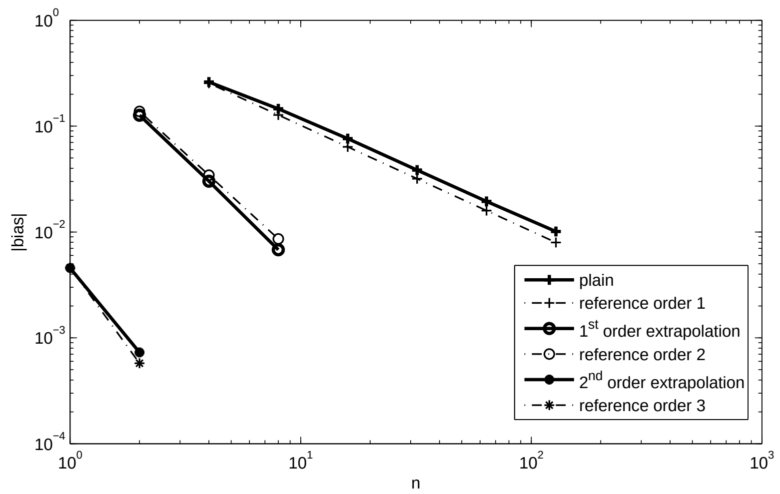


Figure 8: Bias convergence of the Euler scheme (BS model and digital option)

WEML AND MLMC

To compare performance of the estimates we plot the RMSE against CC of the WEML and the MLMC estimates on Figure 9 in a log-log scale. Moreover in Table 8 we present the speed up of the WEML

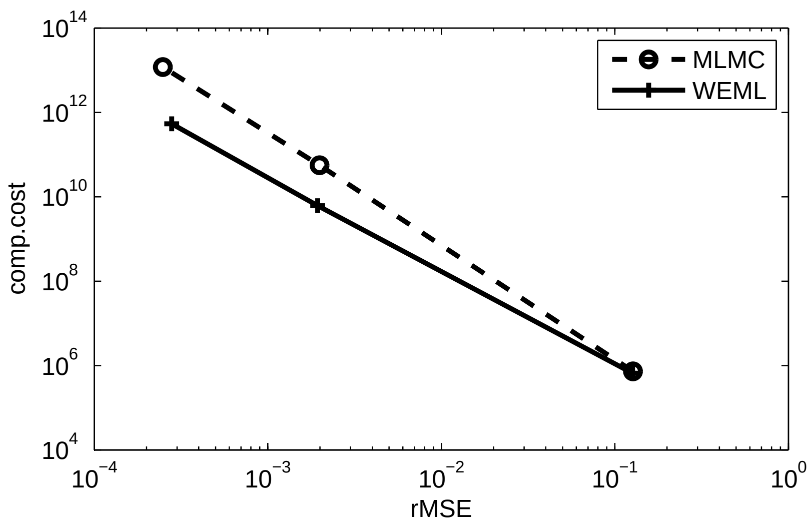


Figure 9: CC vs. RMSE (BS model and digital option)

compared to the MLMC.

As $\beta = 0.5$ the Complexity Theorem 7 and Theorem 8 predict the CC of the WEML to be almost a half an order better than the CC of the MLMC estimate. This is confirmed by a difference in the slopes of the CC graphs on Figure 9. Moreover the WEML algorithm is significantly faster than the MLMC algorithm for small errors (see Table 10)

n	4	8	16	32	64	128
$\bar{E}^0_n$	84.7933	84.9087	84.9785	85.0162	85.0352	85.0445
$\overline{\text{bias}}^0_n$	−0.2614	−0.1459	−0.0761	−0.0384	−0.0194	−0.0101
$\frac{\overline{\text{bias}}^0_n}{\overline{\text{bias}}^0_{2n}}$	1.7916	1.9164	1.9814	1.9775	1.9240	
err_n	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003

(a) Bias convergence of the Euler scheme

n	4	8	16	32	64	128
$\bar{E}^1_n$	84.9281	85.0244	85.0478	85.0536	85.0545	85.0543
$\overline{\text{bias}}^1_n$	−0.1265	−0.0302	−0.0068	−0.0010	−0.0001	−0.0003
$\frac{\overline{\text{bias}}^1_n}{\overline{\text{bias}}^1_{2n}}$	4.1889	4.4410	6.8008	7.7415	0.4460	
err_n	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003

(b) Bias convergence of the first order HOWE

n	8	16	32	64	128
$\bar{E}^2_n$	85.0592	85.0554	85.0552	85.0548	85.0544
$\overline{\text{bias}}^2_n$	0.0046	0.0007	0.0006	0.0002	0.0002
$\frac{\overline{\text{bias}}^2_n}{\overline{\text{bias}}^2_{2n}}$	6.2728	1.2229	3.1740	0.8724	
err_n	0.0004	0.0004	0.0004	0.0004	0.0004

(c) Bias convergence of the second order HOWE

Table 9: Bias convergence for the Euler scheme (BS model and digital option)

3.5.1.3 *Asian Call Option*

We consider a European Asian call option with maturity $T = 1$ and strike $K = 100$, which has a true value of

$$c \approx 5.0510\ .$$

Remark 24. *This reference value is obtained via the MLMC algorithm with* $\text{r}\mathbb{MSE} = 0.00002$*.*

Note, instead of the plain MC estimation of the starting level $\mathbb{E}\left(\hat{P}^1\right)$ *of the MLMC estimator we use its exact value. It can be computed via a BS type formula [45], because in the case of the trapezoidal integration rule*

$$\hat{P}^1 = e^{-rT}\left(\frac{1}{2}\left(S\left(0\right) + S\left(T\right)\right) - K\right)^+ ,$$

where $S\left(T\right)$ *is log-normally distributed. This idea leads to substantial savings in terms of CC. In general, it might be extended to some other options as well, e.g. barrier options.*

rMSE	0.1	0.05	0.01	0.005	0.001	0.0005	0.0001
speed up	1.3	1.8	3.8	5.1	10.6	14.4	29.7

Table 10: Speed up of WEML compared to MLMC (BS model and digital option)

The payoff of an Asian options involves an integral over time of the price of a stock, which is approximated numerically by a trapezoidal rule [18]

$$\int_0^T S(t)\,dt \approx \bar{S}^n := \sum_{i=1}^{n} \frac{\hat{S}^n((i-1)\Delta t) + \hat{S}^n(i\Delta t)}{2} \Delta t\,,$$

where $\hat{S}^n$ is the Euler scheme with n discretisation points. Hence,

$$\hat{P}^n = e^{-rT}\left(\frac{1}{T}\bar{S}^n - K\right)^+ .$$

POWER SERIES BIAS EXPANSION

Unfortunately the power series bias expansion (45) is not satisfied in this case. Even more the "$\overline{\text{bias}}_n^0/\overline{\text{bias}}_{2n}^0$" row of Table 11 indicates that the bias error of the Euler scheme does not converges as any power of n^{-1}. And Figure 10 suggests that the bias behaves as

$$\text{bias}_n^0 \asymp \frac{1}{ne^{-an}}$$

for some $a > 0$. As a result even the first order weak extrapolation

n	4	8	16	32
$\bar{E}_n^0$	5.0138	5.0432	5.0499	5.0511
$\overline{\text{bias}}_n^0$	0.0372	0.0078	0.0011	0.0001
$\frac{\overline{\text{bias}}_n^0}{\overline{\text{bias}}_{2n}^0}$	4.7882	7.2125	14.5507	
err_n	0.0001	0.0001	0.0001	0.00001

Table 11: Bias convergence for the Euler scheme (BS model and Asian option)

scheme cannot be constructed and, in particular, this explains unreliability of the weak extrapolation in this setting mentioned in [28].

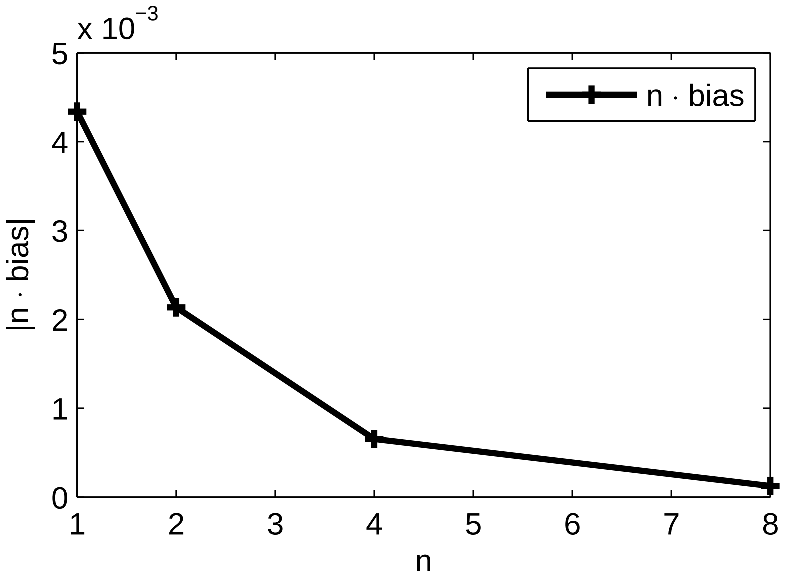


Figure 10: Bias convergence of the Euler scheme (BS model and Asian option): n bias plot

3.5.1.4 *Asian Call Option and Exact Price Simulations*

To overcome the above problem we make use of a closed form solution of the BS SDE. Then the integral over time of the price of a stock will be approximated numerically by the trapezoidal rule [18]

$$\int_0^T S(t)\,dt \approx \bar{S}^n := \sum_{i=1}^{n} \frac{S((i-1)\Delta) + S(i\Delta)}{2}\Delta\,, \qquad \text{for} \quad \Delta = \frac{T}{n}\,,$$

where S is a solution (2) of the BS SDE (1). Hence,

$$\hat{P}^n = e^{-rT}\left(\frac{1}{T}\bar{S}^n - K\right)^+ .$$

We consider the same European Asian call option with maturity $T = 1$ and strike $K = 100$, which has a true value of

$$C \approx 5.0510\,.$$

POWER SERIES BIAS EXPANSION

The results of checking assumption (45) are summarised in Table 12. Note, in this case we only consider the first order HOWE scheme. On Figure 11 we plot some estimated bias errors of the Euler scheme, the first order and the second order extrapolation schemes against the number of discretisation points n in a log-log scale.

As we simulate process S exactly and the trapezoidal integration rule has an error of the second order [18] the bias power expansion (45) is expected to be fulfilled with $\alpha = 2$. The values in the

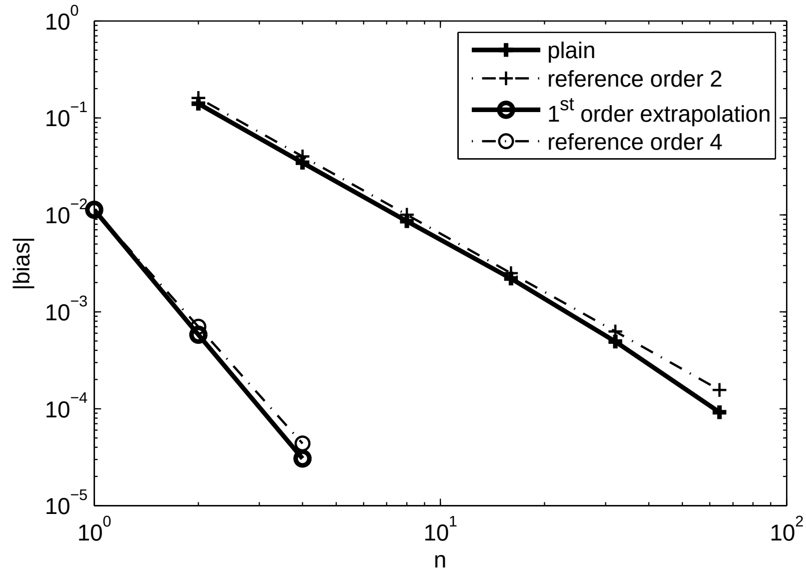


Figure 11: Bias convergence of the Euler scheme (BS model and Asian option)

"$\overline{\text{bias}}_n^0/\overline{\text{bias}}_{2n}^0$" row of Table 12a are all about 4, what confirms our expectations. Moreover the first two values in the "$\overline{\text{bias}}_n^1/\overline{\text{bias}}_{2n}^1$" row of Table 12b are about 16 suggesting that the bias has an order of convergence 4. Then we say that assumption (45) is fulfilled with $\alpha = 2$.

