Electronic Thesis and Dissertation Repository

10-28-2010 12:00 AM

Simulation-based Valuation and Counterparty Exposure Estimation of American Options

Kin Hung Kan, *The University of Western Ontario*

Supervisor: Mark Reesor, *The University of Western Ontario*
A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Applied Mathematics
© Kin Hung Kan 2010

Follow this and additional works at: https://ir.lib.uwo.ca/etd

🌀 Part of the Management Sciences and Quantitative Methods Commons

SIMULATION-BASED VALUATION AND COUNTERPARTY EXPOSURE
ESTIMATION OF AMERICAN OPTIONS

(Spine title: Simulation and American Options)

(Thesis format: Integrated-Article)

by

Kin Hung (Felix) <u>Kan</u>

Graduate Program
in
Applied Mathematics

A thesis submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

THE UNIVERSITY OF WESTERN ONTARIO
School of Graduate and Postdoctoral Studies

CERTIFICATE OF EXAMINATION

Supervisor

Examiners

_____

Mark Reesor

Supervisory Committee

_____

Matt Davison

_____

Adam Metzler

_____

Matt Davison

_____

Colin Denniston

_____

Hao Yu

_____

Lars Stentoft

The thesis by

KIN HUNG (FELIX) KAN

entitled:

SIMULATION-BASED VALUATION AND COUNTERPARTY EXPOSURE ESTIMATION OF AMERICAN OPTIONS

is accepted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Date _____

_____
Chair of the Thesis Examination Board

Abstract and Keywords

Valuing American options is a central problem in option pricing since the early- exercise feature is very common among financial or insurance derivatives products. For high-dimensional American options, Monte Carlo simulation is generally regarded as the only viable approach to price them, and this is the focus of our work. We propose a new regression-based Monte Carlo algorithm for pricing American options. This method typically generates an upper bound of the option value. It is computationally efficient and generates accurate price estimates.

To improve the convergence rate, we apply a bias reduction technique to the least-squares Monte Carlo estimators of American option value. It works by subtracting a bias approximation from the original option value estimators at each exercise opportunity. The bias approximation is derived using large sample properties of the least-squares regression estimators. The resulting expression is easy to evaluate, and is applicable to any payoff structures and underlying processes. Numerical results show that this technique can significantly reduce the bias. However, it introduces non-negligible computational costs, thus careful treatment is required when it is adopted in practice.

Finally, we extend the least-squares Monte Carlo algorithm to estimate the counterparty exposures of American options. The new algorithm is termed optimized least-squares Monte Carlo (OLSM), which is combined with variance reduction techniques, initial state dispersion and multiple bucketing to enhance its performance. The biggest advantage of OLSM is that it avoids nested simulations, allowing for the computation of risk measures on various time

horizons under a reasonable computational budget.

iv

Authorship

All the work presented in this thesis is my own. Although I benefited from the comments of my supervisor Dr. Mark Reesor, and my colleagues at CIBC, Greg Frank and Victor Mozgin, I derived all the mathematical proofs and implemented all the algorithms.

The content of Chapter 2 is submitted to The Journal of Computational Finance. Partial results in Chapter 3 are published in Monte Carlo and Quasi-Monte Carlo Methods 2008, and much of the content of this chapter forms another independent paper which is submitted to Applied Mathematical Finance. The content of Chapter 4 is published in Mathematics-in-Industry Case Studies Journal. Mark Reesor is my co-author for the manuscript of Chapter 2. The published manuscript based on Chapters 3 is co-authored with Mark Reesor, Tyson Whitehead and Matt Davison, and the unpublished one is co-authored with Mark Reesor only. The manuscript of Chapter 4 is co-authored with Greg Frank, Victor Mozgin and Mark Reesor.

He's always hands-on; he's so good at

everything he does!

_____

A colleague of Michael Jackson

To my parents: Shuen Hei Kan and Chun Yuk Choi

and my wife: Hui Ying (Lucy) Chen

Acknowledgements

Contents

List of Tables

List of Figures

Chapter 1

Introduction

We begin the introduction with an overview of the major contributions. All the work in this thesis is centered around American options. In this thesis, we

- develop a new, efficient regression-based Monte Carlo algorithm which typically generates an upper bound for the true value of an American option;

- provide the proofs for the properties of the option value estimator generated by the new algorithm;

- extend a recently developed bias reduction technique to the least-squares Monte Carlo (LSM) estimators;

- use numerical results to show that the bias reduction method works effectively for the low-biased LSM estimator, but not so well for the high-biased LSM estimator;

- design an optimized least-squares Monte Carlo (OLSM) method that avoids the use of nested simulations for estimating counterparty exposures of American options;

- demonstrate that OLSM can generate reasonably accurate exposure estimates.

An American option is a contract that allows the holder to exercise at any time before and on the maturity date. This early-exercise feature makes it difficult to derive a fair price for

the American options. As of today, no analytical closed-form solution exists for an American option written on just one underlying asset, although several analytical approximations have been developed. As a result, people often resort to numerical methods, such as finite difference methods, transform methods, and variational inequalities, etc. Unfortunately, all these methods break down for high-dimensional problems, which leaves us no choice but Monte Carlo simulation, since its convergence rate is independent of the dimension.

There is a long history about pricing American options by simulations in the literature. It started with Tilley [16] which dispelled the belief that simulations could not be used for valuing American options. A great deal of subsequent work was inspired by Tilley's pioneering paper. Representative works include Broadie and Glasserman's random tree method [2] and stochastic mesh method [3], the regression-based methods in Carrière [4], Longstaff and Schwartz [13] and Tsitsiklis and Van Roy [17], and the duality approach in Andersen and Broadie [1], Haugh and Kogan [6] and Rogers [15]. In particular, the regression-based methods are the heart of this thesis. It is reasonable that the regression-based methods are at the forefront of research since they are the most efficient among the aforementioned approaches, simple to implement and applicable to many payoff structures and underlying processes.

The value of an American option is determined by the optimal exercise strategy. Owing to this nature, it is relatively easy to generate a lower bound for the option value. One standard approach is to estimate the exercise strategy by cross-sectional regression and approximate dynamic programming on the first set of simulated paths, then simulate the second (independent) set of paths on which exercise decisions are made based on the estimated exercise strategy. Since the estimated exercise strategy can never beat the optimal exercise strategy by definition, it leads to a lower bound for the option value.

On the contrary, the establishment of an upper bound for the option value is much more challenging for the regression-based methods. High-biased estimators can be constructed in the context of random tree or stochastic mesh by using the same continuation value estimator to make exercise decisions and propagate backward. However, the same approach does not

work as well for the regression-based methods as the sign of the regression model error, and hence the direction of bias of the option value estimator, is unknown. As a consequence, there is no theoretical guarantee that the option value estimator generates an upper bound. Nonetheless, we note that it typically generates an upper bound in numerical experiments such as those appear in Glasserman [5]. Independent works in Haugh and Kogan [6] and Rogers [15] overcome the deficiency of the above approach by constructing a true upper bound using a duality approach. By that it means the upper bound is valid no matter what the sign of the regression model error is. The major drawback of the duality approach is that it is computationally demanding to generate a dual upper bound, thus it may not serve well in those situations where speed is critical.

The first goal of this thesis is to develop a computationally efficient algorithm which generates a price estimate that helps evaluate the tightness of the lower bound. Ideally, this price estimate forms a true upper bound so that a valid confidence interval can be constructed for the true option value, but a tight upper bound relative to the true approximation value for a given finite set of basis functions is also informative — the smaller the gap between the upper bound and the lower bound, the more accurately the selected set of basis functions approximates the true continuation value function. We attain the goal by proposing a new regression-based algorithm that is similar to the one in Tsitsiklis and Van Roy [17], but it is mixed with some important elements in the Longstaff-Schwartz algorithm. More explicitly, in the new algorithm, continuation values on the out-of-the-money paths are estimated by the discounted cash flows, while those on the in-the-money (ITM) paths are given by a regression fit that makes use of the ITM paths only. Numerical results demonstrate that these slight modifications can significantly improve the accuracy of the price estimates compared to the Tsitsiklis-Van Roy algorithm. The new algorithm typically generates high-biased estimates that can be used to evaluate the tightness of the lower bound. Furthermore, the new algorithm has a competitive advantage over the duality approach in terms of computational efficiencies. Intuitively, a fast and reasonably accurate algorithm is more sound than a slow but very accurate algorithm in

practical situations. Therefore, the new algorithm is a valuable addition to the literature.

We move on to look at the American option pricing problem from the efficiency point of view. Given a regression-based Monte Carlo estimator, we would like to design a method that makes it converge faster to the true option value. A standard approach for this purpose is to apply variance reduction methods. Interested readers may refer to Juneja and Kalra [7], Lemieux and La [12] and Moreni [14], to name a few. Variance reduction is not the focus of our work, though. Noting that all Monte Carlo estimators of American option value are consistent but biased, we are presented with an opportunity to accelerate the convergence of the estimators by reducing estimator bias.

Whitehead [18] has recently developed a general bias reduction technique in the context of a stochastic tree estimator. We extend his work to the popular LSM estimators. Incorrect exercise decisions lead to bias expressions that cannot be evaluated analytically. To get around this problem, we make use of the large sample properties of the least-squares estimators of the regression coefficients to derive an approximation to the bias. Particularly, a normal distribution is used to approximate the distribution of the estimators of the regression coefficients. The resulting expression is easy to evaluate, and works for virtually any payoff function and underlying process. By recursively subtracting the bias approximation from the uncorrected option value estimator, we obtain a bias-reduced estimator at time zero. We apply the bias reduction method to both high- and low-biased LSM estimators. Numerical results show that the bias-corrected estimators outperform their uncorrected counterpart across all combinations of number of exercise opportunities, option moneyness, and sample size. Although the method can significantly reduce the relative error of either estimator, it does not benefit the high-biased estimator as far as computational speed is concerned. The reason is that it doubles the computing time while it reduces the relative error of the high-biased estimator by a factor less than two in some cases. On the contrary, the relative error reduction factor for the low-biased estimator is much greater than two. Thus, the benefits of the bias reduction method outweigh the costs. Significant computational efficiency increases can be realized through trivial parallel

implementations using the corrected estimator.

Besides valuing American options, we are also interested in measuring the counterparty credit exposure of these contracts. The counterparty risk is the risk that a party to an over-the-counter (OTC) derivative contract (traded directly between two parties) may fail to perform on its contractual obligations, causing losses to the other party. Losses are often quantified in terms of the replacement cost of the defaulted derivatives. Counterparty exposure is the larger of zero and the market value of a derivative that would be lost if the counterparty were to default and there were zero recovery. We exclusively focus on the Monte Carlo approach because this is generally applicable to various types of products and risk factor processes. In order to find the quantile of the exposures at a future time point, we need to estimate the market value of the option at every simulated state at that specific time point. Since the valuation of American options is done by simulation, a straightforward implementation of the Monte Carlo approach for estimating counterparty exposures requires nested simulations, which is computationally formidable, especially when the risk horizon is long.

We contribute to the area of counterparty exposure estimation by proposing a regression-based method that avoids the time-consuming nested simulations. This method, which we term OLSM, is actually inherited from Longstaff and Schwartz's LSM method. The exposure at a future time point on a particular path can be estimated by the maximum of the options's intrinsic value and the estimated continuation value, where the latter is a by-product of the LSM method. By that we mean the regression coefficients are estimated using LSM, then the exposure is obtained by using these estimated regression coefficients together with the risk factors simulated under the real-world measure. It is seen that the biggest underlying task here is to estimate the regression coefficients, or the continuation value function, accurately. This task is of utmost importance for estimating the exposure of American options since an inaccurate continuation value estimate may cause an option to be exercised early, in which case the exposure is underestimated because exposures at all the future dates would be set to zero after exercise. We find that initial state dispersion and multiple- bucketing (piecewise-linear regression)

are two useful techniques for improving the regression fit. Initial state dispersion widens the support for regression, whereas multiple-bucketing address the fact that the continuation value function, in general, cannot be well approximated by a single polynomial model throughout the whole space. Furthermore, variance reduction techniques can be applied to reduce the standard error of the exposure estimates. Numerical results reveal that OLSM can dramatically reduce the number of incorrect exercise decisions, leading to a major enhancement of the accuracy of the exposure estimates. The efficiency and effectiveness of OLSM render it an excellent candidate for practitioners who are involved in measuring counterparty exposures.

Throughout the thesis, our work is based on three types of existing regression-based estimators for valuing American options, including the interleaving estimator in Longstaff and Schwartz [13], the high-biased estimator in Tsitsiklis and Van Roy [17] and its associated low-biased standard path estimator. Owing to the fact that the estimator proposed by Tsitsiklis and Van Roy often produces severely high-biased estimates (see, for example, the numerical results in Chapter 8 of Glasserman [5]), while the magnitude of bias for the low-biased estimates is relatively small, we are interested in developing a high-biased estimator that results in a smaller magnitude of bias. This work is done in Chapter 2. In Chapter 3, we consider applying a bias reduction technique to the high- and low-biased estimators, but not the interleaving estimator because the dependence of the sample paths that are used to make exercise decisions and propagate backward renders difficulties in deriving a bias approximation. Lastly, the interleaving estimator remains the best one to date. It is the best since it is economical, uses one set of sample paths only and guarantees to generate a low-biased estimator for a large sample size. On the other hand, the high-biased estimator does not always generate a valid upper bound, no matter how large the sample size is. Therefore, we decide to adopt the interleaving estimator in estimating counterparty exposures in Chapter 4.

This introductory chapter closes with a short outline of the rest of the thesis. Chapter 2 develops a new regression-based algorithm that typically generates an upper bound for the American option value. Theoretical properties of the resulting estimator are studied in this

chapter. The materials in this chapter appear in Kan and Reesor [9]. Chapter 3 extends a bias reduction technique to the context of the least-squares Monte Carlo estimators for valuing American options. The content of this chapter is based on Kan and Reesor [10] and Kan et al. [11]. Chapter 4 considers the problem of measuring counterparty exposures of American options. A version of this chapter is published in Kan et al. [8]. Finally, Chapter 5 summarizes the main results of the thesis and recommends future directions for research related to this work.

Bibliography

[1] Andersen, L., and Broadie, M. (2004) *A primal-dual simulation algorithm for pricing multi-dimensional American options*. Management Science **50** (9): 1222-1234.

[2] Broadie, M., and Glasserman, P. (1997) *Pricing American-style securities using simulation*. Journal of Economic Dynamics and Control **21** (8-9): 1323-1352.

[3] Broadie, M., and Glasserman, P. (2004) *A stochastic mesh method for pricing high-dimensional American options*. The Journal of Computational Finance **7** (4): 35-72.

[4] Carrière, J.F. (1996) *Valuation of the early-exercise price for options using simulations and nonparametric regression*. Insurance: Mathematics and Economics **19** (1): 19-30.

[5] Glasserman, P. (2004) *Monte Carlo Methods in Financial Engineering*. Springer-Verlag, New York.

[6] Haugh, M., and Kogan, L. (2004) *Pricing American options: A duality approach*. Operations Research **52** (2): 258-270.

[7] Juneja, S., and Kalra, H. (2009) *Variance reduction techniques for pricing American options using function approximations*. The Journal of Computational Finance **12** (3): 79-102.

[8] Kan, K.H.F., Frank, G., Mozgin, V., and Reesor, R.M. (2010) *Optimized least-squares Monte Carlo for measuring counterparty credit exposure of American-style options*. Mathematics-in-Industry Case Studies Journal **2**: 1-21.

[9] Kan, K.H.F., and Reesor, R.M. (2010a) *An alternative regression-based method for American option valuation*. Submitted.

[10] Kan, K.H.F., and Reesor, R.M. (2010b) *Bias reduction for pricing American options by least-squares Monte Carlo*. Submitted.

[11] Kan, K.H.F., Reesor, R.M., Whitehead, T., and Davison, M. (2009) *Correcting the bias in Monte Carlo estimators of American-style option values*. Monte Carlo and Quasi-Monte Carlo Methods 2008, ed. P. L'Ecuyer and A. B. Owen, Springer-Verlag, 439-454.

[12] Lemieux, C., and La, J. (2005) *A study of variance reduction techniques for American option pricing*. Proceedings of the 37th conference on Winter simulation, 1884-1891.

[13] Longstaff, F.A., and Schwartz, E.S. (2001) *Valuing American options by simulation: A simple least-squares approach*. Review of Financial Studies **14** (1): 113-147.

[14] Moreni, N. (2004) *A variance reduction technique for American option pricing*. Physica A: Statistical Mechanics and its Applications **338** (1-2): 292-295.

[15] Rogers, L.C.G. (2002) *Monte Carlo valuation of American options*. Mathematical Finance **12** (3): 271-286.

[16] Tilley, J.A. (1993) *Valuing American options in a path simulation model*. Transactions of the Society of Actuaries **45**: 83-104.

[17] Tsitsiklis, J., and Van Roy, B. (2001) *Regression methods for pricing complex American-style options*. IEEE Transactions on Neural Networks **12** (4): 694-703.

[18] Whitehead, T. (2008) *Correcting the Monte Carlo optimal-stopping bias*. PhD Thesis. University of Western Ontario.

Chapter 2

An alternative regression-based method for American option valuation

2.1 Introduction

Pricing American options is challenging as the option holder has the right to exercise the option before the maturity date. The problem becomes more difficult under a high-dimensional setting, for which lattice and finite difference methods are often impractical due to the curse of dimensionality. Monte Carlo simulation is the only viable means to price high-dimensional American options since its convergence rate is independent of the number of dimensions. Many simulation-based algorithms have been proposed for pricing American options in the last two decades, beginning with Tilley [12] that first dispelled the belief that simulation could not be used for American option valuation.

Regression-based methods for pricing American options were proposed in Carrière [4], Tsitsiklis and Van Roy [13], and Longstaff and Schwartz [8]. The algorithms proposed in Tsitsiklis and Van Roy [13] and Carrière [4] typically generate an upper bound, whereas the method in Longstaff and Schwartz [8] generates a lower bound. Clèment et al. [5] and Stentoft

---

[1]A version of this chapter has been submitted for publication [7].

[11] analyzed the convergence of Longstaff and Schwartz's least-squares method. Tsitsiklis and Van Roy [13] also gave convergence results for their proposed algorithm.

True upper bounds based on a duality approach were independently proposed in Rogers [10] and Haugh and Kogan [6]. Andersen and Broadie [1] developed a practical duality-based algorithm which uses only the approximation to the optimal exercise strategy to generate an upper bound. The drawback of the Andersen-Broadie algorithm is that it requires nested simulation, which means it is very computationally demanding. To remedy this drawback, Broadie and Cao [3] designed variance reduction techniques to improve the efficiency of computing the dual upper bound. Furthermore, Belomestny et al. [2] proposed a method that computes the dual upper bound without nested simulation. This method entails a regression estimator that preserves the martingale property.

The regression-based algorithm we propose in this chapter is similar to the one that appears in Tsitsiklis and Van Roy [13]. We suggest that on the in-the-money (ITM) paths continuation values are obtained by regression, whereas on the non-ITM paths the continuation values are the discounted approximate option values from the next time point along that sample path. This differs from the Tsitsiklis and Van Roy approach which uses regression to estimate continuation values on all the paths. Somewhat surprisingly, these slight modifications help generate much more accurate price estimates. The effects of these modifications are twofold:

- regression only on the ITM paths gives better fit, hence more accurate continuation value estimates;

- discounting the approximate option values on the non-ITM paths prevents the introduction of additional approximation error in regression.

By way of construction, we expect that the new method works better for out-of-the-money (OTM) options than ITM options. Numerical results, however, show that it is effective across option moneyness.

The rest of this chapter is organized as follows. Section 2.2 describes the American option

pricing problem. Section 2.3 introduces and discusses the properties of the new regression-based Monte Carlo estimator for American option value. Numerical results in Section 2.4 demonstrate the performance of the new estimator. We conclude the chapter in Section 2.5.

2.2   The Pricing Problem

In this section we set up the notation for the Bermudan option pricing problem. Bermudan options are the discrete-time version of American options, i.e., options where the holder has the right to exercise at a finite number of dates prior to the option maturity. Most simulation algorithms, including the one presented here, for pricing American options effectively price Bermudan options. In this chapter, we refer to Bermudan options as American options.

Suppose the option has n underlying assets represented by $S_t = (S_t^1, \ldots, S_t^n)$, which is a vector-valued Markov process on $\mathbb{R}^n$ with fixed initial state S_0. All processes are defined on a risk-neutral filtered probability space, $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}, \mathbb{P})$, satisfying the usual conditions. In particular, the filtration $\{\mathscr{F}_t\}$ is generated by $\{S_t\}$. Let h_t be the payoff upon exercise at time t. The payoff h_t depends on the current state and may, in general, depend on the entire history of the underlying process until time t. The option holder has d exercise opportunities at times $0 < t_1 < t_2 < \ldots < t_d = T$, with T the option maturity. For ease of notation, we write i for t_i. Without loss of generality and with a gain in clarity, we suppress the discount factor throughout the chapter. The option value is

$$Q_0(S_0) = \sup_{\tau \in \mathscr{T}} E_0[h_\tau(S_\tau)], \tag{2.1}$$

where τ is a stopping time taking values in the finite set $\mathscr{T} = \{1, 2, \ldots, d\}$. The notation $E_t[\cdot]$ is short for the expectation conditional on the information available up to time t, i.e., $E_t[\cdot] = E[\cdot | \mathscr{F}_t]$.

Valuing an American option is a discrete optimal stopping time problem that can be de-

composed, for $i = 1, \ldots, d - 1$, into the recursive equations

$$C_i(S_i) = E_i[Q_{i+1}(S_{i+1})] \quad \text{and} \tag{2.2}$$

$$Q_i(S_i) = \max(C_i(S_i), h_i(S_i)), \tag{2.3}$$

where C_i is the value of continuing after some time $i \in \mathscr{T}$, and Q_i represents the option value at time i (i.e., the greater of the value of exercising or continuing). The terminal condition is $C_d = 0$ since there is no value in holding the option past expiry. This translates to the terminal condition on the option value, $Q_d = h_d$.

In parallel to (2.2), we define

$$C_i(S_{i,j}) = E_i\left[Q_{i+1}(S_{i+1,j})\right], \tag{2.4}$$

where $S_{i,j}$ denotes the state along the jth sample path at time i. $\{S_{i,j}\}$ represents N independent sample paths, each conditional on a fixed value of S_0. This will play a role in the proof of Theorem 2.1 in Section 2.3.2.

2.3 Regression-based Estimators

In the literature of Monte Carlo pricing of American options, regression-based estimators are prevalent because of their simplicity and efficiency. In this section, we first review a regression-based estimator described in Tsitsiklis and Van Roy [13]. Then, we will introduce a new regression-based estimator, which is a slight modification of the former. Since they both typically generate an upper bound, we call them 'old high' and 'new high' estimators, respectively.

Before we proceed, we provide an explicit expression for the continuation value estimator. Regression of the option value $Q_{i+1}(S_{i+1})$ on the current state S_i implies an expression for the

continuation value of the form [2]

$$E\left[Q_{i+1}\left(S_{i+1}\right)|S_i = x\right] = \sum_{k=1}^{M} \beta_{ik}\psi_k(x), \tag{2.5}$$

for an arbitrary point $x \in \mathbb{R}^n$, some basis functions $\psi_k : \mathbb{R}^n \to \mathbb{R}$ and some constants β_{ik} for time i, and $k = 1, \dots, M$. We can equivalently write (2.5) as

$$C_i(x) = \beta_i^{\mathrm{T}}\psi(x), \tag{2.6}$$

where

$$\beta_i^{\mathrm{T}} = (\beta_{i1}, \dots, \beta_{iM}) \quad \text{and} \quad \psi(x) = (\psi_1(x), \dots, \psi_M(x))^{\mathrm{T}}.$$

For simplicity, we assume that the same basis functions are used at different exercise dates, thus ψ does not depend on i. Assuming (2.5) holds, the vector β_i is given by

$$\beta_i = \left(E\left[\psi(S_i)\psi(S_i)^{\mathrm{T}}\right]\right)^{-1} E\left[\psi(S_i)Q_{i+1}(S_{i+1})\right]. \tag{2.7}$$

The (unbiased) least-squares estimator of β_i is given by

$$\tilde{\beta}_i = \left(\frac{1}{N}\sum_{j=1}^{N}\psi(S_{i,j})\psi(S_{i,j})^{\mathrm{T}}\right)^{-1} \frac{1}{N}\sum_{j=1}^{N}\psi(S_{i,j})Q_{i+1}(S_{i+1,j}), \tag{2.8}$$

where N is the number of sample paths. As the option values Q_{i+1} are unknown, they are in practice replaced by their estimators $\hat{Q}_{i+1}$, which leads to another estimator of β_i,

$$\hat{\beta}_i = \left(\frac{1}{N}\sum_{j=1}^{N}\psi(S_{i,j})\psi(S_{i,j})^{\mathrm{T}}\right)^{-1} \frac{1}{N}\sum_{j=1}^{N}\psi(S_{i,j})\hat{Q}_{i+1}(S_{i+1,j}). \tag{2.9}$$

---

[2]More generally, the expectation can be conditioned on the history of the state variables up to time i.

The estimator $\hat{\beta}_i$ then defines a continuation value estimator through

$$\hat{C}_i(x) = \hat{\beta}_i^{\mathrm{T}} \psi(x), \tag{2.10}$$

which will be used to construct the 'old high' and 'new high' estimators in the next sections.

2.3.1 Old high estimator

The old high estimator follows the approach of Tsitsiklis and Van Roy [13]. The algorithm is given below:

- For $i = d$,

  $$\hat{Q}_d^{oh}(S_{d,j}) = h_d(S_{d,j}), \quad j = 1, \ldots, N.$$

- For $i = d-1, \ldots, 1$,

  $$\hat{C}_i^{oh}(S_{i,j}) = \hat{\beta}_i^{\mathrm{T}} \psi(S_{i,j}) \quad \text{and}$$

  $$\hat{Q}_i^{oh}(S_{i,j}) = \max(\hat{C}_i^{oh}(S_{i,j}), h_i(S_{i,j})), \quad j = 1, \ldots, N,$$

  with $\hat{\beta}_i$ as in (2.9).

- At time 0,

  $$\hat{Q}_0^{oh} = \frac{1}{N} \sum_{j=1}^{N} \hat{Q}_1^{oh}(S_{1,j}).$$

Note that for this estimator (a) all paths are used in estimating regression coefficients; and (b) the continuation value estimator on all paths is provided by the regression fit.

2.3.2 New high estimator

We propose a new high estimator which differs from the old high estimator in two aspects: (a) only ITM paths are used in estimating regression coefficients; (b) on the non-ITM paths, the time-i continuation value estimators are set to the discounted time-$(i+1)$ option value estimators. Modification (a) aims at improving the regression fit, while modification (b) prevents the introduction of additional error in approximating the continuation value by regression.

Define $J_i^N = \bigcup_{j=1}^N J_{ij}$ and $J_{ij} = \{h_i(S_{i,j}) > 0\}$ for $i = 1, \ldots, d-1$. The set, J_{ij}, and its complement, J_{ij}^c, represents that path j is ITM and non-ITM at time i, respectively. Let $\{S_{i,1'}, S_{i,2'}, \ldots, S_{i,|J_i^N|'}\}$ be the states along the paths in J_i^N, with $|J_i^N|$ the number of elements in the set J_i^N, and let $\mathbb{1}_A$ be the indicator function of the set A, taking the value of 1 on A and 0 otherwise. The algorithm for the new high estimator is presented as follows:

- For $i = d$,

  $$\hat{Q}_d^{nh}(S_{d,j}) = h_d(S_{d,j}), \quad j = 1, \ldots, N.$$

- For $i = d-1, \ldots, 1$,

  $$\hat{C}_i^{nh}(S_{i,j}) = (\hat{\beta}_i')^{\mathrm{T}} \psi(S_{i,j}) \quad \text{and}$$

  $$\hat{Q}_i^{nh}(S_{i,j}) = \hat{Q}_{i+1}^{nh}(S_{i+1,j}) \cdot \mathbb{1}_{J_{ij}^c} + \max(\hat{C}_i^{nh}(S_{i,j}), h_i(S_{i,j})) \cdot \mathbb{1}_{J_{ij}},$$

  with

  $$\hat{\beta}_i' = \left(\frac{1}{|J_i^N|} \sum_{j=1'}^{|J_i^N|'} \psi(S_{i,j}) \psi(S_{i,j})^{\mathrm{T}} \right)^{-1} \frac{1}{|J_i^N|} \sum_{j=1'}^{|J_i^N|'} \psi(S_{i,j}) \hat{Q}_{i+1}^{nh}(S_{i+1,j}). \tag{2.11}$$

- At time 0,

  $$\hat{Q}_0^{nh} = \frac{1}{N} \sum_{j=1}^N \hat{Q}_1^{nh}(S_{1,j}).$$

In what follows, we will discuss the properties of the new high estimator. We assume the following conditions are satisfied for the rest of this section.

A1: For every i and every k, $\psi_k(S_i)$ is non-negative.

A2: For every i and every k, $\psi_k(S_i)$ and $h_i(S_i)$ are in L^2.