WEML AND MLMC

To compare performance of the estimates we plot the RMSE against CC of the WEML and the MLMC on Figure 12 in a log-log scale and present the speed up in Table 13.

For $\beta = 2$ according to the Complexity Theorem 7 and Theorem 8 CC of the WEML and the MLMC differs only in a constant. This explains the same slopes of the CC graphs on Figure 12 and the speed up that is uniform in RMSE (see Table 13).

3.5.1.5 *Single Barrier Option*

We consider a European down and out call option with maturity $T = 1$, strike $K = 100$ and barrier $B = 90$, which has a true value of

$$C = 7.3004 .$$

The exact price of the option is obtained via a closed form formula [61]. We approximate the barrier option payoff by means of the Euler scheme $\hat{S}^n$ and barrier cross monitoring on the discretisation points of $\hat{S}^n$.

n	2	4	8	16	32	64
$\bar{E}^0_n$	4.9112	5.0165	5.0424	5.0488	5.0505	5.0509
$\overline{\text{bias}}^0_n$	−0.1398	−0.0345	−0.0086	−0.0022	−0.0005	−0.0001
$\frac{\overline{\text{bias}}^0_n}{\overline{\text{bias}}^0_{2n}}$	4.0531	4.0213	3.8897	4.4848	5.3510	
err_n	0.0001	0.0001	0.0001	0.0001	0.00001	0.0001

(a) Bias convergence of the Euler scheme

n	2	4	8	16	32	64
$\bar{E}^1_n$	5.0623	5.0516	5.0510	5.0509	5.0510	5.0510
$\overline{\text{bias}}^1_n$	0.0113	0.0006	0.0000	−0.0001	0.0000	0.0000
$\frac{\overline{\text{bias}}^1_n}{\overline{\text{bias}}^1_{2n}}$	19.4358	18.9150	0.4867	1.2740	1.2026	
err_n	0.0001	0.0001	0.0001	0.0001	0.00001	0.0001

(b) Bias convergence of the first order HOWE

Table 12: Bias convergence for the Euler scheme (BS model and Asian option)

rMSE	0.1	0.05	0.01	0.005	0.001	0.0005	0.0001
speed up	1.9	1.9	1.9	1.9	1.9	1.9	1.9

Table 13: Speed up of WEML compared to MLMC (BS model and Asian option)

POWER SERIES BIAS EXPANSION

We summarise investigation of the power series bias expansion (45) in Table 14. The reliable estimations therein are visualised on Figure 13, where we plot the bias errors of the Euler scheme, the first order and the second order extrapolation schemes against the number of discretisation points n in a log-log scale.

It is known [30] that barrier options exhibit a slower order of weak convergence 0.5. In particular the "$\overline{\text{bias}}^0_n/\overline{\text{bias}}^0_{2n}$" row of Table 14a indicates that the bias error has an order of convergence 0.5, as values therein starting form $n = 16$ are all around $\sqrt{2} \approx 1.4142$. Further the "$\overline{\text{bias}}^1_n/\overline{\text{bias}}^1_{2n}$" values for $n \geqslant 16$ (see Table 14b) suggest that we gain a half of order, i.e. the next term in the bias expansion is n^{-1}. Thus if the expansion (45) holds, $\alpha = 0.5$. However the $\overline{\text{bias}}^2_{16}/\overline{\text{bias}}^2_{32}$ value in Table 14c seems to indicate that the next order of the bias expansion is 2. Note, the rest values of Table 14c are of a bad quality, as the error err_n is relatively high. As a result the bias seems to converge

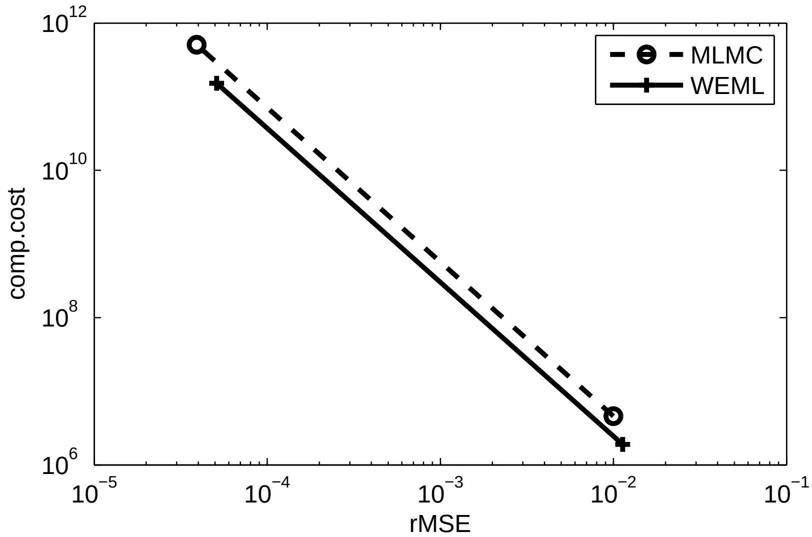


Figure 12: CC vs. RMSE (BS model and Asian option)

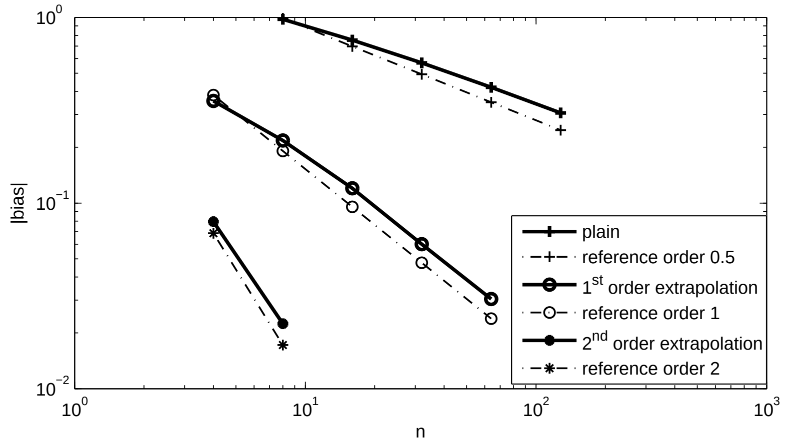


Figure 13: Bias convergence of the Euler scheme (BS model and barrier option)

smoothly. Although, does not satisfy assumption (45) in the usual meaning, as the first orders in the bias expansion are $n^{-1/2}$, n^{-1} and n^{-2}.

The HOWE scheme can be adopted to this bias expansion by finding corresponding weights. Therefore we can construct an appropriate WEML estimator, but this is out of the scope for this thesis.

Remark 25. *In some cases the $n^{-1/2}$ term appearing in the bias expansion can be eliminated by using advanced techniques [30], as e.g.*

- *the barrier shifting [13];*
- *the Brownian bridge [30].*

As a consequence the original WEML approach might be applied to price barrier options.

n	8	16	32	64	128
$\bar{E}_n^0$	8.2783	8.0555	7.8697	7.7204	7.6062
$\overline{\text{bias}}_n^0$	0.9778	0.7551	0.5692	0.4199	0.3057
$\frac{\overline{\text{bias}}_n^0}{\overline{\text{bias}}_{2n}^0}$	1.2950	1.3266	1.3556	1.3735	
err_n	0.0002	0.0002	0.0002	0.0001	0.0001

(a) Bias convergence of the Euler scheme

n	8	16	32	64	128
$\bar{E}_n^1$	7.6555	7.5176	7.4204	7.3605	7.3309
$\overline{\text{bias}}_n^1$	0.3551	0.2172	0.1200	0.0600	0.0304
$\frac{\overline{\text{bias}}_n^1}{\overline{\text{bias}}_{2n}^1}$	1.6349	1.8102	1.9981	1.9732	
err_n	0.0003	0.0003	0.0003	0.0003	0.0003

(b) Bias convergence of the first order HOWE

n	8	16	32	64	128
$\bar{E}_n^2$	7.3552	7.3798	7.3228	7.3013	7.3016
$\overline{\text{bias}}_n^2$	0.0548	0.0794	0.0224	0.0008	0.0011
$\frac{\overline{\text{bias}}_n^2}{\overline{\text{bias}}_{2n}^2}$	0.6903	3.5475	26.5369	0.7422	
err_n	0.0003	0.0003	0.0003	0.0001	0.0001

(c) Bias convergence of the second order HOWE

Table 14: Bias convergence for the Euler scheme (BS model and barrier option)

3.5.2 *Heston Model*

We continue with the Heston model for the following set of parameters

$$\begin{aligned} &S(0) = 100\,, \qquad r = 0.02\,, \qquad \rho = -0.8\,, \\ &V(0) = 0.1\,, \qquad \kappa = 3\,, \qquad \theta = 0.16\,, \qquad \sigma = 0.4\,. \end{aligned}$$

Note, this set of the parameters fulfils the Feller condition (see Section 1.1.2.2). We use the Euler discretisation scheme applied to the Heston SDE (3). To ensure positivity the volatility CIR process simulations we use the full truncation approach [53].

As we will see below for the above approximation scheme the bias power expansion (45) does not hold even for call options.

Remark 26. *The breach of the bias expansion* (45) *might be caused by many reasons, just to name some heuristic full truncation approach to simulation of a CIR process, non Lipschitz coefficients of the Heston SDE* (3)*, etc.*

To overcome this problem one can think of "log"-price simulations, some other techniques to simulate a CIR process [53] or even exact simulations [12], which are known to be slow, but combined with the WEML might be efficiently used.

3.5.2.1 *Call Option*

We consider a European call option with maturity $T = 1$ and strike $K = 100$, which has a true value of

$$C = 15.4300 .$$

POWER SERIES BIAS EXPANSION

In Table 15 we present estimated values of the bias error, that are also plotted on Figure 14. The values of "$\overline{\text{bias}}^0_n / \overline{\text{bias}}^0_{2n}$" for $n = 4$ and $n = 8$ might be interpreted as indicator for a second order weak convergence, but this is impossible. Otherwise the "$\overline{\text{bias}}^0_n / \overline{\text{bias}}^0_{2n}$" row of Table 15 rather indicates that bias error convergence is irregular for n under investigation. Hence, the power series bias expansion (45)

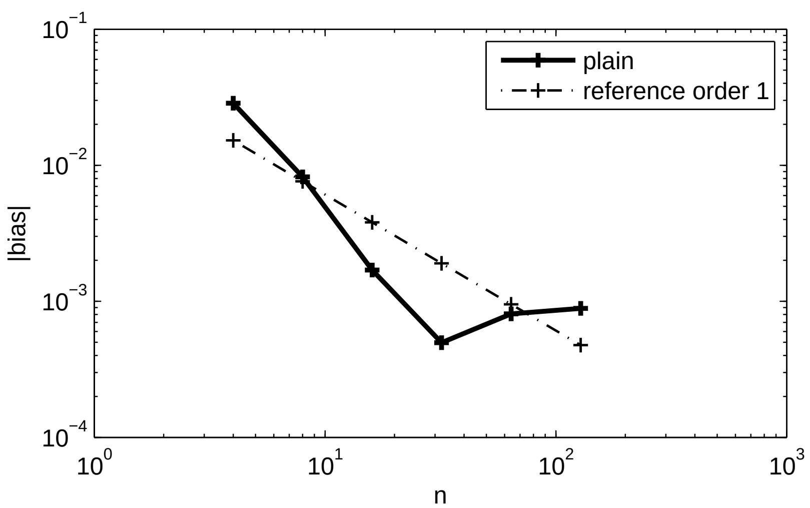


Figure 14: Bias convergence of the Euler scheme (Heston model and call option)

is not satisfied. As a result even the first order weak extrapolation method cannot be applied in this setting.

n	4	8	16	32	64	128
$\bar{E}_n^0$	15.4014	15.4218	15.4283	15.4295	15.4292	15.4291
$\overline{\text{bias}}_n^0$	0.0286	−0.0082	−0.0017	−0.0005	−0.0008	−0.0009
$\frac{\overline{\text{bias}}_n^0}{\overline{\text{bias}}_{2n}^0}$	3.4833	4.8290	3.4205	0.6152	0.9086	
err_n	0.0002	0.0002	0.0002	0.0002	0.0003	0.0003

Table 15: Bias convergence for the Euler scheme (Heston model and call option)

A

OPTIMISATION

A.1 OPTIMAL NUMBERS OF SIMULATED PATHS

We use the Lagrange multiplier optimisation technique [14] to solve the following problem

$$\sum_{i=0}^{L} N_i C_i \to \min$$
$$\text{s.t. } \sum_{i=0}^{L} N_i^{-1} V_i = \varepsilon^2 \,.$$

The corresponding Lagrange function Λ is equal to

$$\Lambda\left(N_0, N_1, \ldots, N_L, \lambda\right) = \sum_{i=0}^{L} N_i C_i + \lambda \left(\sum_{i=0}^{L} N_i^{-1} V_i - \varepsilon^2 \right) \,.$$

Then

$$\frac{\partial \Lambda}{\partial N_i} = C_i - \lambda V_i N_i^{-2}$$

implies that stationary point satisfy

$$N_i = \lambda^{-1/2} C_i^{-1/2} V_i^{1/2} \,.$$

Substituting N_i with the above expression in the constraint we obtain that

$$\lambda^{1/2} \sum_{i=0}^{L} C_i^{1/2} V_i^{1/2} = \varepsilon^2 \,,$$

what implies that

$$\lambda^{1/2} = \frac{\varepsilon^2}{\sum_{i=0}^{L} C_i^{1/2} V_i^{1/2}} \,.$$

Hence,

$$N_i = \frac{\sum_{i=0}^{L} \sqrt{C_i V_i}}{\varepsilon^2} \sqrt{\frac{V_i}{C_i}}$$

and the corresponding CC is equal to

$$C.C.\left[E\right] = \left(\sum_{i=0}^{L} \sqrt{C_i V_i} \right)^2 \,.$$

Remark 27. *Note, we omit checking second-order sufficient conditions for optimality, as this solution is known to be optimal, see e.g. [55].*

A.2 OPTIMAL NUMBERS OF DISCRETISATION POINT FOR EACH SCHEME

Instead of solving the original problem of

$$\left(n_0^{\gamma/2} + \sum_{i=1}^{L} n_i^{\gamma/2} n_{i-1}^{-\beta/2}\right)^2 \to \min$$

we are considering

$$f = n_0^{\gamma/2} + \sum_{i=1}^{L} n_i^{\gamma/2} n_{i-1}^{-\beta/2} \to \min\,,$$

that has the same point of optimum.

The gradient of the objective function f is of the following form

$$\begin{aligned}
\frac{\partial f}{\partial n_0} &= \frac{\gamma}{2}\, n_0^{\frac{\gamma}{2}-1} - \frac{\beta}{2}\, n_1^{\frac{\gamma}{2}} n_0^{-\frac{\beta}{2}-1}\,,\\
\frac{\partial f}{\partial n_i} &= \frac{\gamma}{2}\, n_i^{\frac{\gamma}{2}-1} n_{i-1}^{-\frac{\beta}{2}} - \frac{\beta}{2}\, n_i^{-\frac{\beta}{2}-1} n_{i+1}^{\frac{\gamma}{2}} \quad \text{for } i = 1, 2, \ldots, L-1\,.
\end{aligned}$$

Then the first order condition can be written as

$$\begin{aligned}
n_0^{\gamma+\beta} &= \left(\frac{\beta}{\gamma}\right)^2 n_1^{\gamma}\,,\\
n_i^{\gamma+\beta} &= \left(\frac{\beta}{\gamma}\right)^2 n_{i+1}^{\gamma} n_{i-1}^{\beta} \quad \text{for } i = 1, 2, \ldots, L-1\,.
\end{aligned} \tag{64}$$

A.2.1 *Solving System of Equations for n_i*

Defining

$$a = \frac{\beta}{\gamma} \qquad \text{and} \qquad b = \left(\frac{\beta}{\gamma}\right)^{\frac{2}{\gamma}}$$

we can rewrite the system of equations (64) as

$$\begin{aligned}
n_0^{\gamma+\beta} &= \left(\frac{\beta}{\gamma}\right)^2 n_1^{\gamma} & &\Rightarrow & n_1\, n_0^{-a} &= b^{-1}\, n_0\\
n_1^{\gamma+\beta} &= \left(\frac{\beta}{\gamma}\right)^2 n_2^{\gamma}\, n_0^{\beta} & &\Rightarrow & n_2\, n_1^{-a} &= b^{-1}\, n_1\, n_0^{-a}\\
n_2^{\gamma+\beta} &= \left(\frac{\beta}{\gamma}\right)^2 n_3^{\gamma}\, n_1^{\beta} & &\Rightarrow & n_3\, n_2^{-a} &= b^{-1}\, n_2\, n_1^{-a}\\
n_3^{\gamma+\beta} &= \left(\frac{\beta}{\gamma}\right)^2 n_4^{\gamma}\, n_2^{\beta} & &\Rightarrow & n_4\, n_3^{-a} &= b^{-1}\, n_3\, n_2^{-a}\\
&\ldots & &\ldots & &\ldots
\end{aligned}$$

$$
\begin{array}{rclcrcl}
n_i^{\gamma+\beta} &=& \left(\dfrac{\beta}{\gamma}\right)^2 n_{i+1}^{\gamma}\, n_{i-1}^{\beta} &\Rightarrow& n_{i+1}\, n_i^{-a} &=& b^{-1}\, n_i\, n_{i-1}^{-a} \\
\dots & & & \dots & & \dots & \\
n_{L-3}^{\gamma+\beta} &=& \left(\dfrac{\beta}{\gamma}\right)^2 n_{L-2}^{\gamma}\, n_{L-4}^{\beta} &\Rightarrow& n_{L-2}\, n_{L-3}^{-a} &=& b^{-1}\, n_{L-3}\, n_{L-4}^{-a} \\
n_{L-2}^{\gamma+\beta} &=& \left(\dfrac{\beta}{\gamma}\right)^2 n_{L-1}^{\gamma}\, n_{L-3}^{\beta} &\Rightarrow& n_{L-1}\, n_{L-2}^{-a} &=& b^{-1}\, n_{L-2}\, n_{L-3}^{-a} \\
n_{L-1}^{\gamma+\beta} &=& \left(\dfrac{\beta}{\gamma}\right)^2 n_{L}^{\gamma}\, n_{L-2}^{\beta} &\Rightarrow& n_{L}\, n_{L-1}^{-a} &=& b^{-1}\, n_{L-1}\, n_{L-2}^{-a}
\end{array}
$$

Then starting form the first equation we substitute the RHS of each equation with the LHS of the previous that gives the following system equations with immediate implications

$$
\begin{array}{rclcrcl}
n_1\, n_0^{-a} &=& b^{-1}\, n_0 &\Rightarrow& n_1 &=& b^{-1}\, n_0\, n_0^{a} \\
n_2\, n_1^{-a} &=& b^{-2}\, n_0 &\Rightarrow& n_2 &=& b^{-2}\, n_0\, n_1^{a} \\
n_3\, n_2^{-a} &=& b^{-3}\, n_0 &\Rightarrow& n_3 &=& b^{-3}\, n_0\, n_2^{a} \\
n_4\, n_3^{-a} &=& b^{-4}\, n_0 &\Rightarrow& n_4 &=& b^{-4}\, n_0\, n_3^{a} \\
& \dots & & \dots & & \dots & \\
n_{i+1}\, n_i^{-a} &=& b^{-(i+1)}\, n_0 &\Rightarrow& n_{i+1} &=& b^{-(i+1)}\, n_0\, n_i^{a} \\
& \dots & & \dots & & \dots & \\
n_{L-2}\, n_{L-3}^{-a} &=& b^{-(L-2)}\, n_0 &\Rightarrow& n_{L-2} &=& b^{-(L-2)}\, n_0\, n_{L-3}^{a} \\
n_{L-1}\, n_{L-2}^{-a} &=& b^{-(L-1)}\, n_0 &\Rightarrow& n_{L-1} &=& b^{-(L-1)}\, n_0\, n_{L-2}^{a} \\
n_{L}\, n_{L-1}^{-a} &=& b^{-L}\, n_0 &\Rightarrow& n_{L} &=& b^{-L}\, n_0\, n_{L-1}^{a}
\end{array} \tag{65}
$$

Going backwards (i.e. starting with the last equation) through the above system we express n_i in terms of n_0

$$n_i = b^{-\sum_{j=0}^{i-1}\left((i-j)a^j\right)}\, n_0^{\sum_{j=0}^{i} a^j} . \tag{66}$$

For the later we have distinguish between two case

- $a = 1$,
- $a \neq 1$.

A.2.1.1 *Case of* $a = 1$

As $b = 1$, Equation 66 gives that

$$n_0 = n_L^{\frac{1}{L+1}} ,$$

Moreover the system of equations (65) implies that

$$n_i = \frac{n_{i+1}}{n_0} .$$

Hence,

$$n_i = n_L^{\frac{i+1}{L+1}} \quad \text{for } i = 0, 1, \dots, L-1 .$$

A.2.1.2 *Case of* $a \neq 1$

Note, that

$$\sum_{j=0}^{i} a^j = \frac{a^{i+1}-1}{a-1} \ ,$$

as the sum of a geometric progression. Moreover differentiating the formula for a sum of a geometric progression we first derive that

$$\sum_{j=1}^{i}\left(j\left(\frac{1}{a}\right)^{j-1}\right) = \frac{d}{d\frac{1}{a}}\left(\sum_{j=0}^{i}\left(\frac{1}{a}\right)^{j}\right) = \frac{d}{d\frac{1}{a}}\left(\frac{\left(\frac{1}{a}\right)^{i+1}-1}{\frac{1}{a}-1}\right)$$

$$= \frac{i\ \left(\frac{1}{a}\right)^{i+1} - (i+1)\ \left(\frac{1}{a}\right)^{i} + 1}{\left(\frac{1}{a}-1\right)^2} \ ,$$

and thus

$$\Rightarrow a^{i-1} \sum_{j=1}^{i}\left(j\left(\frac{1}{a}\right)^{j-1}\right) = \frac{a^{i+1}-(i+1)\ a+i}{(a-1)^2} \ .$$

Then rewriting (66) as

$$n_i = b^{-a^{i-1}\ \sum_{j=1}^{i}\left\{j\ \left(\frac{1}{a}\right)^{j-1}\right\}}\ n_0^{\sum_{j=0}^{i} a^j} \ ,$$

and plugging in the above expressions for the sums we obtain

$$n_i = b^{-\frac{a^{i+1}-(i+1)\ a+i}{(a-1)^2}}\ n_0^{\frac{a^{i+1}-1}{a-1}} \ .$$