A3: For every i and a fixed $M > 1$, if $\sum_{k=1}^M \beta_{ik} \psi_k(S_i) = 0$ almost surely, then $\beta_{ik} = 0$ for $k = 1, \ldots, M$.

Theorem 2.1 shows that for all sample sizes, the new high estimator is an upper bound on the true price given that the representation (2.5) holds. That is, the new high estimator has positive bias under this condition.

Theorem 2.1. *(Bias) If a relation of the form (2.5) is valid at all $i = 1, \ldots, d-1$, then the new high estimator is biased high, i.e., for all N,*

$$E_0 \left[\hat{Q}_0^{nh}(S_0) \right] \geq Q_0(S_0).$$

Proof: We prove this theorem by induction. At maturity we have $\hat{Q}_d^{nh}(S_{d,j}) = h_d(S_{d,j}) = Q_d(S_{d,j})$ for all j. Take as the induction hypothesis that $E_{i+1}[\hat{Q}_{i+1}^{nh}(S_{i+1,j})] \geq Q_{i+1}(S_{i+1,j})$ for all j. Now we have

$$
\begin{aligned}
& E_i \left[\hat{Q}_i^{nh}(S_{i,j}) \right] \\
&= E_i \left[\hat{Q}_{i+1}^{nh}(S_{i+1,j}) \, \mathbb{1}_{J_{ij}^c} + \max \left(\hat{C}_i^{nh}(S_{i,j}), h_i(S_{i,j}) \right) \mathbb{1}_{J_{ij}} \right] \\
&= E_i \left[E_{i+1} \left[\hat{Q}_{i+1}^{nh}(S_{i+1,j}) \right] \right] \mathbb{1}_{J_{ij}^c} + E_i \left[\max \left(\hat{C}_i^{nh}(S_{i,j}), h_i(S_{i,j}) \right) \right] \mathbb{1}_{J_{ij}} \\
&\geq E_i \left[Q_{i+1}(S_{i+1,j}) \right] \mathbb{1}_{J_{ij}^c} + \max \left(E_i \left[\hat{C}_i^{nh}(S_{i,j}) \right], h_i(S_{i,j}) \right) \mathbb{1}_{J_{ij}} \\
&\geq \max(C_i(S_{i,j}), 0) \, \mathbb{1}_{J_{ij}^c} + \max \left(C_i(S_{i,j}), h_i(S_{i,j}) \right) \mathbb{1}_{J_{ij}} \\
&= Q_i(S_{i,j}) \, \mathbb{1}_{J_{ij}^c} + Q_i(S_{i,j}) \, \mathbb{1}_{J_{ij}} \\
&= Q_i(S_{i,j}).
\end{aligned}
$$

The first step follows from the definition of $\hat{Q}_i^{nh}(S_{i,j})$. The second uses the tower property of conditional expectation and the fact that the events J_{ij} and J_{ij}^c are known given $S_{i,j}$. The third uses the induction hypothesis and Jensen's inequality. In the fourth step, the first term follows from the definition of $C_i(S_{i,j})$ and the fact that $h_i(S_{i,j}) = 0$ on the set J_{ij}^c, whereas the second term uses the inequality $E_i[\hat{C}_i^{nh}(S_{i,j})] \geq C_i(S_{i,j})$, which is due to the induction hypothesis and **A1**, the unbiasedness of $\tilde{\beta}_i'$ (analogously defined as $\hat{\beta}_i'$) and the representation (2.5). That is, $E_i[\hat{C}_i^{nh}(S_{i,j})] = E_i[(\hat{\beta}_i')^{\mathrm{T}} \psi(S_{i,j})] \geq E_i[(\tilde{\beta}_i')^{\mathrm{T}} \psi(S_{i,j})] = \beta_i^{\mathrm{T}} \psi(S_{i,j}) = C_i(S_{i,j})$. The fifth follows from the definition of $Q_i(S_{i,j})$. The last step combines the two disjoint sets.

Finally, at time 0,

$$E_0\left[\hat{Q}_0^{nh}(S_0)\right] = E_0\left[\frac{1}{N}\sum_{j=1}^N \hat{Q}_1^{nh}(S_{1,j})\right] = E_0\left[\hat{Q}_1^{nh}(S_1)\right] = E_0\left[E_1\left[\hat{Q}_1^{nh}(S_1)\right]\right]$$

$$\geq E_0\left[Q_1(S_1)\right] = Q_0(S_0).$$

The second equality uses the fact that the state variables at time 1 are identically distributed.

$\square$

Before proceeding to state the convergence result for the new high estimator, we give some additional notation. For $i = 1, \ldots, d-1$ and $j = 1, \ldots, N$, define

$$\tilde{Q}_i^{nh}(S_{i,j}) = \tilde{Q}_{i+1}^{nh}(S_{i+1,j}) \cdot \mathbb{1}_{J_{ij}^c} + \max\left(C_i(S_{i,j}), h_i(S_{i,j})\right) \cdot \mathbb{1}_{J_{ij}},$$

with $\tilde{Q}_d^{nh}(S_{d,j}) = h_d(S_{d,j})$. Note that $E_i[\tilde{Q}_i^{nh}(S_{i,j})] = Q_i(S_{i,j})$. Moreover, we denote by $\|x\|$ the Euclidean norm of a vector x. Convergence of the new high estimator is given by Theorem 2.2.

Theorem 2.2. *(Convergence) If the representation (2.5) holds at all $i = 1, \ldots, d-1$, then the new high estimator $\hat{Q}_0^{nh}(S_0)$ converges almost surely to the true value $Q_0(S_0)$ as N approaches infinity.*

The proof is based on the following lemmas.

Lemma 2.3. *For $i = 1, \ldots, d-1$, we have:*

$$|\hat{Q}_i^{nh}(S_{i,j}) - \tilde{Q}_i^{nh}(S_{i,j})| \leq \|\hat{\beta}_i' - \beta_i\| \cdot \|\psi(S_{i,j})\| \cdot \mathbb{1}_{J_{ij}} + \sum_{k=i+1}^{d-1} \|\hat{\beta}_k' - \beta_k\| \cdot \|\psi(S_{k,j})\| \cdot \mathbb{1}_{J_{ij}^c \cdots J_{(k-1)j}^c J_{kj}}.$$

Proof:

$$|\hat{Q}_i^{nh}(S_{i,j}) - \tilde{Q}_i^{nh}(S_{i,j})|$$

$$= |\max(\hat{C}_i^{nh}(S_{i,j}), h_i(S_{i,j}))\mathbb{1}_{J_{ij}} + \sum_{k=i+1}^{d-1}\max(\hat{C}_k^{nh}(S_{k,j}), h_k(S_{k,j}))\mathbb{1}_{J_{ij}^c...J_{(k-1)j}^c J_{kj}} + h_d\mathbb{1}_{J_{ij}^c...J_{(d-1)j}^c}$$

$$-\max\left(C_i(S_{i,j}), h_i(S_{i,j})\right)\mathbb{1}_{J_{ij}} - \sum_{k=i+1}^{d-1}\max\left(C_k(S_{k,j}), h_k(S_{k,j})\right)\mathbb{1}_{J_{ij}^c...J_{(k-1)j}^c J_{kj}} - h_d\mathbb{1}_{J_{ij}^c...J_{(d-1)j}^c}|$$

$$\leq |\hat{C}_i^{nh}(S_{i,j}) - C_i(S_{i,j})|\cdot\mathbb{1}_{J_{ij}} + \sum_{k=i+1}^{d-1}|\hat{C}_k^{nh}(S_{k,j}) - C_k(S_{k,j})|\cdot\mathbb{1}_{J_{ij}^c...J_{(k-1)j}^c J_{kj}}$$

$$\leq \|\hat{\beta}_i' - \beta_i\|\cdot\|\psi(S_{i,j})\|\cdot\mathbb{1}_{J_{ij}} + \sum_{k=i+1}^{d-1}\|\hat{\beta}_k' - \beta_k\|\cdot\|\psi(S_{k,j})\|\cdot\mathbb{1}_{J_{ij}^c...J_{(k-1)j}^c J_{kj}}.$$

The second step uses the triangle inequality and the inequality $|\max(a,c)-\max(b,c)| \leq |a-b|$. The last step is a result of repeated applications of the Cauchy-Schwartz inequality. $\square$

Lemma 2.4. *Assume that for $i = 1,\ldots,d-1$, the representation (2.5) holds, then $\hat{\beta}_i'$ converges almost surely to β_i.*

Proof: We proceed by induction on i. For $i = d-1$, the result is a direct consequence of the strong law of large numbers, since $|J_{d-1}^N|$ tends to infinity as N approaches infinity. Assuming that the result is true for $i = l,\ldots,d-1$, we want to prove that it is true for $i = l-1$. We have

$$\hat{\beta}_{l-1}' = \left(\frac{1}{|J_{l-1}^N|}\sum_{j=1'}^{|J_{l-1}^N|'}\psi(S_{l-1,j})\psi(S_{l-1,j})^{\mathrm{T}}\right)^{-1}\frac{1}{|J_{l-1}^N|}\sum_{j=1'}^{|J_{l-1}^N|'}\psi(S_{l-1,j})\hat{Q}_l^{nh}(S_{l,j}).$$

We know $\frac{1}{|J_{l-1}^N|}\sum_{j=1'}^{|J_{l-1}^N|'}\psi(S_{l-1,j})\psi(S_{l-1,j})^{\mathrm{T}}$ converges almost surely to $E[\psi(S_{l-1})\psi(S_{l-1})^{\mathrm{T}}]$ by the strong law of large numbers, thus it remains to show that $\frac{1}{|J_{l-1}^N|}\sum_{j=1'}^{|J_{l-1}^N|'}\psi(S_{l-1,j})\hat{Q}_l^{nh}(S_{l,j})$ converges almost surely to $E[\psi(S_{l-1})Q_l(S_l)]$. Note that $E[\psi(S_{l-1})\psi(S_{l-1})^{\mathrm{T}}]$ is finite by **A2**, and its inverse exists by **A3**. The strong law of large numbers implies the almost sure convergence of $\frac{1}{|J_{l-1}^N|}\sum_{j=1'}^{|J_{l-1}^N|'}\psi(S_{l-1,j})\tilde{Q}_l^{nh}(S_{l,j})$ to $E[\psi(S_{l-1})Q_l(S_l)]$ since $E[\psi(S_{l-1})\tilde{Q}_l^{nh}(S_l)] =$

$E[\psi(S_{l-1})Q_l(S_l)]$. Therefore, it suffices to prove that

$$\lim_{N\to\infty}\frac{1}{|J_{l-1}^N|}\sum_{j=1'}^{|J_{l-1}^N|'}\left(\psi(S_{l-1,j})\hat{Q}_l^{nh}(S_{l,j})-\psi(S_{l-1,j})\tilde{Q}_l^{nh}(S_{l,j})\right)=0.$$

Denote $G_N=\frac{1}{|J_{l-1}^N|}\sum_{j=1'}^{|J_{l-1}^N|'}\left(\psi(S_{l-1,j})\hat{Q}_l^{nh}(S_{l,j})-\psi(S_{l-1,j})\tilde{Q}_l^{nh}(S_{l,j})\right)$. Using Lemma 2.3, we obtain

$$\|G_N\|$$
$$\leq\frac{1}{|J_{l-1}^N|}\sum_{j=1'}^{|J_{l-1}^N|'}\|\psi(S_{l-1,j})\|\cdot|\hat{Q}_l^{nh}(S_{l,j})-\tilde{Q}_l^{nh}(S_{l,j})|$$
$$\leq\frac{1}{|J_{l-1}^N|}\sum_{j=1'}^{|J_{l-1}^N|'}\|\psi(S_{l-1,j})\|\left(\|\hat{\beta}_l'-\beta_l\|\cdot\|\psi(S_{l,j})\|\,\mathbb{1}_{J_{lj}}\right.$$
$$\left.+\sum_{k=l+1}^{d-1}\|\hat{\beta}_k'-\beta_k\|\cdot\|\psi(S_{k,j})\|\,\mathbb{1}_{J_{lj}^c\ldots J_{(k-1)j}^c J_{kj}}\right).$$

Since the induction hypothesis says that for $i=l,\ldots,d-1$, $\hat{\beta}_i'$ converges almost surely to β_i, we have for each $\varepsilon>0$

$$\limsup_{N\to\infty}\|G_N\|$$
$$\leq\limsup_{N\to\infty}\frac{1}{|J_{l-1}^N|}\sum_{j=1'}^{|J_{l-1}^N|'}\|\psi(S_{l-1,j})\|\left(\varepsilon\cdot\|\psi(S_{l,j})\|\,\mathbb{1}_{J_{lj}}+\sum_{k=l+1}^{d-1}\varepsilon\cdot\|\psi(S_{k,j})\|\,\mathbb{1}_{J_{lj}^c\ldots J_{(k-1)j}^c J_{kj}}\right)$$
$$=\varepsilon\cdot E\left[\|\psi(S_{l-1,j})\|\left(\|\psi(S_l)\|\,\mathbb{1}_{\tilde{J}_l}+\sum_{k=l+1}^{d-1}\|\psi(S_k)\|\,\mathbb{1}_{\tilde{J}_l^c,\ldots,\tilde{J}_{k-1}^c,\tilde{J}_k}\right)\right],$$

where $\tilde{J}_i=\{h_i(S_i)>0\}$ and $\tilde{J}_i^c=\{h_i(S_i)=0\}$, and the last equality follows from the strong law of large numbers. Letting ε go to 0, we obtain the convergence to 0. $\square$

Proof of Theorem 2.2: For each $\varepsilon > 0$, We have

$$
\begin{aligned}
&\limsup_{N\to\infty} \left|\hat{Q}_0(S_0) - Q_0(S_0)\right| \\
&\leq \limsup_{N\to\infty} \left(\frac{1}{N}\sum_{j=1}^{N}\left|\hat{Q}_1^{nh}(S_{1,j}) - \tilde{Q}_1^{nh}(S_{1,j})\right| + \left|\frac{1}{N}\sum_{j=1}^{N}\tilde{Q}_1^{nh}(S_{1,j}) - Q_0(S_0)\right| \right) \\
&\leq \limsup_{N\to\infty} \left(\frac{1}{N}\sum_{j=1}^{N}\frac{\varepsilon}{2} + \left|\frac{1}{N}\sum_{j=1}^{N}\tilde{Q}_1^{nh}(S_{1,j}) - Q_0(S_0)\right| \right) \\
&= \frac{\varepsilon}{2} + \limsup_{N\to\infty} \left(\left|\frac{1}{N}\sum_{j=1}^{N}\tilde{Q}_1^{nh}(S_{1,j}) - E_0\left[\tilde{Q}_1^{nh}(S_1)\right]\right| \right) \\
&\leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \\
&= \varepsilon
\end{aligned}
$$

The first step subtracts and adds $\frac{1}{N}\sum_{j=1}^{N}\tilde{Q}_1^{nh}(S_{1,j})$, and uses the triangle inequality. In the second step, the first term follows from Lemmas 2.3 and 2.4, which implies that $\hat{Q}_1^{nh}(S_{1,j})$ converges to $\tilde{Q}_1^{nh}(S_{1,j})$ almost surely. The third step uses the fact that $E_0\left[\tilde{Q}_1^{nh}(S_1)\right] = Q_0(S_0)$. The fourth is a direct consequence of the strong law of large numbers. Letting ε go to 0, we obtain the convergence to 0. $\qquad\square$

2.4   Numerical Study

In this section we test the new estimator on a well-studied example in the literature — an American max-call option. We suppose that the risk-neutral dynamics of the n underlying assets follow correlated geometric Brownian motion processes, i.e.,

$$
\frac{dS_t^i}{S_t^i} = (r - \delta)dt + \sigma dW_t^i, \tag{2.12}
$$

where $W_t^i, i = 1,\ldots,n$, are standard Brownian motion processes and the instantaneous correlation of W^i and W^j is ρ for all $i,j = 1,\ldots,n$ and $i \neq j$. The interest rate, r, the dividend yield,

δ, and the volatility, σ, are also assumed to be constant. Exercise opportunities are equally spaced at times with a time interval T/d. The payoff of the max-call option upon exercise at time t is

$$h_t(S_t) = (\max(S_t^1, ..., S_t^n) - K)^+, \tag{2.13}$$

where the notation x^+ means $\max(x, 0)$.