In particular, we can express n_0 as function of n_L

$$n_L = b^{-\frac{a^{L+1}-(L+1)\ a+L}{(a-1)^2}}\ n_0^{\frac{a^{L+1}-1}{a-1}}$$

$$\Rightarrow n_0 = b^{\frac{1}{a-1}-\frac{L+1}{a^{L+1}-1}}\ n_L^{\frac{a-1}{a^{L+1}-1}} \ .$$

As a consequence

$$n_i = b^{\frac{(i+1)\ \left(a^{L+1}-1\right)-(L+1)\ \left(a^{i+1}-1\right)}{\left(a^{L+1}-1\right)\ (a-1)}}\ n_L^{\frac{a^{i+1}-1}{a^{L+1}-1}} \ .$$

A.2.2 *Value at Optima*

Note that the system (64) implies that

$$n_{i+1}^{\gamma/2}\ n_i^{-\beta/2} = \frac{1}{a}\ n_i^{\gamma/2}\ n_{i-1}^{-\beta/2} \qquad \text{for } i = 1, 2, \ldots, L-1 \ .$$

Hence

$$f = n_0^{\gamma/2} + \sum_{i=1}^{L} n_i^{\gamma/2} n_{i-1}^{-\beta/2} = n_0^{\frac{\gamma}{2}} \sum_{j=0}^{L}\left(\frac{1}{a}\right)^j \ ,$$

and in particular

$$f = \begin{cases} (L+1)\ n_0^{\frac{\gamma}{2}} & \text{if} \quad a = 1\,, \\ \frac{\left(\frac{1}{a}\right)^{L+1}-1}{\frac{1}{a}-1}\ n_0^{\frac{\gamma}{2}} & \text{if} \quad a \neq 1\,. \end{cases}$$

Then substituting n_0 with the findings of the previous section we obtain

$$f = \begin{cases} (L+1)\, n_L^{\frac{1}{L+1}\ \frac{\gamma}{2}} & \text{if} \quad a = 1\,, \\ \frac{a^{L+1}-1}{a^{L+1}-a^L}\, a^{\left(\frac{1}{a-1}-\frac{(L+1)}{a^{L+1}-1}\right)}\ n_L^{\frac{a-1}{a^{L+1}-1}\ \frac{\gamma}{2}} & \text{if} \quad a \neq 1\,. \end{cases} \tag{67}$$

Remark 28. *Note that the objective function* f *is a* posynomial. *Then using techniques of the* geometric programming *[11], and in particular log-exponent transformation, it can be shown that* f *has an unique global minimum.*

As a consequence the stationary point found in the above considerations is the global minimum of the objective function f.

A.3 OPTIMAL NUMBER OF LEVELS

Treating L as a real number we will solve

$$G\,(L) \to \min\,,$$

where the objective function G is as defined in Equation 29. In the following we distinguish between two case: $a = 1$ and $a \neq 1$.

Case of $a = 1$

We will consider the equivalent problem of

$$f = (L+1)\, n_L^{\frac{1}{L+1}\ \frac{\gamma}{2}} \to \min\ .$$

As the derivative of the objective function is equal to

$$\begin{aligned} f' &= n_L^{\frac{1}{L+1}\ \frac{\gamma}{2}} + (L+1)\, n_L^{\frac{1}{L+1}\ \frac{\gamma}{2}}\ \frac{-\gamma \ln(n_L)}{2\ (L+1)^2} \\ &= \frac{1}{2}\, n_L^{\frac{1}{L+1}\ \frac{\gamma}{2}} \left(2 - \frac{\gamma}{L+1}\ \ln(n_L)\right)\,, \end{aligned}$$

the first order condition implies that

$$L = \frac{\gamma}{2} \ln(n_L) - 1\ .$$

Hence,

$$n_i = n_L^{\frac{i+1}{L+1}} = \left(n_L^{\frac{1}{\ln(n_L)}}\right)^{\frac{i+1}{\gamma/2}} = \left(e^{2/\gamma}\right)^{i+1}\,,$$

and moreover

$$n_i = \left(e^{2/\gamma}\right) n_{i-1} \ .$$

And the corresponding

$$\mathrm{C.C.}\left[\hat{E}\right] \asymp \left(e\frac{\gamma}{2}\ln\left(n_L\right)\right)^2 \ .$$

Moreover as

$$\begin{aligned}
f'' &= \frac{d}{dL}\left(\frac{1}{2}\, n_L^{\frac{1}{L+1}\ \frac{\gamma}{2}}\ \left(2 - \frac{\gamma}{L+1}\ \ln\left(n_L\right)\right)\right) \\
&= -\frac{\gamma}{4} n_L^{\frac{1}{L+1}\frac{\gamma}{2}} \frac{\ln\left(n_l\right)}{\left(L+1\right)^2}\left(2 - \frac{\gamma}{L+1}\ln\left(n_L\right)\right) + \frac{1}{2} n_L^{\frac{1}{L+1}\frac{\gamma}{2}} \frac{\gamma\ln\left(n_L\right)}{\left(L+1\right)^2} \\
&= \frac{1}{2} n_L^{\frac{1}{L+1}\frac{\gamma}{2}} \frac{\gamma\ln\left(n_L\right)}{\left(L+1\right)^2}\left(-1 + \frac{1}{2}\frac{\gamma\ln\left(n_L\right)}{L+1} + 1\right) \\
&= \frac{1}{4} n_L^{\frac{1}{L+1}\frac{\gamma}{2}} \frac{\gamma^2\left(\ln\left(n_L\right)\right)^2}{\left(L+1\right)^3} \ > 0
\end{aligned}$$

for positive L and $n_L > 1$. The objective function f is convex on this domain. Hence the above found stationary point is the unique global minimum of f. Furthermore the function f does not have any other local minimums.

Case of $a \neq 1$

Analogously to the previous case, the original problem is substituted by

$$f = \frac{a^{L+1} - 1}{a^{L+1} - a^L}\ a^{\left(\frac{1}{a-1} - \frac{(L+1)}{a^{L+1}-1}\right)}\ n_L^{\frac{a-1}{a^{L+1}-1}\cdot\frac{\gamma}{2}} \to \min \ .$$

For the sake of notation convenience we define

$$\begin{aligned}
X &= \frac{a^{L+1} - 1}{a^{L+1} - a^L} \ , \\
Y &= a^{\left(\frac{1}{a-1} - \ \frac{(L+1)}{a^{L+1}-1}\right)} \ , \\
Z &= n_L^{\frac{a-1}{a^{L+1}-1}\cdot\frac{\gamma}{2}} \ .
\end{aligned}$$

Then the derivative of the objective function is equal to

$$f' = X'YZ + XY'Z + XYZ' \ .$$

In the following we calculate derivatives of X, Y, Z and combine them at the very last step.

$$\frac{d}{dL}\left(X\right) = \frac{\ln\left(a\right)}{a^{L+1} - a^L} \ .$$

$$\frac{d}{dL}(Y) = Y\ln(a)\frac{d}{dL}\left(\frac{1}{a-1} - \frac{(L+1)}{a^{L+1}-1}\right)$$
$$= Y\ln(a)\,\frac{(L+1)\ \ln(a)\ a^{L+1} - a^{L+1} + 1}{(a^{L+1}-1)^2} .$$

$$\frac{d}{dL}(Z) = Z\ln(n_L)\,(a-1)\,\frac{\gamma}{2}\frac{d}{dL}\left(\frac{1}{a^{L+1}-1}\right)$$
$$= Z\ln(n_L)\,(a-1)\,\frac{\gamma}{2}\left(-\frac{\ln(a)\,a^{L+1}}{(a^{L+1}-1)^2}\right) .$$

Then combining all together we obtain

$$\begin{aligned}
f' &= YZ\ln(a)\left(\frac{1}{a^{L+1}-a^{L}}\right.\\
&+ \frac{a^{L+1}-1}{a^{L+1}-a^{L}}\frac{(L+1)\ln(a)\,a^{L+1} - a^{L+1} + 1}{(a^{L+1}-1)^2}\\
&\left.- \frac{a^{L+1}-1}{a^{L+1}-a^{L}}\ln(n_L)\,(a-1)\ \frac{\gamma}{2}\ \frac{a^{L+1}}{(a^{L+1}-1)^2}\right)\\
&= YZ\ln(a)\ \frac{1}{(a^{L+1}-a^{L})}\left(1 + \frac{(L+1)\ \ln(a)\ a^{L+1}}{a^{L+1}-1}\right.\\
&\left.-1 - \ln(n_L)\,(a-1)\ \frac{\gamma}{2}\ \frac{a^{L+1}}{a^{L+1}-1}\right)\\
&= YZ\ln(a)\,\frac{a^{L+1}\left((L+1)\ln(a) - \ln(n_L)\,(a-1)\,\frac{\gamma}{2}\right)}{(a^{L+1}-1)\,(a^{L+1}-a^{L})}
\end{aligned}$$

Then the first order condition implies that

$$L = \frac{(a-1)\,\gamma}{2\ln(a)}\ln(n_L) - 1 .$$

For this choice of L

$$\begin{aligned}
n_i &= b^{\frac{(i+1)\ \left(a^{L+1}-1\right) - (L+1)\ \left(a^{i+1}-1\right)}{\left(a^{L+1}-1\right)\ (a-1)}}\ n_L^{\frac{a^{i+1}-1}{a^{L+1}-1}}\\
&= \left(b^{\frac{1}{a-1}}\right)^{i+1}\ \left(b^{\frac{L+1}{a-1}}\right)^{-\frac{a^{i+1}-1}{a^{L+1}-1}}\ n_L^{\frac{a^{i+1}-1}{a^{L+1}-1}}\\
&= \left(b^{\frac{1}{a-1}}\right)^{i+1}\ \left(b^{\frac{\gamma\ \ln(n_L)}{2\ \ln(a)}}\right)^{-\frac{a^{i+1}-1}{a^{L+1}-1}}\ n_L^{\frac{a^{i+1}-1}{a^{L+1}-1}}\\
&= \left(b^{\frac{1}{a-1}}\right)^{i+1}\ \left(e^{\ln(n_L)}\right)^{-\frac{a^{i+1}-1}{a^{L+1}-1}}\ n_L^{\frac{a^{i+1}-1}{a^{L+1}-1}}\\
&= \left(b^{\frac{1}{a-1}}\right)^{i+1} = R^{i+1} .
\end{aligned}$$

Using $n_i^{\gamma/2} n_{i-1}^{-\beta/2} = a^{\frac{1}{a-1}} a^{-i}$ we obtain the corresponding

$$\begin{aligned} \text{C.C.}\left[\hat{E}\right] &\asymp \left(n_0^{\gamma/2} + \sum_{i=1}^{L} n_i^{\gamma/2} n_{i-1}^{-\beta/2}\right)^2 = \left(a^{\frac{1}{a-1}} \sum_{i=0}^{L} a^{-i}\right)^2 \\ &= a^{\frac{2}{a-1}} \left(\frac{a^{-L-1}-1}{a^{-1}-1}\right)^2 = \left(\frac{a^{\frac{a}{a-1}}}{a-1}\right)^2 \left(n_L^{\frac{(1-a)\gamma}{2}} - 1\right)^2 . \end{aligned}$$

Note that for $\forall n_L \geqslant R^2$

- $f' < 0$ for $L \in \left(0, \frac{(a-1)\gamma}{2\ln(a)} \ln(n_L) - 1\right)$;
- $f' > 0$ for $L \in \left(\frac{(a-1)\gamma}{2\ln(a)} \ln(n_L) - 1, +\infty\right)$.

Hence the above found stationary point is the unique global minimum of f. Moreover the function f does not have any other local minimums.

A.4 COMPUTATIONAL COST OF THE MLC ESTIMATE

Proof of Proposition 1 on the CC of the MLC estimate. As the values of the numbers of simulated paths N_i, given by Equation 36, are bigger than

$$\left\lceil \frac{\sum_{i=0}^{L} \sqrt{n_i^{\gamma} V_i}}{\varepsilon^2} \sqrt{n_i^{-\gamma} V_i} \right\rceil \geqslant \frac{\sum_{i=0}^{L} \sqrt{n_i^{\gamma} V_i}}{\varepsilon^2} \sqrt{n_i^{-\gamma} V_i} \,,$$

the solution of the continuous optimisation problem, feasibility of the optimal solution ensures that

$$\mathbb{V}\text{ar}\left(\hat{E}\right) \leqslant \varepsilon^2 .$$

Further we will establish bounds for the CC of the estimate.

$$\begin{aligned} \text{C.C.}\left[\hat{E}\right] &= \sum_{i=0}^{L} N_i C_i \leqslant \sum_{i=0}^{L} \left\{ \left(\frac{\sum_{i=0}^{L} \sqrt{C_i V_i}}{\varepsilon^2} \sqrt{\frac{V_i}{C_i}} + 1 \right) C_i \right\} \\ &= \sum_{i=0}^{L} C_i + \varepsilon^{-2} \left(\sum_{i=1}^{L} \sqrt{C_i V_i} \right)^2 . \end{aligned}$$

Note that for a fixed n_L the n_i and L are fixed and thus C_i and V_i as well. Then for sufficiently small ε

$$\sum_{i=0}^{L} C_i \leqslant \varepsilon^{-2} \left(\sum_{i=1}^{L} \sqrt{C_i V_i} \right)^2 .$$

As a result

$$\text{C.C.}\left[\hat{E}\right] \leqslant 2\varepsilon^{-2} \left(\sum_{i=1}^{L} \sqrt{C_i V_i} \right)^2 . \tag{68}$$

In the following we will derive a bound for $\left(\sum_{i=1}^{L}\sqrt{C_i V_i}\right)^2$.
As

$$\mathbb{V}\mathrm{ar}\left(P^{n_i}-P^{n_{i-1}}\right)\leqslant\left(\sqrt{\mathbb{V}\mathrm{ar}\left(P-P^{n_i}\right)}+\sqrt{\mathbb{V}\mathrm{ar}\left(P-P^{n_{i-1}}\right)}\right)^2,$$

under the assumption of the proposition we obtain that

$$V_i\leqslant\begin{cases} c_v & \text{for}\quad i=0\,,\\ c_v\left(n_i^{-\beta/2}+n_{i-1}^{-\beta/2}\right)^2 & \text{for}\quad i=1,2,\dots,L\,.\end{cases}$$

Moreover the assumption on CC of the scheme implies that

$$C_i\leqslant\begin{cases} c_c\ n_0^{\gamma} & \text{for}\quad i=0\,,\\ c_c\ \left(n_i^{\gamma}+n_{i-1}^{\gamma}\right) & \text{for}\quad i=1,2,\dots,L\,.\end{cases}$$

Then using the following inequality

$$\sqrt{n_i^{\gamma}+n_{i-1}^{\gamma}}\leqslant n_i^{\gamma/2}+n_{i-1}^{\gamma/2}\,,$$

we can establish that

$$\left(\sum_{i=1}^{L}\sqrt{C_iV_i}\right)^2\leqslant c_cc_v\left(n_0^{\frac{\gamma}{2}}+\sum_{i-1}^{L}\left(n_i^{\frac{\gamma}{2}}+n_{i-1}^{\frac{\gamma}{2}}\right)\left(n_i^{-\frac{\beta}{2}}+n_{i-1}^{-\frac{\beta}{2}}\right)\right)^2\,.$$

As n_i is increasing w.r.t. i and γ, β are positive

$$\left(\sum_{i=1}^{L}\sqrt{C_iV_i}\right)^2\leqslant 16c_cc_v\left(n_0^{\frac{\gamma}{2}}+\sum_{i-1}^{L}n_i^{\frac{\gamma}{2}}n_{i-1}^{-\frac{\beta}{2}}\right)^2\,.$$

To make use of the optimisation results we bound $\left(\sum_{i=1}^{L}\sqrt{C_iV_i}\right)^2$ in terms of the "continuous" $\bar{n}_i$. Then using

$$\bar{n}_i-1\leqslant n_i\leqslant\bar{n}_i+1$$

we derive

$$\begin{aligned}\left(\sum_{i=1}^{L}\sqrt{C_i\ V_i}\right)^2&\leqslant 16\ c_c\ c_v\left((\bar{n}_0+1)^{\frac{\gamma}{2}}+\sum_{i-1}^{L}(\bar{n}_i+1)^{\frac{\gamma}{2}}(\bar{n}_{i-1}-1)^{-\frac{\beta}{2}}\right)^2\\&\leqslant 16\ c_c\ c_v\ 2^{\gamma+\beta}\left(\bar{n}_0^{\frac{\gamma}{2}}+\sum_{i=1}^{L}\bar{n}_i^{\frac{\gamma}{2}}\bar{n}_{i-1}^{-\frac{\beta}{2}}\right)^2\,.\end{aligned}$$

The results obtained during the optimisation, in particular Equation 67, imply that

$$\left(\sum_{i=1}^{L}\sqrt{C_i\ V_i}\right)^2\leqslant 16\ c_c\ c_v\ 2^{\gamma+\beta}\left(f\left(L\right)\right)^2\,, \tag{69}$$

where f defined as in Equation 67

$$f(L) = \begin{cases} (L+1)\, n_L^{\frac{1}{L+1}\,\frac{\gamma}{2}} & \text{if} \quad a = 1\,, \\ \frac{a^{L+1}-1}{a^{L+1}-a^{L}}\, a^{\left(\frac{1}{a-1} - \frac{(L+1)}{a^{L+1}-1}\right)} n_L^{\frac{a-1}{a^{L+1}-1}\,\frac{\gamma}{2}} & \text{if} \quad a \neq 1\,. \end{cases}$$

As $f(L)$ has only local minimum and

$$\log_{[R]}(n_L) - 2 = \bar{L} - 1 \leqslant \lfloor L \rfloor \leqslant \bar{L} \leqslant \lceil L \rceil \leqslant \bar{L} + 1 = \log_{[R]}(n_L)$$

it follows that

$$f(L) \leqslant \max\left\{ f\left(\bar{L}-1\right), f\left(\bar{L}+1\right) \right\}\ .$$

In the following we have to distinguish between tow cases.

CASE OF $a = 1$

First, let us deal with $f\left(\bar{L}+1\right)$

$$\begin{aligned} f\left(\bar{L}+1\right) &= \left(\frac{\gamma}{2}\ \ln(n_L) + 1\right)\ n_L^{\frac{1}{\ln(n_L)+2/\gamma}} \leqslant \left(\frac{\gamma}{2}\ \ln(n_L) + 1\right)\ n_L^{\frac{1}{\ln(n_L)}} \\ &\leqslant e\ \left(\frac{\gamma}{2}\ \ln(n_L) + 1\right)\ . \end{aligned}$$

On the other hand

$$f\left(\bar{L}-1\right) = \left(\frac{\gamma}{2}\ \ln(n_L) - 1\right)\ n_L^{\frac{1}{\ln(n_L)-2/\gamma}}\ .$$

As $n_L \geqslant R^2 = e^{4/\gamma}$ implies that $2/\ln(n_L) \geqslant 1/(\ln(n_L) - 2/\gamma)$ it holds that

$$\begin{aligned} f\left(\bar{L}-1\right) &\leqslant \left(\frac{\gamma}{2}\ \ln(n_L) - 1\right)\ n_L^{\frac{2}{\ln(n_L)}} \\ &\leqslant e^2\ \left(\frac{\gamma}{2}\ \ln(n_L) - 1\right)\ . \end{aligned}$$

Hence,

$$f(L) \leqslant e^2\ \left(\frac{\gamma}{2}\ \ln(n_L) + 1\right)\ .$$

Then plugging it into Equation 69 and combining with Equation 68 we derive

$$\begin{aligned} C.C.\left[\hat{E}\right] &\leqslant 32\varepsilon^{-2} c_c c_\nu 2^{\gamma+\beta} \left(f(L)\right)^2 \\ &\leqslant \tilde{C}\left(\frac{\gamma}{2}\ln(n_L) + 1\right)^2 \varepsilon^{-2}\ . \end{aligned}$$

Furthermore for $n_L \geqslant R^2$

$$C.C.\left[\hat{E}\right] \leqslant C\left(\ln(n_L)\right)^2 \varepsilon^{-2}\ ,$$

what proves the assertion in the case of $\beta = \gamma$.

CASE OF $a \neq 1$

Note that $f(L)$ can be rewritten as

$$f(L) = \frac{a^{L+1}-1}{a^{L+1}-a^{L}} a^{\left(\frac{1}{a-1}-\frac{(L+1)}{a^{L+1}-1}\right)} n_L^{\frac{a-1}{a^{L+1}-1}\frac{\gamma}{2}}$$

$$= \frac{a^{\frac{a}{a-1}}}{a-1}\left(a^{-L-1}-1\right)\left(a^{-L-1}\, n_L^{\frac{\beta-\gamma}{2}}\right)^{\frac{1}{a^{L+1}-1}} . \qquad (70)$$

We begin with $f\left(\bar{L}+1\right)$ by noting that

$$a^{L} = \left(a^{\frac{\beta-\gamma}{2\ln(a)}}\right)^{\ln(n_L)} = n_L^{\frac{\beta-\gamma}{2}} .$$

Then Equation 70 can be transformed as

$$f\left(\bar{L}+1\right) = \frac{a^{\frac{a}{a-1}}}{a-1}\left(\frac{n_L^{\frac{\gamma-\beta}{2}}}{a}-1\right) a^{\frac{1}{1-a^{L+1}}}$$

$$= \frac{a^{\frac{a}{a-1}}}{a-1}\left(n_L^{\frac{\gamma-\beta}{2}}-a\right) a^{\frac{a^{L+1}}{1-a^{L+1}}}$$

$$= \frac{a^{\frac{a}{a-1}}}{a-1}\left(n_L^{\frac{\gamma-\beta}{2}}-a\right) a^{\frac{a\, n_L^{\frac{\beta-\gamma}{2}}}{1-a\, n_L^{\frac{\beta-\gamma}{2}}}}$$

$$= \frac{a^{\frac{a}{a-1}}}{a-1}\left(n_L^{\frac{\gamma-\beta}{2}}-a\right) a^{\frac{n_L^{\frac{\beta-\gamma}{2}}}{a^{-1}-n_L^{\frac{\beta-\gamma}{2}}}} .$$

Moreover for $n_L \geqslant R^2$

$$a^{\frac{n_L^{\frac{\beta-\gamma}{2}}}{a^{-1}-n_L^{\frac{\beta-\gamma}{2}}}} \leqslant 1$$

and thus

$$f\left(\bar{L}+1\right) \leqslant \frac{a^{\frac{a}{a-1}}}{a-1}\left(n_L^{\frac{\gamma-\beta}{2}}-a\right) .$$