In the regressions, we use a set of ten basis functions. Let $S_{(i),t}$ be the i-th highest asset price at time t. The basis functions include the polynomials up to the third order, and are

$$c, S_{(1),t}, S_{(2),t}, S_{(1),t}^2, S_{(2),t}^2, S_{(1),t}S_{(2),t}, S_{(1),t}^3, S_{(2),t}^3, S_{(1),t}^2 S_{(2),t}, S_{(1),t}S_{(2),t}^2.$$

Note that the constant c is counted as one of the basis functions. Andersen and Broadie [1] used these ten basis functions, and polynomials in a European max-call option on $S_{(1),t}$ and $S_{(2),t}$ up to the third order. Their choice of 13 basis functions should give a better approximation (and, perhaps, smaller standard errors) to the continuation value function. Our opinion is that the computation of the European max-call option takes too much time in practice, thus we omit those three additional basis functions in our numerical experiments.

To approximate the optimal exercise strategy more efficiently, we disperse the initial states for regression, an idea first proposed in Rasmussen [9]. Initial state dispersion (ISD) is particularly helpful for estimating the continuation value function at the early stage of the option's life since it generates a wider support for regression. Rasmussen [9] suggested simulating the state variables from a distribution of initial states instead of from S_0 at time 0 given by

$$S_0 \exp\left(-\frac{1}{2}\sigma^2(T/2) + \sigma\sqrt{T/2}\varepsilon\right), \quad \varepsilon \sim \mathcal{N}(0, 1). \tag{2.14}$$

We propose a slightly different ISD scheme for the 'new high' estimator. Specifically, we put 20% of the initial states equally spaced on the interval $\left[S_0 \exp(-\sigma\sqrt{T/\alpha}), S_0 \exp(\sigma\sqrt{T/\alpha})\right]$, where α controls the width of the interval and is specified below; the rest of the initial states

are S_0. We can then get a high-biased estimate by averaging the option value estimates on the paths that are simulated from S_0. It remains to specify the value of α to make the ISD scheme effective. From our extensive numerical experiments, we find that $\alpha = 9$ works well in most situations, although some other values of α also give reasonably good estimates. Thus, we take $\alpha = 9$ for pricing various max-call options. The same ISD scheme is also applied to obtain the 'old high' estimates. We note, to the best of our knowledge, that there does not exist a systematic method in the literature to derive the value of a dispersion parameter, such as α in our ISD scheme.

After the exercise rule has been determined by the 'new high' estimator, we can simulate independent paths from S_0 to obtain a low-biased estimate.

Numerical results for $n = 2, 3$, and 5 assets are given in Table 2.1. For comparison purposes, we also obtain Longstaff and Schwartz estimates by simulation, but quote the binomial point estimates, and the point estimates and standard errors for the duality-based estimators from Andersen and Broadie [1]. It is clear that the 'old high' estimates are way off the true value (approximated by the binomial estimate). The 'new high' estimates are significantly more accurate than the 'old high' estimates and comparable to the Longstaff and Schwartz estimates in all cases. However, the 'new high' estimates occasionally drop below the true value, indicating that the new algorithm does not always generate an upper bound for the true value. It is worth noting that, for $n = 2$, the 'new high' and 'new low' estimates are all within sampling error of one another. This could happen when the selected set of basis functions is a good approximation of the true continuation value. This result is not surprising given our choice of basis functions. Therefore, it offers a means of choosing basis functions. In general, the closer the high- and low-biased estimates, the better the basis functions. As a rule of thumb, we recommend accepting a set of basis functions when the corresponding high- and low-biased estimates are within sampling error of one another.

The results reported in Table 2.1 are remarkable given the simplicity and efficiency of the method, the absence of any variance reduction techniques, and the exclusion of European op-

tion values as basis functions. Despite the positive numerical results, conservatively speaking, the 'new high' estimator cannot replace the dual upper bound. This is because the representation 2.5 does not hold exactly in practice, so Theorem 2.1, which assumes this representation, cannot guarantee an upper bound. It is, however, a nice complement to the dual upper bound. We suggest substituting the 'new high' estimator for the Longstaff and Schwartz estimator in determining the exercise strategy in the Andersen and Broadie algorithm. We can then use the 'new high' and 'new low' estimates to judge whether the selected set of basis functions is good enough before committing additional time to compute the dual upper bound. Having said that, the 'new high' and 'new low' estimates can give us a concrete idea where the true value lies with significantly less computational effort than the duality method. This may be more useful in practical situations.

2.5   Conclusions

In this chapter, we developed a regression-based method for pricing American options. Numerical results show that this method often generates an upper bound, and it is as accurate and computationally efficient as the Longstaff and Schwartz approach. The new estimator can also be used to determine whether the selected set of basis functions is good enough before spending more effort to compute the dual upper bound.

S_0	LS	OL	OH	NL	NH	ABL	ABH	BIN
$n = 2$ assets:								
90	8.063	8.040	8.588	8.060	8.068	8.065	8.069	8.075
	0.010	0.009	0.003	0.009	0.005	0.006	0.007	
	4.747	3.391	9.254	3.416	4.647			
100	13.861	13.877	14.368	13.905	13.867	13.907	13.915	13.902
	0.012	0.011	0.005	0.011	0.006	0.008	0.010	
	6.391	4.160	8.532	4.185	6.213			
110	21.333	21.286	21.850	21.329	21.323	21.333	21.340	21.345
	0.014	0.013	0.006	0.013	0.006	0.009	0.010	
	7.792	4.984	8.452	4.977	7.670			
$n = 3$ assets:								
90	11.256	11.244	11.680	11.251	11.291	11.279	11.290	11.290
	0.011	0.010	0.003	0.010	0.005	0.007	0.009	
	6.322	4.557	9.750	4.480	6.385			
100	18.671	18.668	19.137	18.686	18.700	18.678	18.703	18.690
	0.014	0.012	0.005	0.012	0.005	0.009	0.013	
	8.185	5.466	9.157	5.406	8.098			
110	27.531	27.488	27.983	27.529	27.544	27.531	27.627	27.580
	0.016	0.014	0.006	0.014	0.006	0.010	0.019	
	8.935	5.737	8.669	5.710	8.678			
$n = 5$ assets:								
90	16.624	16.612	17.009	16.619	16.714			
	0.013	0.012	0.003	0.012	0.005			
	8.349	5.989	10.225	6.012	8.142			
100	26.106	26.092	26.490	26.114	26.176			
	0.016	0.014	0.004	0.014	0.004			
	10.353	7.156	10.166	7.181	10.188			
110	36.720	36.644	36.971	36.677	36.759			
	0.018	0.016	0.005	0.016	0.005			
	11.738	7.994	10.625	7.981	11.333			

Table 2.1: Bermudan max-call option in multiasset Black-Scholes models. For each panel, the top, middle, and bottom row represent the point estimates, standard errors, and computing times (if available) for 8 estimators, namely Longstaff & Schwartz (LS), old low (OL), old high (OH), new low (NL), new high (NH), Andersen & Broadie's low (ABL) and high (ABH), and binomial (BIN). The values in the columns of ABH, ABL and BIN are quoted from Andersen and Broadie [1] for comparison. The option parameters are $K = 100, r = 5\%, \delta = 10\%, \rho = 0, T = 3$, and $\sigma = 20\%$. Exercise opportunities are equally spaced at times $t_i = iT/d, i = 1, \ldots, d$, with $d = 9$. $N = 200,000$ paths were used to obtain the LS, OH and NH estimates; $N_L = 2,000,000$ paths were used to obtain the OL and NL estimates. All the experiments were done using MATLAB with a CPU at 1.60GHz.

Bibliography

[1] Andersen, L., and Broadie, M. (2004) *A primal-dual simulation algorithm for pricing multi-dimensional American options*. Management Science **50** (9): 1222-1234.

[2] Belomestny, D., Bender, C., and Schoenmakers, J. (2009) *True upper bounds for Bermudan products via non-nested Monte Carlo*. Mathematical Finance **19** (1): 53-71.

[3] Broadie, M., and Cao, M. (2008) *Improved lower and upper bound algorithms for pricing American options by simulation*. Quantitative Finance **8** (8): 845-861.

[4] Carrière, J.F. (1996) *Valuation of the early-exercise price for options using simulations and nonparametric regression*. Insurance: Mathematics and Economics **19** (1): 19-30.

[5] Clèment, E., Lamberton, D., and Protter, P. (2002) *An analysis of a least squares regression method for American option pricing*. Finance and Stochastics **6** (4): 449-471.

[6] Haugh, M., and Kogan, L. (2004) *Pricing American options: A duality approach*. Operations Research **52** (2): 258-270.

[7] Kan, K.H.F., and Reesor, R.M. (2010) *An alternative regression-based method for American option valuation*. Submitted.

[8] Longstaff, F.A., and Schwartz, E.S. (2001) *Valuing American options by simulation: A simple least-squares approach*. Review of Financial Studies **14** (1): 113-147.

[9] Rasmussen, N. (2005) *Control variates for Monte Carlo valuation of American options*. Journal of Computational Finance **9** (1): 84-102.

[10] Rogers, L.C.G. (2002) *Monte Carlo valuation of American options*. Mathematical Finance **12** (3): 271-286.

[11] Stentoft, L. (2004) *Convergence of the least squares Monte Carlo approach to American option valuation*. Management Science **50** (9): 1193-1203.

[12] Tilley, J.A. (1993) *Valuing American options in a path simulation model*. Transactions of the Society of Actuaries **45**: 83-104.

[13] Tsitsiklis, J., and Van Roy, B. (2001) *Regression methods for pricing complex American-style options*. IEEE Transactions on Neural Networks **12** (4): 694-703.

Chapter 3

Bias reduction for pricing American options by least-squares Monte Carlo

3.1 Introduction

There is a rich history of pricing American options by simulation in the literature, beginning with Tilley [21] that first dispelled the belief that simulation could not be used for their valuation. Much subsequent work was inspired by this paper. Many methods estimate the continuation value of the contingent claim to solve the optimal-stopping-time problem arising from this application. These include Broadie and Glasserman's stochastic tree [5] and stochastic mesh [6], and the least-squares Monte Carlo (LSM) method in Longstaff and Schwartz [16]. The LSM method is very popular in practice since it is easy to implement and represents nothing more than a cross-sectional regression to estimate the continuation value is required. Earlier variants appeared in Carrière [7] and Tsitsiklis and Van Roy [22]. The regression-based approach is the focus of this chapter and we collectively refer to the various forms of regression estimators as LSM estimators.

Monte Carlo simulation is generally regarded as a method to overcome the curse of dimensionality because its convergence rate is typically independent of the number of state variables. Moreover, it is flexible enough to handle different types of models and payoff structures. However, its convergence rate can be slow. The convergence rate of a biased and consistent Monte Carlo estimator such as an LSM estimator can be enhanced in two directions, variance reduction or bias reduction. Although variance reduction for pricing American options by simulation has been widely studied (see for example [2, 6, 12, 15, 17]), little work has been done on reducing the bias in the estimators of American option values. Older bias-correction techniques in the literature are the ones that result in significantly increased estimator variance and/or computational time (see Broadie and Glasserman [4], and Carrière [7]). An exception is the one suggested by Avramidis and Hyden [2], which recursively averages the stochastic mesh high- and low-biased estimator at each exercise opportunity.

Recent works by Whitehead et al. [24, 25, 26] have made a remarkable contribution to bias reduction for Monte Carlo pricing of early-exercise options. They introduce a general bias-reduction technique, based on large sample theory, that corrects stochastic tree estimators. This technique corrects the bias due to making incorrect exercise decisions, which is the major source of bias in the estimators. In [25], numerical results from a well-studied multivariate pricing problem demonstrate that this technique significantly reduces estimator bias, hence increasing the convergence rate. A rigorous proof justifying the bias-reduction method for the high-biased stochastic tree estimator is given in [24, 26]. Motivated by these results, bias-reduced stochastic mesh estimators are given in Kan et al. [14] and Whitehead [24]. Though not yet rigorously justified, exploratory numerical results in [14, 24] clearly show the effectiveness of the technique in removing the bias from the mesh estimators.

This motivates the extension of the bias-reduction technique to LSM estimators. At each exercise opportunity we derive an approximation to estimator bias using a well-known distributional approximation for least-squares regression estimators. The bias approximation is subtracted from the LSM estimator at each exercise opportunity resulting in a corrected esti-

mator. This is done for each of the high- and low-biased LSM estimators. The bias-corrected LSM estimators are tested on a well-studied multivariate pricing problem used in Broadie and Glasserman [6], specifically an American call option on the maximum of five underlying assets. Numerical results show that the bias-correction technique significantly reduces the bias of these two types of LSM estimators.