On the other hand in the case of $f\left(\bar{L}-1\right)$

$$a^{L+1} = a^{-1} a^{\log_{[R]}(n_L)} = a^{-1} n_L^{\frac{\beta-\gamma}{2}} .$$

Then Equation 70 takes the following form

$$
\begin{aligned}
f\left(\bar{L}-1\right) &= \frac{a^{\frac{a}{a-1}}}{a-1}\left(a\, n_L^{\frac{\gamma-\beta}{2}}-1\right)\, a^{\frac{1}{a^{L+1}-1}} \\
&= \frac{a^{\frac{a}{a-1}}}{a-1}\left(n_L^{\frac{\gamma-\beta}{2}}-a^{-1}\right)\, a^{\frac{a^{L+1}}{a^{L+1}-1}} \\
&= \frac{a^{\frac{a}{a-1}}}{a-1}\left(n_L^{\frac{\gamma-\beta}{2}}-a^{-1}\right)\, a^{\frac{a^{-1}\, n_L^{\frac{\beta-\gamma}{2}}}{a^{-1}\, n_L^{\frac{\beta-\gamma}{2}}-1}} \\
&= \frac{a^{\frac{a}{a-1}}}{a-1}\left(n_L^{\frac{\gamma-\beta}{2}}-a^{-1}\right)\, a^{\frac{n_L^{\frac{\beta-\gamma}{2}}}{n_L^{\frac{\beta-\gamma}{2}}-a}}\,.
\end{aligned}
$$

Moreover for $n_L \geqslant R^2$ the last multiplier can be bounded as

$$
a^{\frac{n_L^{\frac{\beta-\gamma}{2}}}{n_L^{\frac{\beta-\gamma}{2}}-a}} \leqslant \max\left\{a^2, a^{-1}\right\}
$$

and thus

$$
f\left(\bar{L}-1\right) \leqslant \max\left\{a^2, a^{-1}\right\} \frac{a^{\frac{a}{a-1}}}{a-1}\left(n_L^{\frac{\gamma-\beta}{2}}-a^{-1}\right)\,.
$$

Hence,

$$
f\left(L\right) \leqslant \max\left\{a^2, a^{-1}\right\}\ \frac{a^{\frac{a}{a-1}}}{a-1}\ \max\left\{n_L^{\frac{\gamma-\beta}{2}}-a^{-1},\ n_L^{\frac{\gamma-\beta}{2}}-a\right\}
$$

Then according Equation 68 and Equation 69

$$
\begin{aligned}
\mathrm{C.C.}\left[\hat{E}\right] &\leqslant 32\varepsilon^{-2} c_c c_v 2^{\gamma+\beta}\left(f\left(L\right)\right)^2 \\
&\leqslant \tilde{C}\left(\max\left\{n_L^{\frac{\gamma-\beta}{2}}-a^{-1},\ n_L^{\frac{\gamma-\beta}{2}}-a\right\}\right)^2 \varepsilon^{-2}\,.
\end{aligned}
$$

If $a > 1$ for $n_L \geqslant R^2$

$$
\begin{aligned}
\mathrm{C.C.}\left[\hat{E}\right] &\leqslant \tilde{C}\left(a-n_L^{\frac{\gamma-\beta}{2}}\right)^2 \varepsilon^{-2} \\
&\leqslant C\,\varepsilon^{-2}\,,
\end{aligned}
$$

what proves the assertion in the case of $\beta > \gamma$. On the other hand in the case of $a < 1$ for $n_L \geqslant R^2$

$$
\begin{aligned}
\mathrm{C.C.}\left[\hat{E}\right] &\leqslant \tilde{C}\left(n_L^{\frac{\gamma-\beta}{2}}-a^{-1}\right)^2 \varepsilon^{-2} \\
&\leqslant C\, n_L^{\gamma-\beta}\varepsilon^{-2}\,,
\end{aligned}
$$

what proves the assertion in the case of $\beta < \gamma$.

□

JUSTIFICATION OF THE REMARK ON THE CONSTANT C

Here we justify the expression of the constant C given in Remark 15. To simplify consideration we assume that the choices of n_i, N_i and L coincide with their continuous optimal counterparts. First of all note that

$$\left(n_i^{-\beta/2} - n_{i-1}^{-\beta/2}\right)^2 \leqslant \mathbb{V}\mathrm{ar}\left(\hat{P}^{n_i} - \hat{P}^{n_{i-1}}\right) \leqslant \left(n_i^{-\beta/2} - n_{i-1}^{-\beta/2}\right)^2 .$$

The lower bound corresponds to the case of perfect negative correlation and upper to perfect positive correlation. For the later we assume that variance is equal to the geometric average of these bounds

$$\begin{aligned}\mathbb{V}\mathrm{ar}\left(\hat{P}^{n_i} - \hat{P}^{n_{i-1}}\right) &\approx \sqrt{\left(n_i^{-\beta/2} - n_{i-1}^{-\beta/2}\right)^2 \left(n_i^{-\beta/2} + n_{i-1}^{-\beta/2}\right)^2} \\ &= n_{i-1}^{-\beta} - n_i^{-\beta} .\end{aligned}$$

In particular,

$$\begin{aligned}\mathrm{C.C.}\left[\hat{E}\right] &= \varepsilon^{-2}\left(\sum_{i=0}^{L}\sqrt{C_i V_i}\right)^2 \\ &\leqslant \varepsilon^{-2} c_c c_v \left(n_0^{\gamma/2} + \sum_{i=1}^{L}\sqrt{\left(n_{i-1}^{\gamma} + n_i^{\gamma}\right)\left(n_{i-1}^{-\beta} - n_i^{-\beta}\right)}\right)^2 .\end{aligned}$$

Then using the optimal $n_i = R^{-(L-i)} n_L$ from (32) we derive

$$\begin{aligned}&\sum_{i=1}^{L}\sqrt{\left(n_{i-1}^{\gamma} + n_i^{\gamma}\right)\left(n_{i-1}^{-\beta} - n_i^{-\beta}\right)} = \\ &= \sqrt{(1+R^{-\gamma})(R^{\beta}-1)}\sum_{i=1}^{L} n_i^{\frac{\gamma-\beta}{2}} \\ &= \sqrt{(1+R^{-\gamma})(R^{\beta}-1)} n_L^{\frac{\gamma-\beta}{2}} \sum_{i=1}^{L}\left(R^{\frac{\gamma-\beta}{2}}\right)^{-L+i} \\ &= \sqrt{(1+R^{-\gamma})(R^{\beta}-1)} n_L^{\frac{\gamma-\beta}{2}} \sum_{i=1}^{L}\left(R^{\frac{\gamma-\beta}{2}}\right)^{-L+i}\end{aligned}$$

In the following we distinguish between two cases: $\gamma = \beta$ and $\gamma \neq \beta$.

CASE OF $\gamma = \beta$

$$\begin{aligned}&\sum_{i=1}^{L}\sqrt{\left(n_{i-1}^{\gamma} + n_i^{\gamma}\right)\left(n_{i-1}^{-\beta} - n_i^{-\beta}\right)} = L\sqrt{(1+R^{-\gamma})(R^{\beta}-1)} \\ &= L\sqrt{R^{\gamma} - R^{-\gamma}}\end{aligned}$$

Keeping in mind that $R = e^{2/\gamma}$, $n_0 = R$ and $L = \log_{[R]}(n_L) - 1$ we can obtain

$$\begin{aligned}
CC\left[\hat{E}\right] &\leqslant \varepsilon^{-2}\, c_c\, c_v \left(R^{\gamma/2} + \sqrt{R^{\gamma} - R^{-\gamma}} \cdot L\right)^2 \\
&= \varepsilon^{-2}\, c_c\, c_v \left(e + \sqrt{e^2 - e^{-2}} \cdot L\right)^2 \\
&= \varepsilon^{-2}\, c_c\, c_v \,\left(e^2 - e^{-2}\right) \left(\log_{[R]}(n_L) + \sqrt{\frac{e^2}{e^2 - e^{-2}}} - 1\right)^2 \\
&= \varepsilon^{-2}\, c_c\, c_v \,\left(e^2 - e^{-2}\right) \left(\ln\left(n_L^{\frac{\gamma}{2}}\right) + \sqrt{\frac{e^2}{e^2 - e^{-2}}} - 1\right)^2
\end{aligned}$$

CASE OF $\gamma \neq \beta$

$$\begin{aligned}
&\sum_{i=1}^{L} \sqrt{\left(n_{i-1}^{\gamma} + n_i^{\gamma}\right)\left(n_{i-1}^{-\beta} - n_i^{-\beta}\right)} = \\
&= \sqrt{\left(1 + R^{-\gamma}\right)\left(R^{\beta} - 1\right)}\, n_L^{\frac{\gamma-\beta}{2}} \sum_{i=1}^{L} \left(R^{\frac{\gamma-\beta}{2}}\right)^{-L+i} \\
&= \sqrt{\left(1 + R^{-\gamma}\right)\left(R^{\beta} - 1\right)}\, n_L^{\frac{\gamma-\beta}{2}} \sum_{i=0}^{L-1} \left(R^{\frac{\beta-\gamma}{2}}\right)^{i} \\
&= \sqrt{\left(1 + R^{-\gamma}\right)\left(R^{\beta} - 1\right)}\, n_L^{\frac{\gamma-\beta}{2}} \frac{\left(R^{\frac{\beta-\gamma}{2}}\right)^{L} - 1}{R^{\frac{\beta-\gamma}{2}} - 1}
\end{aligned}$$

Keeping in mind that $a = \beta/\gamma$, $R = \left(\frac{\gamma}{\beta}\right)^{\frac{2}{\gamma-\beta}}$, $n_0 = R$ and $L = \log_{[R]}(n_L) - 1$ we can derive

$$\begin{aligned}
&\sum_{i=1}^{L} \sqrt{\left(n_{i-1}^{\gamma} + n_i^{\gamma}\right)\left(n_{i-1}^{-\beta} - n_i^{-\beta}\right)} = \\
&= \sqrt{\left(1 + R^{-\gamma}\right)\left(R^{\beta} - 1\right)}\, n_L^{\frac{\gamma-\beta}{2}} \frac{a^L - 1}{a - 1} \\
&= \frac{\sqrt{\left(1 + R^{-\gamma}\right)\left(R^{\beta} - 1\right)}}{a - 1} n_L^{\frac{\gamma-\beta}{2}} \left(a^{-1}\, n_L^{\frac{\beta-\gamma}{2}} - 1\right) \\
&= \frac{\sqrt{\left(1 + R^{-\gamma}\right)\left(R^{\beta} - 1\right)}}{a - 1} \left(a^{-1} - n_L^{\frac{\gamma-\beta}{2}}\right)
\end{aligned}$$

Then putting this into the expression for CC we obtain

$$
\begin{aligned}
\mathrm{C.C.}\left[\hat{E}\right] &\leqslant \varepsilon^{-2}\, c_c\, c_\nu \left(n_0^{\gamma/2} + \sum_{i=1}^{L} \sqrt{\left(n_{i-1}^{\gamma} + n_i^{\gamma}\right)\left(n_{i-1}^{-\beta} - n_i^{-\beta}\right)} \right)^2 \\
&\leqslant \varepsilon^{-2}\, c_c\, c_\nu \left(R^{\gamma/2} + \frac{\sqrt{(1+R^{-\gamma})\,(R^{\beta}-1)}}{a-1} \cdot \left(a^{-1} - n_L^{\frac{\gamma-\beta}{2}} \right) \right)^2 \\
&\leqslant \varepsilon^{-2}\, c_c\, c_\nu\, \frac{(1+R^{-\gamma})\left(R^{\beta}-1\right)}{(a-1)^2} \\
&\quad \times \left(n_L^{\frac{\gamma-\beta}{2}} - a^{-1} + (a-1)\, \sqrt{\frac{R^{\gamma}}{(1+R^{-\gamma})\,(R^{\beta}-1)}} \right)^2 .
\end{aligned}
$$

B

WEAK EXTRAPOLATION AND MLMC APPENDIX

B.1 SOLVING SYSTEM OF EQUATIONS λ^n

Here we show how the system of equations (49) can be solved.

Remark 29. *In the classic HOWE settings of Section 3.1.2, the corresponding system of equations is solved [59] using properties of Vandermonde matrices [54].*

To solve this problem we consider it as a problem of an exact polynomial fit. We consider the polynomial

$$p(x) = \lambda_0^N + \lambda_1^N x^1 + \lambda_2^N x^2 + \ldots + \lambda_N^N x^N = \sum_{i=0}^{N} \lambda_i^N x^i \tag{71}$$

and the problem of its exact fit to the set of points

$$\left\{ (1,1) ; \left(\frac{1}{M^{1\alpha}}, 0 \right) ; \left(\frac{1}{M^{2\alpha}}, 0 \right) ; \ldots ; \left(\frac{1}{M^{N\alpha}}, 0 \right) \right\} .$$

Then by construction the coefficients λ_i^N of the fitted polynomial and the solution of the the system of equations (49) coincide.

The fitting problem can be solved using the Lagrange's interpolation formula [18], what yields

$$p(x) = \frac{\left(x - \frac{1}{M^{1\alpha}}\right)\left(x - \frac{1}{M^{2\alpha}}\right) \ldots \left(x - \frac{1}{M^{N\alpha}}\right)}{\left(1 - \frac{1}{M^{1\alpha}}\right)\left(1 - \frac{1}{M^{2\alpha}}\right) \ldots \left(1 - \frac{1}{M^{N\alpha}}\right)} ,$$

and after elementary transformation an equivalent form

$$p(x) = \frac{\left(M^{1\alpha}x - 1\right)\left(M^{2\alpha}x - 1\right) \ldots \left(M^{N\alpha}x - 1\right)}{\left(M^{1\alpha} - 1\right)\left(M^{2\alpha} - 1\right) \ldots \left(M^{N\alpha} - 1\right)} . \tag{72}$$

From representation (72) we directly obtain

$$\begin{aligned} \lambda_0^N &= \frac{(-1)(-1)\ldots(-1)}{\left(M^{1\alpha} - 1\right)\left(M^{2\alpha} - 1\right) \ \ldots \left(M^{N\alpha} - 1\right)} \\ &= (-1)^N \prod_{i=1}^{N} \frac{1}{M^{i\alpha} - 1} \end{aligned}$$

and

$$\begin{aligned} \lambda_N^N &= \frac{M^{1\alpha} M^{2\alpha} \ldots M^{N\alpha}}{\left(M^{1\alpha} - 1\right)\left(M^{2\alpha} - 1\right) \ldots \left(M^{N\alpha} - 1\right)} \\ &= \prod_{i=1}^{N} \frac{M^{i\alpha}}{M^{i\alpha} - 1} . \end{aligned}$$

Further collecting coefficients with all $(N-1)^{\text{th}}$ degree x-terms in Equation 72 we derive λ^N_{N-1} as

$$\begin{aligned}\lambda_{N-1} &= \frac{1}{\prod_{i=1}^N (M^{i\alpha}-1)} \sum_{i=1}^N \left(-\prod_{\substack{j=1\\ j\neq i}}^N M^{j\alpha} \right) \\ &= \frac{1}{\prod_{i=1}^N (M^{i\alpha}-1)} \left(-\prod_{j=1}^N M^{j\alpha} \right) \sum_{i=1}^N \frac{1}{M^{i\alpha}} \\ &= -\frac{1}{M^{\alpha}-1} \prod_{i=1}^{N-1} \frac{M^{i\alpha}}{M^{i\alpha}-1} .\end{aligned}$$

This suggests that the general form of the extrapolation coefficient reads as

$$\lambda_{N-j} = (-1)^j \prod_{i=1}^{j} \frac{1}{M^{i\alpha}-1} \prod_{i=1}^{N-j} \frac{M^{i\alpha}}{M^{i\alpha}-1} . \tag{73}$$

However this requires proof of the correctness this solution. It can be given using a formula for the determinant of a Vandermonde matrix, for details see [51].

B.2 BIAS OF THE HOWE SCHEME

Proof of Proposition 3 on the bias of the HOWE scheme. Representation (46) and Equation 47 imply that the first non-vanishing term of the error has the following form

$$\frac{c_{N+1}}{n^{(N+1)\alpha}} \sum_{i=0}^N \lambda_i^n \, M^{-i(N+1)\alpha} = \frac{c_{N+1}}{n^{(N+1)\alpha}} \, p\left(M^{-(N+1)\alpha}\right) ,$$

where p is the polynomial defined in Section B.1. Then using representation (72) of the polynomial p we derive the first term of the bias error as

$$\begin{aligned}\frac{c_{N+1}}{n^{(N+1)\alpha}} \sum_{i=0}^N \lambda_i^n \, M^{-i(N+1)\alpha} &= \frac{c_{N+1}}{n^{(N+1)\alpha}} (-1)^N \prod_{i=1}^N \frac{1}{M^{i\alpha}} \\ &= (-1)^N \, c_{N+1} \frac{1}{n^{(N+1)\alpha} M^{\frac{(N+1)N}{2}\alpha}} .\end{aligned}$$

It remains to show that the remaining terms are of o-small of the above expression. In particular using representation (46) and representation (72) of the polynomial p we show that an absolute value of the this term can be bounded as

$$\begin{aligned}\left|\frac{c_{N+2}}{n^{(N+2)\alpha}}\sum_{i=0}^{N}\lambda_i^n\, M^{-i(N+2)\alpha}\right| &= \left|\frac{c_{N+2}}{n^{(N+2)\alpha}}\, p\left(M^{-(N+2)\alpha}\right)\right| \\ &= \left|\frac{c_{N+2}}{n^{(N+2)\alpha}}(-1)^N\frac{M^{(N+1)\alpha}-1}{M^{\alpha}-1}\prod_{i=2}^{N+1}\frac{1}{M^{i\alpha}}\right| \\ &= |c_{N+2}|\,\frac{1}{n^{(N+2)\alpha}}\frac{M^{(N+1)\alpha}-1}{M^{\alpha}-1}\frac{1}{M^{\frac{(N+3)N}{2}\alpha}} \\ &< |c_{N+2}|\ \frac{1}{n^{(N+2)\alpha}M^{\frac{(N+1)N}{2}\alpha}}\ \frac{M^{\alpha}}{M^{\alpha}-1}\ .\end{aligned}$$

Hence,

$$\frac{c_{N+2}}{n^{(N+2)\alpha}}\sum_{i=0}^{N}\lambda_i^n\, M^{-i(N+2)\alpha} = o\left(\frac{1}{n^{(N+1)\alpha}M^{\frac{(N+1)N}{2}\alpha}}\right)\ ,$$

as $n\to\infty$, what yields the assertion. □

B.3 BOUNDEDNESS OF THE WEIGHTS κ^N

Proof of Proposition 4 on the boundedness of the weights κ^N. First note that

$$\frac{1}{M^{(N-i)\alpha}-1}\left|\lambda_{i+1}^N\right| = \frac{M^{(i+1)\alpha}}{M^{(i+1)\alpha}-1}\left|\lambda_i^N\right| \qquad \text{for } i=0,1,\dots,N-1\ .$$

Hence for M such that $M^{\alpha}\geqslant 2$ holds

$$\left|\lambda_i^N\right| < \left|\lambda_{i+1}^N\right| \qquad \text{for } i=0,1,\dots,N-1\ , \tag{74}$$

i.e. for a fixed N absolute values of λ_i^N monotonously increase with i.

Further we prove that $\lambda_N^N\to C<\infty$ as $N\to\infty$. Note,

$$\ln\left(\lambda_N^N\right) = -\sum_{i=1}^{N}\ln\left(1-\frac{1}{M^{i\alpha}}\right)\ .$$

Then

$$\ln(1-x) = -x + o(-x) \qquad \text{and} \qquad M^{-i\alpha}\to 0$$

imply that

$$\exists N^*\in\mathbb{N}\quad \text{s.t. } \forall i>N^*\quad \ln\left(1-\frac{1}{M^{i\alpha}}\right)\geqslant -\frac{1}{M^{i\alpha}}-\frac{1}{M^{i\alpha}} = -2\frac{1}{M^{i\alpha}}\ .$$

Hence,

$$\ln\left(\lambda_N^N\right)\leqslant -\sum_{i=1}^{N^*}\ln\left(1-\frac{1}{M^{i\alpha}}\right)+2\sum_{i=N^*}^{N\wedge N^*}\frac{1}{M^{i\alpha}}$$

and, thus $\forall N \in \mathbb{N}$

$$\lambda_N^N \leqslant e^{-\sum_{i=1}^{N^*} \ln\left(1-\frac{1}{M^{i\alpha}}\right)} e^{2\sum_{i=1}^{N} \frac{1}{M^{i\alpha}}} \leqslant e^{-\sum_{i=1}^{N^*} \ln\left(1-\frac{1}{M^{i\alpha}}\right)} e^{\frac{2}{1-M^{-\alpha}}} \ .$$

Then, as $\lambda_N^N < \lambda_L^L$ for $N < L$, the monotone convergence theorem implies that $\lambda_N^N \to C < \infty$ as $N \to \infty$.

Now let us prove the main assertion of the proposition, i.e.

$$0 < \kappa_i^N < \kappa_N^N \qquad \text{for } i = 0, 1, \dots, N-1 \ .$$

Note that $\kappa_N^N = \lambda_N^N > 0$ and $\kappa_0^N = 1$ by definition. Alternating signs of λ_i^N and inequality (74) imply that

$$\lambda_{N-2j}^N + \lambda_{N-2j+1}^N < 0 \qquad \text{and} \qquad \lambda_{N-2j-1}^N + \lambda_{N-2j}^N > 0$$

for $j = 1, 2, \dots, \lceil N/2 \rceil$ and $j = 0, 1, \dots, \lfloor N/2 \rfloor$ correspondingly. Then representing κ_{N-i}^N as

$$\kappa_{N-i}^N = \begin{cases} \sum_{j=1}^{i/2} \underbrace{\left(\lambda_{N-2j}^N + \lambda_{N-2j+1}^N\right)}_{<0} + \lambda_N^N & \text{for even } i \ , \\ \underbrace{\lambda_{N-i}^N}_{<0} + \sum_{j=1}^{\lfloor i/2 \rfloor} \underbrace{\left(\lambda_{N-2j}^N + \lambda_{N-2j+1}^N\right)}_{<0} + \lambda_N^N & \text{for odd } i \ , \end{cases}$$

we obtain that

$$\kappa_{N-i}^N < \kappa_N^N = \lambda_N^N \qquad \text{for } i = 1, 2, \dots, N-1 \ .$$

On the other hand κ_{N-i}^N can be represented as

$$\kappa_{N-i}^N = \begin{cases} \underbrace{\lambda_{N-i}^N}_{>0} + \sum_{j=0}^{i/2-1} \underbrace{\left(\lambda_{N-2j-1}^N + \lambda_{N-2j}^N\right)}_{>0} & \text{for even } i \ , \\ \sum_{j=0}^{\lfloor i/2 \rfloor} \underbrace{\left(\lambda_{N-2j-1}^N + \lambda_{N-2j}^N\right)}_{>0} & \text{for odd } i \ , \end{cases}$$

thus

$$\kappa_{N-i}^N > 0 \qquad \text{for } i = 1, 2, \dots, N-1 \ .$$

□

[1] Mohamed Ben Alaya, Ahmed Kebaier, et al. Central limit theorem for the multilevel monte carlo euler method and applications to asian options. *HAL*, 2012. URL http://hal.archives-ouvertes.fr/docs/00/69/31/91/PDF/CLT_Multilevel.pdf.

[2] Aurélien Alfonsi. Strong convergence of some drift implicit Euler scheme. Application to the CIR process. *STATISTICS & PROBABILITY LETTERS*, 83(2):602–607, 2013. URL http://hal-enpc.archives-ouvertes.fr/hal-00709202.