In [14] the bias approximation and the corrected version of the high-biased LSM estimator is presented and partial numerical results show the effectiveness of the corrected high- and low-biased estimators. In this chapter, we derive the bias approximation for the low-biased LSM estimator (not given in MCQMC [14] due to space constraints) and provide extensive numerical results supporting this approach to bias correction. To be complete, we also include the derivation of the bias approximation for the high-biased estimator in this chapter. While our numerical results show that bias correction is not particularly useful for the high-biased estimator, it is very effective for the low-biased estimator. By that we mean the reduction in relative error of the high-biased estimator is not large enough to be useful in improving the computational efficiency since the additional computing time for the bias approximation is not negligible. Therefore, Item 4 on the following list is valid for the low-biased estimator, but not for the high-biased estimator; all the other points hold for both estimators.

1. The bias reduction method

   - is independent of the dimension;

   - holds for very general asset-price process and payoff structures;

   - is simple to implement by modifying a few lines of code in existing algorithms;

   - has little impact on estimator variance; and

   - results in a consistent corrected estimator under fairly general conditions (Corollary 3.2).

2. The numerical results displayed in Section 3.5 show that the corrected estimators are always better (as measured by price, relative error and root mean square error) than the

uncorrected versions across all combinations of option moneyness, number of exercise opportunities and sample size.

3. The relative superiority of the corrected to uncorrected estimators decreases with sample size due to estimator consistency — there is less bias to correct for large sample sizes.

4. By allowing for a better tradeoff between decreasing sample size M in exchange for an increased number of repeated valuations, the technique permits increased computational efficiencies over trivial parallel implementations of existing algorithms. The scale of this benefit is constrained mainly by the number of processors available. Nowadays, it is not uncommon for financial institutions to have thousands of processors available for pricing purposes (see Staum [19]).

It is worth mentioning that the bias-reduction method is designed to correct the LSM estimators to the true approximation value given by a set of basis functions used in the regression. In other words, the corrected LSM estimators suffer from the same problem as the uncorrected LSM estimators — they do not converge to the true option value with a finite number of basis functions. This is a separate problem that is beyond the scope of this chapter. We note, however, that for a given set of basis functions the corrected estimators perform better than their uncorrected counterparts.

The remainder of this chapter is organized as follows. Section 3.2 presents the valuation framework underlying the American option pricing problem. Section 3.3 discusses how to construct high- and low-biased estimators using the least-squares Monte Carlo method. Section 3.4 decomposes the error and derives the bias approximations for the LSM estimators. A proof of convergence is also given for the corrected low-biased estimator. Section 3.5 provides numerical results showing the effectiveness of this technique. Section 3.6 concludes the chapter and outlines future work.

3.2   The Valuation Framework

Assume an underlying complete filtered probability space $(\Omega, \mathscr{F}, \mathbb{Q})$ and finite time horizon $[0, T]$, where the state space Ω is the set of all possible realizations of the stochastic economy between time 0 and T, $\mathscr{F}$ is the sigma-field of distinguishable events at time T, and $\mathbb{Q}$ is a risk-neutral measure defined on $(\Omega, \mathscr{F})$. We assume that $\mathscr{F}_T = \mathscr{F}$.

We are interested in valuing an American-style contingent claim which matures at T. Suppose this claim has a multi-dimensional underlying asset-price process, $\{S_t\}$, and its time-t payoff, P_t, is a functional of the path, $\{S_\tau | 0 \leq \tau \leq t\}$, for any $t \in [0, T]$. The option owner can exercise at any time in the interval $[0, T]$ and the option value is determined by the value-maximizing exercise strategy. The exercise strategy is described by stopping times. Hence, given that the option has not yet been exercised at time t, its time-t value must be

$$B_t = \sup_{t \leq \tau \leq T} \mathbb{E}[e^{-r(\tau - t)} P_\tau | \mathscr{F}_t], \tag{3.1}$$

where the supremum is taken over all possible stopping times with values in the interval $[t, T]$, and r is the risk-free interest rate which is assumed to be constant.

Most Monte Carlo techniques are developed to solve a discrete-time version of equation (3.1), and there is no exception in this chapter. As such, a discrete-time version of the filtered probability space described above is used without any adjustment to the notation. We restrict our attention to the case in which the American option can only be exercised at N equally-spaced discrete times, $\{k\Delta T | k = 1, ..., N\}$, where $\Delta T = T/N$. This type of discrete American-style exercise feature is also sometimes termed a Bermudan-style exercise feature. By taking N to be sufficiently large, the value of a Bermudan option can be used to approximate the value of a corresponding American option. We continue to use the term American option instead of Bermudan option in this chapter. Let $k\Delta T$ be denoted by time k. The time-k value of the

American option, assuming it has yet to be exercised, is

$$B_k = \max_{\tau \in [k,...,N]} \mathbb{E}[e^{-r(\tau - k)} P_\tau | \mathscr{F}_k]. \tag{3.2}$$

This is a discrete optimal-stopping-time problem in stochastic control. It can be decomposed, via Bellman's principle of optimality, into two recursive equations

$$H_k = \mathbb{E}[e^{-r\Delta T} B_{k+1} | \mathscr{F}_k] \quad \text{and} \tag{3.3}$$

$$B_k = \max(H_k, P_k), \tag{3.4}$$

where H_k is the value of holding the option until at least the next exercise opportunity, B_k is the current value of the option (i.e., the greater of the value of holding or exercising), for $k = 1, ..., N - 1$, and $H_N = 0$ since there is no value in holding the option past expiry. The bulk of the work of many MC valuation algorithms is in estimating the continuation value.

3.3   LSM Estimators

In this section, we introduce two LSM estimators — a high-biased (high) and a low-biased (low) estimator. By construction, these estimators differ in how exercise decisions are made and option value estimators are propagated along the paths. An estimator used to make the exercise/hold decision is called a *determiner* while a *propagator* is an option value estimator passed on to the preceding exercise opportunity. High and low estimators use different determiners and propagators, the details of which we discuss in Sections 3.3.2 – 3.3.3.

In the following subsection some standard linear regression results are presented in the context of least-squares Monte Carlo pricing of American options. These results will be useful in constructing the LSM estimators and deriving the bias approximations in Section 3.4. No insight is lost in omitting, as we do, the discount factor for the rest of the chapter.

3.3.1 Linear Regression

The conditional expectation defining the continuation value is approximated by the fitted values of a regression, where the regression coefficients are estimated by the least-squares method (hence the name least-squares Monte Carlo). Specifically, discounted approximate option values at time $(k+1)$ are regressed on a set of basis functions evaluated at time k. These functions should be related to the underlying processes and the payoff function. Consider the linear regression

$$\tilde{B}^i_{k+1} = (x^i_k)' \beta_k + \varepsilon^i_k, \quad i = 1, 2, ..., n, \tag{3.5}$$

where x^i_k is a $(p \times 1)$ vector of basis functions evaluated at time k for path i, β_k is a $(p \times 1)$ vector of regression coefficients, ε^i_k is the time-k, path-i error term, $\tilde{B}^i_{k+1}$ is the time-$(k+1)$, path-i option value estimator, n is the number of sample paths and p is the number of basis functions, and $'$ denotes transpose. In matrix form this becomes

$$\tilde{\mathbf{B}}_{k+1} = X_k \beta_k + \varepsilon_k, \tag{3.6}$$

where $\tilde{\mathbf{B}}_{k+1} = (\tilde{B}^1_{k+1}, \ldots, \tilde{B}^n_{k+1})'$, $X_k = (x^1_k, \ldots, x^n_k)'$, and $\varepsilon_k = (\varepsilon^1_k, \ldots, \varepsilon^n_k)'$.

We use standard assumptions on the errors, namely that $\mathbb{E}[\varepsilon_k | \mathscr{F}_k] = \mathbf{0}$ and $\mathbb{E}[\varepsilon_k \varepsilon'_k | \mathscr{F}_k] = \text{diag}(\sigma^2_{k,1}, \ldots, \sigma^2_{k,n}) \equiv W_k$, where $\mathbf{0}$ is the column vector of zeros, $\text{diag}(a_1, \ldots, a_n)$ is the diagonal matrix with entries $(a_1, \ldots, a_n)$ and $\sigma_{k,i}$'s are constants which could be different for different values of i.

The ordinary least-squares regression estimators are

$$\tilde{\beta}_k = (X'_k X_k)^{-1} X'_k \tilde{\mathbf{B}}_{k+1}. \tag{3.7}$$

With the above assumptions on the errors it is seen that

$$\mathbb{E}\left[\tilde{\beta}_k \,\middle|\, \mathscr{F}_k\right] = \beta_k \quad \text{and} \tag{3.8}$$

$$\text{Var}\left[\tilde{\beta}_k \,\middle|\, \mathscr{F}_k\right] = \left(X_k'X_k\right)^{-1} X_k'W_kX_k \left(X_k'X_k\right)^{-1} \equiv \frac{\bar{V}_k^{\beta}}{n}. \tag{3.9}$$

Under general conditions, standard regression theory dictates a multivariate normal approximation to the $\mathscr{F}_k$-conditional distribution of $\tilde{\beta}_k$ ([23]). Specifically,

$$\tilde{\beta}_k | \mathscr{F}_k \sim \mathscr{MVN}\left(\beta_k, \frac{\bar{V}_k^{\beta}}{n}\right), \tag{3.10}$$

where $\mathscr{MVN}(\tilde{\mu}, \Sigma)$ denotes a multivariate normal random vector with mean vector $\tilde{\mu}$ and variance-covariance matrix Σ. An application of the Cramer-Wold device yields the approximate $\mathscr{F}_k$-conditional distribution of the time-k, path-i continuation value estimator

$$\tilde{H}_k^i = (x_k^i)'\tilde{\beta}_k | \mathscr{F}_k \sim \mathscr{N}\left((x_k^i)'\beta_k, \frac{\bar{V}_k^i}{n}\right), \tag{3.11}$$

where $\bar{V}_k^i/n = (x_k^i)'\bar{V}_k^{\beta} x_k^i/n$ and $\mathscr{N}(\mu, \sigma^2)$ denotes a normal random variable with mean μ and variance σ^2.

A natural estimator for the time-k, path-i variance, $\bar{V}_k^i/n$, is the time-k, path-i sample variance $\tilde{V}_k^i/n = (x_k^i)'\left(X_k'X_k\right)^{-1} X_k'\tilde{W}_kX_k \left(X_k'X_k\right)^{-1} x_k^i$, where $\tilde{W}_k = \text{diag}(\tilde{\varepsilon}_{k,1}^2, \ldots, \tilde{\varepsilon}_{k,n}^2)$ and $\tilde{\varepsilon}_{k,i}^2 = (\tilde{B}_{k+1}^i - (x_k^i)'\tilde{\beta}_k)^2$ for $i = 1, \ldots, n$. This is a consistent estimator for the assumed error variance. However, the assumption of uncorrelated errors might not hold precisely for the high and low estimators since $\tilde{H}_k^i$ in which is a weighted average of dependent random variables and their overall covariance might be nonzero. Thus, the variance estimator could be biased and inconsistent with respect to the true variance. Nevertheless, we use these estimators and in Section 3.5 show that the corrected estimators computed using the above sample variance estimators are significantly better than their uncorrected counterparts.

3.3.2 High Estimator

A high-biased LSM estimator is obtained by using the same estimator as both determiner and propagator. Specifically, the time-k, path-i continuation value estimator is the fitted value of the regression (3.5). The time-k, path-i option value estimator is simply the maximum of the immediate exercise value and the continuation value estimator. Thus, the recursive equations (3.3)and (3.4) become

$$\tilde{H}_k^i = (x_k^i)' \tilde{\beta}_k \quad \text{and} \tag{3.12}$$

$$\tilde{B}_k^i = \max(\tilde{H}_k^i, P_k^i), \tag{3.13}$$

where $\tilde{H}_k^i$ is the time-k, path-i continuation value estimator, P_k^i is the time-k, path-i exercise value, and the terminal condition is given by $\tilde{H}_N^i = 0$ for all $i = 1, \ldots, n$.

The final price estimator is

$$\tilde{B}_0 = \frac{1}{n} \sum_{i=1}^{n} \tilde{B}_1^i. \tag{3.14}$$

These estimators are easily shown to be biased high relative to the true approximation value given by the set of basis functions. They are also consistent for the true approximation value.

3.3.3 Low Estimator

A standard method of constructing a low-biased estimator is to use independent random variables for determiners and propagators, i.e., to use independent sets of information for exercise decisions and value propagation. For LSM estimators, this is done by using one set of $n/2$ simulated paths to construct the determiners used in approximating optimal stopping times for another independent set of $n/2$ paths, on which the propagators are the discounted cash flows. Along a given path, this results in a sub-optimal exercise strategy, hence resulting in a

low-biased estimator. The low estimators are

$$\tilde{H}_k^i = (x_k^i)' \tilde{\beta}_k \quad \text{(determiner)}, \tag{3.15}$$

$$\hat{H}_k^i = \tilde{B}_{k+1}^i \quad \text{(propagator) and} \tag{3.16}$$

$$\tilde{B}_k^i = \begin{cases} \hat{H}_k^i & \text{if } \tilde{H}_k^i > P_k^i \\ P_k^i & \text{if } \tilde{H}_k^i \le P_k^i \end{cases}, \tag{3.17}$$

where x_k^i is a $(p \times 1)$ vector of basis functions evaluated at time k for path i from the set of propagator paths, $\tilde{\beta}_k$ is the time-k regression estimator computed from the set of determiner paths, $\tilde{H}_k^i$ (determiner) and $\hat{H}_k^i$ (propagator) are the time-k, path-i continuation value estimators, and the terminal conditions are given by $\tilde{H}_N^i = \hat{H}_N^i = 0$ for all $i = 1, \ldots, n/2$.

The final price estimator is

$$\tilde{B}_0 = \frac{1}{n/2} \sum_{i=1}^{n/2} \tilde{B}_1^i. \tag{3.18}$$

3.4 Bias Correction

There are two sources of bias in LSM estimators of American option values:

1. Source 1 — incorrect exercise decisions due to simulation error; and

2. Source 2 — the choice of a finite set of basis functions.

Generally, any kind of function can be represented by summing an infinite number of properly chosen basis functions. In theory, therefore, the bias from Source 2 can be reduced by increasing the number of basis functions to produce a better regression approximation to the conditional expectation function. In practice, however, it is undesirable to use too many basis functions in regression, as this would raise the computational effort substantially. We refer the readers to [10] for an analysis of the tradeoffs between the number of paths and the number of basis functions. The bias from Source 2 remains an open problem and we do not deal with it in

this chapter though it is evident in our numerical results. We will instead focus on correcting the bias caused by the simulation error. In other words, we correct option value estimators to the true approximations given a finite set of basis functions, but not to the true option value. It is obvious that the success of both uncorrected and corrected LSM estimators largely relies on a good choice of basis functions.