[3] Martin Altmayer and Andreas Neuenkirch. Multilevel monte carlo quadrature of discontinuous payoffs in the generalized heston model using malliavin integration by parts. *Preprint*, 144, 2013.

[4] Leif Andersen. Simple and efficient simulation of the heston stochastic volatility model. *Journal of Computational Finance*, 11 (3), 2008.

[5] Leif B.G. Andersen and Vladimir V. Piterbarg. Moment explosions in stochastic volatility models. *Finance and Stochastics*, 11 (1):29–50, 2007. ISSN 0949-2984.

[6] Vlad Bally and Denis Talay. The law of the euler scheme for stochastic differential equations: Ii. convergence rate of the density. *Rapport De Recherche-Institut National De Recherche En Informatique Et En Automatique*, 1995.

[7] Vlad Bally and Denis Talay. The law of the euler scheme for stochastic differential equations. *Probability theory and related fields*, 104(1):43–60, 1996.

[8] D. Bates. Jump and stochastic volatility: Exchange rate processes implicit in deutsche mark options. *Review of Financial Studies*, 9: 69–107, 1996.

[9] Jon Louis Bentley and Andrew Chi-Chih Yao. An almost optimal algorithm for unbounded searching. *Information Processing Letters*, 5(3):82 – 87, 1976. ISSN 0020-0190.

[10] Fischer Black and Myron S Scholes. The Pricing of Options and Corporate Liabilities. *Journal of Political Economy*, 81(3):637–54, May-June 1973.

[11] Stephen Boyd, Seung-Jean Kim, Lieven Vandenberghe, and Arash Hassibi. A tutorial on geometric programming. *Optimization and engineering*, 8(1):67–127, 2007.

[12] Mark Broadie and Özgür Kaya. Exact simulation of stochastic volatility and other affine jump diffusion processes. *Operations Research*, 54(2):217–231, 2006.

[13] Mark Broadie, Paul Glasserman, and Steven Kou. A continuity correction for discrete barrier options. *Mathematical Finance*, 7(4): 325–349, 1997.

[14] E.K.P. Chong and S.H. Zak. *An Introduction to Optimization*. Wiley Series in Discrete Mathematics and Optimization. Wiley, 2011. ISBN 9781118211601.

[15] I.J. Clark. *Foreign Exchange Option Pricing: A Practitioners Guide*. The Wiley Finance Series. Wiley, 2011. ISBN 9780470683682.

[16] R. Cont. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1(2):223–236, 2001.

[17] J. C. Cox, J.E. Ingersoll, and S.A. Ross. A theory of the term structure of interest rates. *Econometrica*, 53:385–407, 1985.

[18] G. Dahlquist and Å. Björck. *Numerical Methods in Scientific Computing: Volume 1*. Number v. 1 in SIAM e-books. Society for Industrial and Applied Mathematics, 2008. ISBN 9780898717785.

[19] K. Debrabant and A. Rößler. On the Acceleration of the Multi-Level Monte Carlo Method. *ArXiv e-prints*, January 2013. URL http://arxiv.org/abs/1301.7650.

[20] F. Delbaen and W. Schachermayer. *The Mathematics of Arbitrage*. Springer Finance. Physica-Verlag, 2006. ISBN 9783540312994.

[21] Steffen Dereich, Andreas Neuenkirch, and Lukasz Szpruch. An euler-type method for the strong approximation of the cox-ingersoll-ross process. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, 468(2140):1105–1115, 2012.

[22] Darrell Duffie and Peter Glynn. Efficient monte carlo simulation of security prices. *The Annals of Applied Probability*, 5(4):897–905, 11 1995.

[23] Bruno Dupire. Pricing with a smile. *Risk*, 1994.

[24] Fang Fang and Cornelis W Oosterlee. A novel pricing method for european options based on fourier-cosine series expansions. *SIAM Journal on Scientific Computing*, 31(2):826–848, 2008.

[25] William Feller. Two singular diffusion problems. *The Annals of Mathematics*, 54(1):173–182, 1951.

[26] J. Gatheral and N.N. Taleb. *The Volatility Surface: A Practitioner's Guide*. Wiley Finance. Wiley, 2006. ISBN 9780470068250.

[27] Michael B Giles and Lukasz Szpruch. Antithetic multi-level monte carlo estimation for multidimensional sdes. In *Monte Carlo and Quasi-Monte Carlo Methods 2012*, pages 367–384. Springer, 2013.

[28] Mike B. Giles. Multi-level monte carlo path simulation. *Operations Research*, 56(3):607–617, 2008.

[29] Paul Glasserman. *Monte Carlo Methods in Financial Engineering*. Springer, 2004.

[30] Emmanuel Gobet. Advanced monte carlo methods for barrier and related exotic options. *Mathematical modeling and numerical methods in finance*, pages 497–528, 2009.

[31] Patrick Hagan, Deep Kumar, Andrew Lesniewski, and Diana Woodward. Managing smile risk. *WILMOTT Magazine*, 2002.

[32] E.G. Haug. *The complete guide to option pricing formulas*. McGraw-Hill, 2007.

[33] S. Heinrich. Monte carlo complexity of global solution of integral equations. *Journal of Complexity*, 14(2):151 – 175, 1998. ISSN 0885-064X.

[34] Pierre Henry-Labordere. Calibration of local stochastic volatility models to market smiles: A monte-carlo approach. *Risk Magazine*, 2009.

[35] Steven L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6:327–343, 1993.

[36] Jean Jacod and Philip Protter. Asymptotic error distributions for the euler method for stochastic differential equations. *The Annals of Probability*, 26(1):267–307, 01 1998.

[37] I. Karatzas and S.E. Shreve. *Brownian Motion and Stochastic Calculus*. Graduate Texts in Mathematics. Springer New York, 1991. ISBN 9780387976556.

[38] Ahmed Kebaier. Statistical romberg extrapolation: A new variance reduction method and applications to option pricing. *The Annals of Applied Probability*, 15(4):2681–2705, 2005.

[39] P.E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Applications of mathematics : stochastic modelling and applied probability. Springer, 1992. ISBN 9783540540625.

[40] D.E. Knuth. *The Art of Computer Programming: Sorting and searching*. Addison-Wesley series in computer science and information processing. Addison-Wesley Pub. Co., 1973.

[41] W. Koepf. *Hypergeometric summation: an algorithmic approach to summation and special function identities*. Advanced lectures in mathematics. Vieweg, 1998. ISBN 9783528069506.

[42] R. Korn and E. Korn. *Option Pricing and Portfolio Optimization: Modern Methods of Financial Mathematics*. Crm Proceedings & Lecture Notes. American Mathematical Society, 2001. ISBN 9780821821237.

[43] Ralf Korn and Stefanie Müller. Getting multi-dimensional trees into a new shape. *Wilmott Journal*, 1(3):145–153, 2009. ISSN 1759-636X.

[44] Ralf Korn and Serkan Zeytun. Efficient basket monte carlo option pricing via a simple analytical approximation. *Journal of Computational and Applied Mathematics*, 243(0):48 – 59, 2013. ISSN 0377-0427.

[45] Ralf Korn, Elke Korn, and Gerald Kroisandt. *Monte Carlo Methods and Models in Finance and Insurance*. Chapman & Hall/CRC, 2010.

[46] Thomas G Kurtz and Philip Protter. Weak limit theorems for stochastic integrals and stochastic differential equations. *The Annals of Probability*, pages 1035–1070, 1991.

[47] Thomas G Kurtz and Philip Protter. Characterizing the weak convergence of stochastic integrals. *Stochastic Analysis*, 167:255–259, 1991.

[48] Thomas G Kurtz and Philip Protter. Wong-zakai corrections, random evolutions, and simulation schemes for sdes. *Stochastic analysis*, 331:346, 1991.

[49] Herman Lam and Gregg Cooke. Finrc: Challenges and opportunities for high-performance reconfigurable computing (hprc) in computational finance. In *Proceedings of the Fourth Workshop on High Performance Computational Finance*, WHPCF '11, pages 21–22, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-1108-3.

[50] Bernard Lapeyre and Emmanuel Temam. Competitive monte carlo methods for the pricing of asian options. *Journal of computational finance*, 5(1):39–58, 2001.

[51] Vincent Lemaire and Gilles Pagès. Multilevel Richardson-Romberg extrapolation. Technical Report LPMA_1603, Laboratoire de Probabilités et Modèles Aléatoires, 2013. URL http://hal.archives-ouvertes.fr/hal-00920660. 33 p.

[52] Roger Lord and Christian Kahl. Optimal fourier inversion in semi-analytical option pricing. Technical report, Tinbergen Institute Discussion Paper, 2006.

[53] Roger Lord, Remmert Koekkoek, and Dick Van Dijk. A comparison of biased simulation schemes for stochastic volatility models. *Quantitative Finance*, 10(2):177–194, 2010.

[54] N Macon and A Spitzbart. Inverses of vandermonde matrices. *American Mathematical Monthly*, pages 95–100, 1958.

[55] Henning Marxen. *Aspects of the Application of Multilevel Monte Carlo Methods in the Heston Model and in a Lévy Process Framework*. PhD thesis, University of Kaiserslautern, 2012.

[56] Henning Marxen, Anton Kostiuk, Ralf Korn, Christian de Schryver, Stephan Wurm, Ivan Shcherbakov, and Norbert Wehn. Algorithmic complexity in the heston model: An implementation view. In *Proceedings of the Fourth Workshop on High Performance Computational Finance*, WHPCF '11, pages 5–12, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-1108-3.

[57] Robert C. Merton. Theory of Rational Option Pricing. *Bell Journal of Economics*, 4(1):141–183, Spring 1973.

[58] Thomas Müller-Gronbach, Erich Novak, and Klaus Ritter. Algorithmen, fehler und kosten. In *Monte Carlo-Algorithmen*, pages 13–26. Springer, 2012.

[59] G. Pagès. Multi-step Richardson-Romberg Extrapolation: Remarks on Variance Control and Complexity. *Monte Carlo Methods and Applications*, 13(1):37–70, 2007.

[60] Vladimir Piterbarg. Markovian projection for volatility calibration. *Risk*, 2007.

[61] Mark Rubinstein and Eric Reiner. Breaking down the barriers. *Risk*, 4(8):28–35, 1991.

[62] P.A. Samuelson. Proof that properly anticipated prices fluctuate randomly. *Industrial management review*, 1965.

[63] D. Talay and L. Tubaro. Expansion of the global error for numerical schemes solving stochastic differential equations. *Stochastic Analysis and Applications*, 8(4):483–509, 1990.

DECLARATION OF ORIGINALITY

I hereby declare that this PhD thesis is my own work, except where stated explicitly in the text or references.

To the best of my knowledge, the results stated in this work were not known or considered previously if not indicated otherwise.

This thesis has never been submitted or published anywhere else before.

Kaiserslautern, December 9, 2014

Anton Kostiuk

SCIENTIFIC CAREER

1994-2004	School-leaving Certificate, Chernihiv Lyceum # 15.
2004-2008	Bachelor Degree in Applied Mathematics, NTUU "Kiev Polytechnic Institute", ESC "Institute for Applied System Analysis".
2008-2010	Master of Science in Financial Mathematics, University of Kaiserslautern, Department of Mathematics.
since 2011	Ph.D. student, University of Kaiserslautern, Department of Mathematics.

WISSENSCHAFTLICHER WERDEGANG

1994-2004	Abitur, Tschernigiwer Lyzeum # 15.
2004-2008	Bakkalaureus in Angewandte Mathematik, NTUU "Kiewer Polytechnisches Institut", LWK "Institut für angewandte Systemanalyse".
2008-2010	Master of Science in Finanzmathematik, Technische Universität Kaiserslautern, Fachbereiche Mathematik.
seit 2011	Promotionsstudent , Technische Universität Kaiserslautern, Fachbereiche Mathematik.