Given a finite set of basis functions, LSM estimators of American option value are biased and consistent estimators for the true approximation value. Thus, we attempt to modify the bias reduction technique in [24] to accelerate the convergence of LSM estimators to the true approximation values. A rigorous proof of this technique for the high-biased stochastic tree estimator is provided in [24] and [26]. A rigorous proof justifying the bias-reduction method for stochastic mesh and LSM estimators is the subject of current research. Nonetheless, we apply the bias-reduction technique to LSM estimators and identify steps in the derivations that require further analysis to make them rigorous. In the following for each of the high and low estimators, the time-k, path-i estimator bias is decomposed into two terms (local bias and global bias) and a heuristic derivation for an approximation to the bias caused by simulation error is given. This derivation relies on the normal approximation to the distribution of the regression-based estimator. We argue that subtracting the bias approximation from the uncorrected estimators at each exercise opportunity results in an estimator having significantly reduced bias.

3.4.1   High Estimator

To begin, let $\bar{H}_k^i = \mathbb{E}_k[\tilde{H}_k^i] = \mathbb{E}_k[\tilde{B}_{k+1}^i]$, where $\mathbb{E}_k[\cdot]$ denotes an $\mathscr{F}_k$-conditional expectation. Here we assume there is no model error in the regression to simplify the derivation of an approximate bias expression. With H_k^i being the true time-k, path-i continuation value, the time-k, path-i bias is defined as

$$\bar{H}_k^i - H_k^i = \mathbb{E}_k[\tilde{B}_{k+1}^i - B_{k+1}^i]. \tag{3.19}$$

An estimator is high-biased if $\bar{H}_k^i - H_k^i > 0$ and is low-biased if $\bar{H}_k^i - H_k^i < 0$. Expanding the inner terms of $\mathbb{E}_k[\tilde{B}_{k+1}^i - B_{k+1}^i]$ using the definition of a high estimator (equation(3.13)) gives

$$\mathbb{E}_k\left[\max(\tilde{H}_{k+1}, P_{k+1}) - \max(H_{k+1}^i, P_{k+1}^i)\right]. \tag{3.20}$$

Adding and subtracting $\mathbb{E}_k[\max(\bar{H}_{k+1}^i, P_{k+1}^i)]$ decomposes the estimator bias into local (3.21) and global (3.22) components,

$$\mathbb{E}_k\left[\max(\tilde{H}_{k+1}^i, P_{k+1}^i) - \max(\bar{H}_{k+1}^i, P_{k+1}^i)\right] \tag{3.21}$$

$$+ \mathbb{E}_k\left[\max(\bar{H}_{k+1}^i, P_{k+1}^i) - \max(H_{k+1}^i, P_{k+1}^i)\right]. \tag{3.22}$$

We focus on the local bias first and will return to the global bias at the end of this section. Let $\mathbb{1}_R$ be an indicator function that is equal to one on the set R and is zero otherwise. The $\mathscr{F}_{k+1}$-conditional expectation (and, by nested expectation, the $\mathscr{F}_k$-conditional expectation) of $\mathbb{1}_{\bar{H}_{k+1}^i > P_{k+1}^i}(\tilde{H}_{k+1}^i - \bar{H}_{k+1}^i)$ is zero as $\mathbb{1}_{\bar{H}_{k+1}^i > P_{k+1}^i}$ is $\mathscr{F}_{k+1}$-measurable and $\mathbb{E}_{k+1}[\tilde{H}_{k+1}^i] = \bar{H}_{k+1}^i$. Therefore, this term can be subtracted inside the $\mathscr{F}_k$-conditional expectation without altering the expected value. Doing this, and expressing the max function with indicator functions, gives

$$\mathbb{E}_k\left[\mathbb{1}_{\bar{H}_{k+1}^i > P_{k+1}^i}\mathbb{1}_{\tilde{H}_{k+1}^i \leq P_{k+1}^i}(P_{k+1}^i - \tilde{H}_{k+1}^i)\right.$$
$$\left. + \mathbb{1}_{\bar{H}_{k+1}^i \leq P_{k+1}^i}\mathbb{1}_{\tilde{H}_{k+1}^i > P_{k+1}^i}(\tilde{H}_{k+1}^i - P_{k+1}^i)\right] \tag{3.23}$$

as an equivalent local bias expression. Rewritten using $\tilde{Y}_{k+1}^i = \tilde{H}_{k+1}^i - P_{k+1}^i$ and $\bar{Y}_{k+1}^i = \bar{H}_{k+1}^i - P_{k+1}^i$, this is

$$\mathbb{E}_k\left[\mathbb{1}_{\bar{Y}_{k+1}^i > 0}\mathbb{1}_{\tilde{Y}_{k+1}^i \leq 0}(-\tilde{Y}_{k+1}^i) + \mathbb{1}_{\bar{Y}_{k+1}^i \leq 0}\mathbb{1}_{\tilde{Y}_{k+1}^i > 0}(\tilde{Y}_{k+1}^i)\right]. \tag{3.24}$$

Table 3.1 summarizes (3.24). It is evident that the local bias is solely due to exercising incorrectly (i.e., choosing between holding and exercising incorrectly). Note the the local bias

is non-negative in all cases.

	Held: $\tilde{Y}^i_{k+1} > 0$	Stopped: $\tilde{Y}^i_{k+1} \leq 0$
Should Hold: $\bar{Y}^i_{k+1} > 0$	0	$-\tilde{Y}^i_{k+1}$
Should Stop: $\bar{Y}^i_{k+1} \leq 0$	$\tilde{Y}^i_{k+1}$	0

Table 3.1: The local bias in the time-k, path-i high-biased continuation value estimator.

Equivalently, the local bias given by (3.24) can be simplified to

$$\mathbb{E}_k \left[\mathbb{1}_{(\bar{Y}^i_{k+1}, \tilde{Y}^i_{k+1}) \in R} |\tilde{Y}^i_{k+1}| \right], \tag{3.25}$$

where $R = (-\infty, 0) \times (0, \infty) \cup (0, \infty) \times (-\infty, 0)$. By (3.11), for a large sample size n, we have the $\mathscr{F}_k$-conditional distributional approximation

$$\tilde{Y}^i_{k+1} | \mathscr{F}_k \sim \mathscr{N} \left(\bar{Y}^i_{k+1}, \frac{\bar{V}^i_{k+1}}{n} \right). \tag{3.26}$$

With the normal distributional approximation, the local bias (3.25) can be written in integral form as

$$\int_0^\infty \int \int_R |\tilde{y}| \frac{1}{\sqrt{\bar{v}/n}} \, \phi \left(\frac{\tilde{y} - \bar{y}}{\sqrt{\bar{v}/n}} \right) f_{\bar{Y}^i_{k+1}, \bar{V}^i_{k+1} | \mathscr{F}_k}(\bar{y}, \bar{v}) \, d\tilde{y} \, d\bar{y} \, d\bar{v}, \tag{3.27}$$

where $\phi(\cdot)$ is the standard normal density function, and $f_{\bar{Y}^i_{k+1}, \bar{V}^i_{k+1} | \mathscr{F}_k}(\cdot, \cdot)$ is the $\mathscr{F}_k$-conditional joint density function of $\bar{Y}^i_{k+1}$ and $\bar{V}^i_{k+1}$. Substituting $\bar{z} = \bar{y}\sqrt{n}$ and $\tilde{z} = \tilde{y}\sqrt{n}$ gives

$$\frac{1}{n} \int_0^\infty \int \int_R |\tilde{z}| \frac{1}{\sqrt{\bar{v}}} \, \phi \left(\frac{\tilde{z} - \bar{z}}{\sqrt{\bar{v}}} \right) f_{\bar{Y}^i_{k+1}, \bar{V}^i_{k+1} | \mathscr{F}_k} \left(\frac{\bar{z}}{\sqrt{n}}, \bar{v} \right) d\tilde{z} \, d\bar{z} \, d\bar{v}. \tag{3.28}$$

We see the time-k local bias is expected to be $O(1/n)$ due to the combined ($O(1/\sqrt{n})$) effects of a decreasing probability of making incorrect stopping decisions and an increasing probability

of those that are made being less significant.[4]

Since $\tilde{Y}^i_{k+1}$ is a consistent estimator for $\bar{Y}^i_{k+1}$, we assume both $f_{\bar{Y}^i_{k+1}, \bar{V}^i_{k+1}|\mathscr{F}_k}(\bar{z}/\sqrt{n}, \bar{v})$ and

$f_{\tilde{Y}^i_{k+1}, \bar{V}^i_{k+1}|\mathscr{F}_k}(\tilde{z}/\sqrt{n}, \bar{v})$ converge to $f_{\tilde{Y}^i_{k+1}, \bar{V}^i_{k+1}|\mathscr{F}_k}(0, \bar{v})$ as n gets large. Assuming $\lim_n$ commutes

with the integration, (3.28) then becomes asymptotically equivalent to

$$\frac{1}{n} \int_0^\infty \int \int_R |\tilde{z}| \frac{1}{\sqrt{\bar{v}}} \, \phi\left(\frac{\tilde{z} - \bar{z}}{\sqrt{\bar{v}}}\right) f_{\tilde{Y}^i_{k+1}, \bar{V}^i_{k+1}|\mathscr{F}_k}\left(\frac{\tilde{z}}{\sqrt{n}}, \bar{v}\right) \, d\tilde{z} \, d\bar{z} \, d\bar{v}. \tag{3.29}$$

Undoing the $\bar{z}$ and $\tilde{z}$ substitutions gives

$$\int_0^\infty \int \int_R |\tilde{y}| \frac{1}{\sqrt{\bar{v}/n}} \, \phi\left(\frac{\tilde{y} - \bar{y}}{\sqrt{\bar{v}/n}}\right) f_{\tilde{Y}^i_{k+1}, \bar{V}^i_{k+1}|\mathscr{F}_k}(\tilde{y}, \bar{v}) \, d\tilde{y} \, d\bar{y} \, d\bar{v}. \tag{3.30}$$

This expression is special because the integral with respect to $\bar{y}$ can be performed. This yields

$$\int_0^\infty \int_{-\infty}^\infty |\tilde{y}| \, \Phi\left(\frac{-|\tilde{y}|}{\sqrt{\bar{v}/n}}\right) f_{\tilde{Y}^i_{k+1}, \bar{V}^i_{k+1}|\mathscr{F}_k}(\tilde{y}, \bar{v}) \, d\tilde{y} \, d\bar{v}, \tag{3.31}$$

where $\Phi(\cdot)$ is the $\mathscr{N}(0,1)$ cumulative distribution function. In expectation form and using the

original random variables an approximation to the local bias is

$$\mathbb{E}_k\left[|\tilde{H}^i_{k+1} - P^i_{k+1}| \, \Phi\left(\frac{-|\tilde{H}^i_{k+1} - P^i_{k+1}|}{\sqrt{\bar{V}^i_{k+1}/n}}\right)\right]. \tag{3.32}$$

In order to utilize (3.32), it is necessary to estimate the theoretical variance $\bar{V}^i_{k+1}$ by the

sample variance $\tilde{V}^i_{k+1}$ specified in Section 3.3.1, which yields

$$\mathbb{E}_k\left[|\tilde{H}^i_{k+1} - P^i_{k+1}| \, \Phi\left(\frac{-|\tilde{H}^i_{k+1} - P^i_{k+1}|}{\sqrt{\tilde{V}^i_{k+1}/n}}\right)\right]. \tag{3.33}$$

---

[4]The statement X_n is $O(f(n))$ or $o(f(n))$ means, respectively, $\limsup_n |X_n/f(n)| <_1 \infty$ or $\lim_n |X_n/f(n)| =_1 0$ (the subscript ones indicating an almost everywhere [a.e.] sense for random variables).

Subtracting (3.33) from the bias gives

$$\mathbb{E}_k[\tilde{H}_k^i] - H_k^i - \mathbb{E}_k\left[|\tilde{H}_{k+1}^i - P_{k+1}^i| \, \Phi\left(\frac{-|\tilde{H}_{k+1}^i - P_{k+1}^i|}{\sqrt{\tilde{V}_{k+1}/n}}\right)\right] \tag{3.34}$$

$$= \mathbb{E}_k\left[\max(\tilde{H}_{k+1}^i, P_{k+1}^i) - \max(\bar{H}_{k+1}^i, P_{k+1}^i)\right] \tag{3.35}$$

$$+ \mathbb{E}_k\left[\max(\bar{H}_{k+1}^i, P_{k+1}^i) - \max(H_{k+1}^i, P_{k+1}^i)\right] \tag{3.36}$$

$$- \mathbb{E}_k\left[|\tilde{H}_{k+1}^i - P_{k+1}^i| \, \Phi\left(\frac{-|\tilde{H}_{k+1}^i - P_{k+1}^i|}{\sqrt{\tilde{V}_{k+1}/n}}\right)\right], \tag{3.37}$$

where (3.35), (3.36), and (3.37) are the time-k, path-i, local and global bias, and correction components, respectively.

The local bias and correction components cancel as the distributional approximation for the continuation value estimator becomes sharper, leaving just the global bias component. Applying Jensen's inequality to move the absolute value inside the conditional expectation and applying the inequality $|\max(x,y) - \max(u,v)| \le |x - u| + |y - v|$ to the absolute value of the conditional expectation in the global bias component gives

$$\left|\mathbb{E}_k\left[\max(\bar{H}_{k+1}^i, P_{k+1}^i) - \max(H_{k+1}^i, P_{k+1}^i)\right]\right| \tag{3.38}$$

$$\le \mathbb{E}_k\left[|\bar{H}_{k+1}^i - H_{k+1}^i|\right] = \mathbb{E}_k\left[|\mathbb{E}_{k+1}[\tilde{H}_{k+1}^i] - H_{k+1}^i|\right], \tag{3.39}$$

which shows it to be bound by the time-$(k+1)$, path-i bias. Similarly the time-$(k+1)$ global bias is bound by the time-$(k+2)$ bias. Continue in this fashion through to the next-to-last exercise opportunity ($k = N - 1$). Thus, the propagation of error across exercise opportunities is at most of the same order as the difference between the local bias and correction components. As a result, we propose a corrected option value estimator that is obtained by subtracting the local bias approximation from the original high estimator in (3.13), namely

$$\tilde{B}_k^i = \max(\tilde{H}_k^i, P_k^i) - |\tilde{H}_k^i - P_k^i| \, \Phi\left(\frac{-|\tilde{H}_k^i - P_k^i|}{\sqrt{\tilde{V}_k^i/n}}\right), \tag{3.40}$$

for $k = 1, \ldots, N - 1$, and

$$\tilde{B}_0^c = \frac{1}{n} \sum_{i=1}^n \tilde{B}_1^i. \tag{3.41}$$

3.4.2 Low Estimator

Given the uncorrected determiner $\tilde{H}_k^i$, we study the possibilities to reduce the bias of the propagator, $\hat{H}_k^i$, of a low estimator (refer to (3.15)–(3.17)). The time-k, path-i bias is defined as

$$\bar{H}_k^i - H_k^i = \mathbb{E}_k \left[\hat{H}_k^i \right] - H_k^i \tag{3.42}$$

$$= \mathbb{E}_k \left[\tilde{B}_{k+1}^i - B_{k+1}^i \right]. \tag{3.43}$$

Expanding the inner terms of $\mathbb{E}_k \left[\tilde{B}_{k+1}^i - B_{k+1}^i \right]$ using the definition of a low estimator gives

$$\mathbb{E}_k \left[\hat{H}_{k+1}^i \, \mathbb{1}_{\tilde{H}_{k+1}^i > P_{k+1}^i} + P_{k+1}^i \, \mathbb{1}_{\tilde{H}_{k+1}^i \leq P_{k+1}^i} - \max(H_{k+1}^i, P_{k+1}^i) \right]. \tag{3.44}$$

We add and subtract $\mathbb{E}_k \left[\max(\bar{H}_{k+1}^i, P_{k+1}^i) \right]$ to split this expression into a local bias (3.45) and a global bias (3.46),

$$\mathbb{E}_k \left[\hat{H}_{k+1}^i \, \mathbb{1}_{\tilde{H}_{k+1}^i > P_{k+1}^i} + P_{k+1}^i \, \mathbb{1}_{\tilde{H}_{k+1}^i \leq P_{k+1}^i} - \max(\bar{H}_{k+1}^i, P_{k+1}^i) \right] \tag{3.45}$$

$$+ \mathbb{E}_k \left[\max(\bar{H}_{k+1}^i, P_{k+1}^i) - \max(H_{k+1}^i, P_{k+1}^i) \right]. \tag{3.46}$$

As with the high estimator, we will return to the global bias at the end of this section and work on the local bias for now. The $\mathscr{F}_{k+1}$-conditional expectation (and, by nested expectation, the $\mathscr{F}_k$-conditional expectation) of $\mathbb{1}_{\tilde{H}_{k+1}^i > P_{k+1}^i} (\hat{H}_{k+1}^i - \bar{H}_{k+1}^i)$ is zero as $\mathbb{1}_{\tilde{H}_{k+1}^i > P_{k+1}^i}$ is $\mathscr{F}_{k+1}$-measurable and $\mathbb{E}_{k+1}[\hat{H}_{k+1}^i] = \bar{H}_{k+1}^i$. Therefore, this term can be subtracted inside the local

	Held: $\tilde{Y}_{k+1}^i > 0$	Stopped: $\tilde{Y}_{k+1}^i \leq 0$
Should Hold: $\bar{Y}_{k+1}^i > 0$	0	$-\bar{Y}_{k+1}^i$
Should Stop: $\bar{Y}_{k+1}^i \leq 0$	$\bar{Y}_{k+1}^i$	0

Table 3.2: The local bias in the time-k, path-i low-biased continuation value estimator.

bias without altering the $\mathscr{F}_k$-conditional expected value. Doing this gives

$$
\begin{aligned}
\mathbb{E}_k \Big[&\mathbb{1}_{\bar{H}_{k+1}^i > P_{k+1}^i} \mathbb{1}_{\tilde{H}_{k+1}^i \leq P_{k+1}^i} (P_{k+1}^i - \hat{H}_{k+1}^i) \\
&+ \mathbb{1}_{\bar{H}_{k+1}^i \leq P_{k+1}^i} \mathbb{1}_{\tilde{H}_{k+1}^i > P_{k+1}^i} (\hat{H}_{k+1}^i - P_{k+1}^i) \Big]
\end{aligned}
\tag{3.47}
$$

as an equivalent local bias expression. Furthermore, since $\tilde{H}_{k+1}^i$ and $\hat{H}_{k+1}^i$ are independent conditional on $\mathscr{F}_{k+1}$, and $\mathbb{E}_{k+1}[\hat{H}_{k+1}^i] = \bar{H}_{k+1}^i$, taking an $\mathscr{F}_{k+1}$-conditional expectation inside the $\mathscr{F}_k$-conditional expectation reduces (3.47) to

$$
\begin{aligned}
\mathbb{E}_k \Big[&\mathbb{1}_{\bar{H}_{k+1}^i > P_{k+1}^i} \mathbb{1}_{\tilde{H}_{k+1}^i \leq P_{k+1}^i} (P_{k+1}^i - \bar{H}_{k+1}^i) \\
&+ \mathbb{1}_{\bar{H}_{k+1}^i \leq P_{k+1}^i} \mathbb{1}_{\tilde{H}_{k+1}^i > P_{k+1}^i} (\bar{H}_{k+1}^i - P_{k+1}^i) \Big].
\end{aligned}
\tag{3.48}
$$

Rewritten using $\tilde{Y}_{k+1}^i = \tilde{H}_{k+1}^i - P_{k+1}^i$ and $\bar{Y}_{k+1}^i = \bar{H}_{k+1}^i - P_{k+1}^i$, this is

$$
\mathbb{E}_k \Big[\mathbb{1}_{\bar{Y}_{k+1}^i > 0} \mathbb{1}_{\tilde{Y}_{k+1}^i \leq 0} (-\bar{Y}_{k+1}^i) + \mathbb{1}_{\bar{Y}_{k+1}^i \leq 0} \mathbb{1}_{\tilde{Y}_{k+1}^i > 0} (\bar{Y}_{k+1}^i) \Big].
\tag{3.49}
$$

Table 3.2 summarizes (3.49). It is clear that the sign of the local bias is non-positive in all cases.

Applying the same arguments used to derive an approximation for the local bias of the high

estimator, we derive an approximation to (3.49).

$$
\mathbb{E}_k \left[\mathbb{1}_{\bar{Y}_{k+1}^i > 0} \, \mathbb{1}_{\tilde{Y}_{k+1}^i \leq 0} \left(-\bar{Y}_{k+1}^i \right) + \mathbb{1}_{\bar{Y}_{k+1}^i \leq 0} \, \mathbb{1}_{\tilde{Y}_{k+1}^i > 0} \left(\bar{Y}_{k+1}^i \right) \right] \tag{3.50}
$$

$$
\approx \int_0^\infty \int \int_R -|\bar{y}| \frac{1}{\sqrt{\bar{v}/(n/2)}} \, \phi\left(\frac{\tilde{y} - \bar{y}}{\sqrt{\bar{v}/(n/2)}} \right) f_{\bar{Y}_{k+1}^i, \bar{V}_{k+1}^i | \mathscr{F}_k}(\bar{y}, \bar{v}) \, \mathrm{d}\tilde{y} \, \mathrm{d}\bar{y} \, \mathrm{d}\bar{v} \tag{3.51}
$$

$$
\approx \int_0^\infty \int \int_R -|\bar{y}| \frac{1}{\sqrt{\bar{v}/(n/2)}} \, \phi\left(\frac{\tilde{y} - \bar{y}}{\sqrt{\bar{v}/(n/2)}} \right) f_{\tilde{Y}_{k+1}^i, \bar{V}_{k+1}^i | \mathscr{F}_k}(\tilde{y}, \bar{v}) \, \mathrm{d}\tilde{y} \, \mathrm{d}\bar{y} \, \mathrm{d}\bar{v} \tag{3.52}
$$

$$
= \mathbb{E}_k \left[|\tilde{Y}_{k+1}^i| \Phi\left(\frac{-|\tilde{Y}_{k+1}^i|}{\sqrt{\bar{V}_{k+1}^i/(n/2)}} \right) - \sqrt{\bar{V}_{k+1}^i/(n/2)} \, \phi\left(\frac{\tilde{Y}_{k+1}^i}{\sqrt{\bar{V}_{k+1}^i/(n/2)}} \right) \right]. \tag{3.53}
$$

The two approximations are the normal distributional approximation for the continuation value estimator and the approximation of the $\mathscr{F}_k$-conditional joint density function of $(\bar{Y}_{k+1}^i, \bar{V}_{k+1}^i)$ by the $\mathscr{F}_k$-conditional joint density function of $(\tilde{Y}_{k+1}^i, \bar{V}_{k+1}^i)$. Equality follows by integrating out the unknown variable $\bar{y}$.

Using the original random variables and the sample variance estimator yields

$$
\mathbb{E}_k \left[|\tilde{H}_{k+1}^i - P_{k+1}^i| \Phi\left(\frac{-|\tilde{H}_{k+1}^i - P_{k+1}^i|}{\sqrt{\tilde{V}_{k+1}^i/(n/2)}} \right) - \sqrt{\tilde{V}_{k+1}^i/(n/2)} \, \phi\left(\frac{\tilde{H}_{k+1}^i - P_{k+1}^i}{\sqrt{\tilde{V}_{k+1}^i/(n/2)}} \right) \right]. \tag{3.54}
$$

This variance estimator is the same as that for the high estimator, except that the x_k^i's correspond to the propagator paths.

Subtracting (3.54) from the bias gives

$$
\mathbb{E}_k \left[\hat{H}_{k+1}^i \, \mathbb{1}_{\tilde{H}_{k+1}^i > P_{k+1}^i} + P_{k+1}^i \, \mathbb{1}_{\tilde{H}_{k+1}^i \leq P_{k+1}^i} - \max(\bar{H}_{k+1}^i, P_{k+1}^i) \right] \tag{3.55}
$$

$$
+ \mathbb{E}_k \left[\max(\bar{H}_{k+1}^i, P_{k+1}^i) - \max(H_{k+1}^i, P_{k+1}^i) \right]. \tag{3.56}
$$

$$
- \mathbb{E}_k \left[|\tilde{H}_{k+1}^i - P_{k+1}^i| \Phi\left(\frac{-|\tilde{H}_{k+1}^i - P_{k+1}^i|}{\sqrt{\tilde{V}_{k+1}^i/(n/2)}} \right) - \sqrt{\tilde{V}_{k+1}^i/(n/2)} \, \phi\left(\frac{\tilde{H}_{k+1}^i - P_{k+1}^i}{\sqrt{\tilde{V}_{k+1}^i/(n/2)}} \right) \right],
$$

$$
\tag{3.57}
$$

where (3.55), (3.56) and (3.57) are the local bias, global bias and correction components, re-

spectively. As with the high estimator, the local bias and correction components cancel as the distributional approximation for the continuation value estimator becomes sharper, leaving only the global bias.

As for the global bias, applying Jensen's inequality and then the inequality $|\max(x, y) - \max(u, v)| \leq |x - u| + |y - v|$ gives

$$\left| \mathbb{E}_k \left[\max(\bar{H}_{k+1}^i, P_{k+1}^i) - \max(H_{k+1}^i, P_{k+1}^i) \right] \right| \tag{3.58}$$

$$\leq \mathbb{E}_k \left[\left| \bar{H}_{k+1}^i - H_{k+1}^i \right| \right] = \mathbb{E}_k \left[\left| \mathbb{E}_{k+1} \left[\hat{H}_{k+1}^i \right] - H_{k+1}^i \right| \right], \tag{3.59}$$

which shows it to be bound by the time-$(k+1)$, path-i bias. Using similar arguments as those used for the high estimator, we see that the propagation of error across exercise opportunities is at most of the same order as the difference between the local bias and the correction component.

The corrected option value estimator is obtained by subtracting the approximate local bias (3.54) from the original low estimator in (3.17), namely

$$\tilde{B}_k^i = \begin{cases} \hat{H}_k^i & \text{if } \tilde{H}_k^i > P_k^i, \\ P_k^i & \text{if } \tilde{H}_k^i \leq P_k^i, \end{cases}$$

$$- |\tilde{H}_k^i - P_k^i| \Phi \left(\frac{-|\tilde{H}_k^i - P_k^i|}{\sqrt{\tilde{V}_k^i/(n/2)}} \right) + \sqrt{\tilde{V}_k^i/(n/2)} \, \phi \left(\frac{\tilde{H}_k^i - P_k^i}{\sqrt{\tilde{V}_k^i/(n/2)}} \right), \tag{3.60}$$

for $k = 1, \ldots, N-1$, and

$$\tilde{B}_0^c = \frac{1}{n/2} \sum_{i=1}^{n/2} \tilde{B}_1^i. \tag{3.61}$$

3.4.3 Comments

The derivations of the bias-correction terms are based on distributional approximations for the continuation value estimators (propagators and determiners). For example, in the high estimator case, a distributional approximation is used to approximate (3.24) with (3.27) and a second distributional approximation is made to approximate (3.27) with (3.29). Similar approxima-

tions are made in the low estimator case. These distributional approximations introduce an error of order $o(1/n)$. The local and global bias components are of order $O(1/n)$. Thus the distributional approximation error is negligible compared with the local and global bias components. Furthermore, recursively subtracting the approximate local bias from the original LSM estimator reduces the local and global bias order from $O(1/n)$ to $o(1/n)$.

Since the uncorrected estimators are consistent for the true approximation value, the approximations for the local bias tend to zero as the sample size increases. Therefore, although the arguments presented here in deriving the local bias approximations rely on large sample theory, there is less bias to correct as the sample size increases. Thus we expect and, in Section 3.5, numerically show for small values of n the corrected estimators to be significantly better than the uncorrected estimators, with this relative superiority decreasing with n. This is evident in Figures 3.2–3.3 which show that the corrected estimators converge for a smaller n than is required for the convergence of the uncorrected estimators. A rigorous analysis of the relative rates of convergence of the corrected and uncorrected estimators is a subject of ongoing research.

3.4.4 Convergence

In the following, we prove that the bias approximation for the LSM low estimator converges to zero in probability. This theorem implies that the bias-corrected LSM low estimator is consistent. The reason is that [8] shows the uncorrected LSM interleaving estimator converges almost surely, and this result is directly applicable to the uncorrected LSM low estimator. Since almost-sure convergence implies convergence in probability, the corrected LSM low estimator is consistent. The proof of convergence for the corrected high-biased estimator is very similar, so is omitted.

Theorem 3.1. *Consider $\tilde{B}^i_{k+1} = (x^i_k)'\beta_k + \varepsilon^i_k$ for $i = 1,\ldots,n/2, k = 1,\ldots,N-1$. Assume that for all i and k,*

1. $\mathbb{E}\left[(P_k^i)^2\right] < \infty$, $\mathbb{E}\left[(x_k^i)'x_k^i\right] < \infty$;

2. $\mathbb{E}\left[x_k^i(x_k^i)'\right]$ *is non-singular;*

3. $\mathbb{E}\left[\varepsilon_k^i \middle| \mathscr{F}_k\right] = 0$, $\mathbb{E}\left[(\varepsilon_k^i)^2 \middle| \mathscr{F}_k\right] = \sigma_{k,i}^2$;

4. $\mathbb{E}\left[(\varepsilon_k^i)^4\right] < \infty$, $\mathbb{E}\left[\left(x_k^{ij}x_k^{il}\right)^2\right] < \infty$, *for all j, l* $(=1,\ldots,p)$;

5. $\mathbb{P}\left((x_k^i)'\beta_k = P_k^i\right) = 0$.

Then, the approximation to the local bias, $|\tilde{H}_k^i - P_k^i|\,\Phi\left(\frac{-|\tilde{H}_k^i - P_k^i|}{\sqrt{\tilde{V}_k^i/(n/2)}}\right) - \sqrt{\frac{\tilde{V}_k^i}{n/2}}\,\phi\left(\frac{\tilde{H}_k^i - P_k^i}{\sqrt{\tilde{V}_k^i/(n/2)}}\right),$ *converges to zero in probability.*

Proof: By Lemma 3.2 of [8], $\tilde{\beta}_k$ converges to β_k almost surely for $k = 1,\ldots,N-1$. Then, $\tilde{H}_k^i = (x_k^i)'\tilde{\beta}_k$ converges to $\bar{H}_k^i = (x_k^i)'\beta_k$ in probability since $(x_k^i)'$ does not vary with the sample size $n/2$ and almost sure convergence implies convergence in probability. Note that there is an abuse of notation as $(x_k^i)'$ is evaluated on the second set of $n/2$ paths. Also, standard regression theory in [23] states that $\tilde{V}_k^i$ converges in probability to $\bar{V}_k^i$ under conditions 1 - 4. In particular, condition 4 guarantees the finiteness of $\bar{V}_k^i$.

$\tilde{V}_k^i/(n/2)$ converges in probability to zero since $\tilde{V}_k^i$ converges in probability to $\bar{V}_k^i$ which is finite. $\tilde{H}_k^i - P_k^i$ converges in probability to $\bar{H}_k^i - P_k^i$ which is bounded by condition 1. Moreover, $\bar{H}_k^i - P_k^i$ is not equal to zero almost surely by condition 5. Therefore, by Slutzky's theorem and the continuous mapping theorem, $-|\tilde{H}_k^i - P_k^i|/\sqrt{\tilde{V}_k^i/(n/2)}$ converges in probability to $-\infty$ and $(\tilde{H}_k^i - P_k^i)/\sqrt{\tilde{V}_k^i/(n/2)}$ to $-\infty/+\infty$ depending on the sign of $\bar{H}_k^i - P_k^i$, respectively. Furthermore, $\Phi(\cdot)$ and $\phi(\cdot)$ are continuous functions. Another application of the continuous mapping theorem results in the bias approximation converging to zero in probability.

$\square$

Corollary 3.2. *Under the conditions of Theorem 3.1, the bias-corrected LSM low estimator (3.61) is consistent.*

Proof: Let n_d and n_p be the number of sample paths that are used to estimate the regression coefficients and the discounted cash flows (propagators), respectively. (In this chapter, $n_d = n_p = n/2$.)

Letting n_d go to infinity, for $k = 1, \ldots, N-1$, by Theorem 3.1, $\tilde{B}_k^i$ converges in probability to

$$\hat{B}_k^i = \begin{cases} \hat{H}_k^i & \text{if } \bar{H}_k^i > P_k^i \\ P_k^i & \text{if } \bar{H}_k^i \leq P_k^i \end{cases},$$

and $\tilde{B}_0^c$ converges in probability to

$$\hat{B}_0 = \frac{1}{n_p} \sum_{i=1}^{n_p} \hat{B}_1^i. \tag{3.62}$$

Moreover, $\hat{B}_1^i$ are independent, identically distributed random variables. Therefore, by law of large numbers, $\hat{B}_0$ converges in probability to the expectation of $\hat{B}_1^i$, which is

$$\mathbb{E}\left[\mathbb{E}_1\left[\hat{B}_1^i\right]\right] = \mathbb{E}\left[\mathbb{E}_1\left[\hat{H}_1^i\right] \mathbb{1}_{\bar{H}_1^i > P_1^i} + P_1^i \mathbb{1}_{\bar{H}_1^i \leq P_1^i}\right] \tag{3.63}$$

$$= \mathbb{E}\left[\bar{H}_1^i \mathbb{1}_{\bar{H}_1^i > P_1^i} + P_1^i \mathbb{1}_{\bar{H}_1^i \leq P_1^i}\right] \tag{3.64}$$

$$= B_0, \tag{3.65}$$

where the first equality holds because $\bar{H}_1^i$ and P_1^i belong to the filtration $\mathscr{F}_1$, the second equality holds by condition 3 in Theorem 3.1, and the third equality holds by definition of the option value. Hence, the bias-corrected LSM low estimator is consistent. $\quad\square$

3.5   Numerical Results

The bias-corrected LSM high and low estimators are tested on a well-studied example from [6] — an American-style max-call option with five underlying assets and a maturity of three years.

The value of such an option at time t is

$$\sup_{t \leq \tau \leq T} \mathbb{E}\left[e^{-r(\tau-t)}(\max(S_\tau^1,...,S_\tau^5) - K)^+|\mathscr{F}_t\right], \qquad (3.66)$$

where K is the strike price and $S_t^1,...,S_t^5$ are the underlying asset-prices for $t \in [0,T]$. These processes are modeled with independent geometric Brownian motions, pay a continuous dividend of 10% and have a volatility of 20%. The strike price of the option is $100 and the risk-free interest rate is 5%. We consider cases in which the initial stock prices S_0 are all set to the same value either of $90, $100 or $110 – encompassing a range of moneyness. The number of exercise opportunities N is one of 3, 6, 9, 12, 15 or 18. In addition to moneyness and number of exercise opportunities, we investigate how the sample size n affects the performance of the corrected estimators.

The implementation of the LSM method can be divided into two parts; (i) simulation of the underlying sample paths; and (ii) dynamic programming. To avoid unnecessary variations due to random number generation, the same sets of sample paths for all combinations of S_0 and N are used to compute price estimators. Thus, there are 18 sets of $(1,280,000 \times 5)$ independent sample paths in total, where $1,280,000$ is the product of the number of repeated valuations (m) and the sample size (n), and 5 is the dimension of the underlying. Note that we use the total number of sample paths as the budget constraint for all test cases.

To focus attention on estimator bias, independent repeated valuations are performed to control estimator variance. Specifically, on each set of n sample paths we compute LSM estimators. This is repeated using m independent sets of n sample paths, resulting in m independent values of the estimators. These m values are averaged to give the price estimator. The standard error of the price estimator is the standard deviation of the estimators divided by the square root of m. Increasing the number of repeated valuations only affects the estimator standard error and has no effect on the bias.

In the cross-sectional regressions, seven basis functions are used, consisting of a constant,
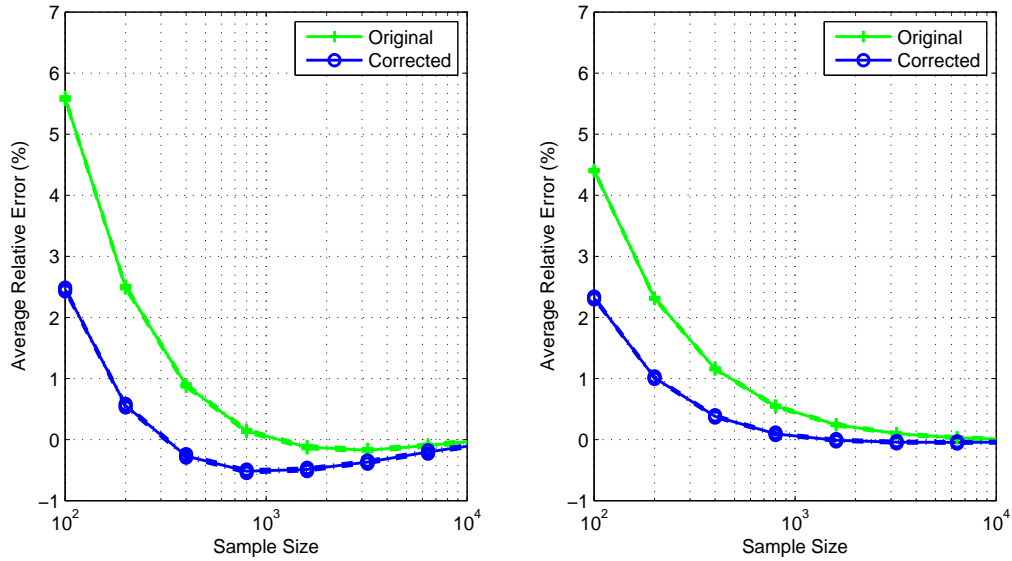
Figure 3.1: Average relative errors (%) of the high estimators versus sample size (n) for $N = 18$ exercise opportunities. Left and right panels correspond to initial stock price $S_0 = \$90$ and $S_0 = \$110$, respectively.

the first three monomials in the maximum of the values of the five assets, the first two monomials in the second largest value, and the third largest value. This set of basis functions might not be best, but it suffices for demonstration purposes. Computational experiments were done using MATLAB with a CPU at 1.60GHz. All numerical results for the LSM estimators, including the (N, S_0, m, n) combinations used, are provided in Tables 3.3, 3.4, 3.5 and 3.6. In this section, we provide figures for illustrative purposes that are produced using a subset of the results in Tables 3.3 and 3.4.

3.5.1 Discussions

It is useful to evaluate estimators using average relative error (ARE) as this metric focuses solely on estimator bias. The relative error of a high or low estimator is given by $(\tilde{B}_0 - B_0^{\text{true}})/B_0^{\text{true}}$ or $(B_0^{\text{true}} - \tilde{B}_0)/B_0^{\text{true}}$, where B_0^{true} denotes the true approximation value which is taken to be the uncorrected option value estimate with the largest sample size $n = 12800$. Averaging the relative errors across m replications gives the ARE. A true approximation value
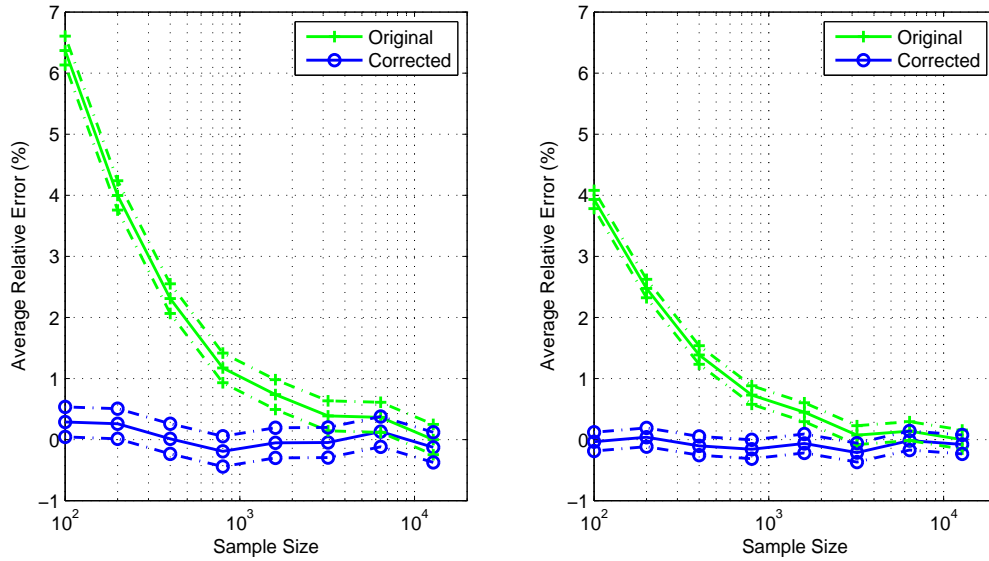
Figure 3.2: Average relative errors (%) of the low estimators versus sample size (n) for $N = 18$ exercise opportunities. Left and right panels correspond to initial stock price $S_0 = \$90$ and $S_0 = \$110$, respectively.

instead of a true option value is used in the ARE calculations since our bias reduction method does not close the gap between the true approximation value and the true option value, which is due to approximation errors in the basis functions.

First, we fix $N = 18$ and investigate the effect of moneyness and sample size on the estimators. Figure 3.1 plots the AREs of the high estimators and their corresponding 95% confidence intervals against sample size, where the left and right panels correspond to $S_0 = \$90$ and $S_0 = \$110$, respectively. Clearly, the corrected estimators outperform their uncorrected counterparts regardless of option moneyness. The slight over-correction for $S_0 = \$90$ may not be a bad thing provided that the true approximation value for the high estimator is typically an upper bound for the true option value. Nonetheless, the bias reduction method is not very useful here as it reduces the ARE by a factor of about two for the small sample sizes, while the computing time for the corrected high estimator is doubled compared to the uncorrected version. The benefit of the application of the bias reduction method is essentially wiped out by the additional computational costs for evaluating the bias approximation.
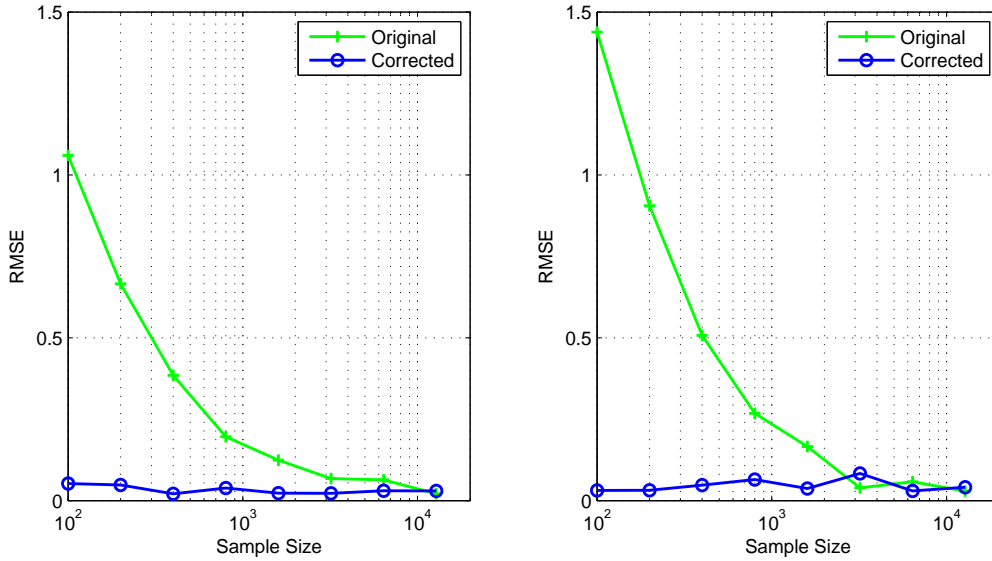
Figure 3.3: Root mean square errors of the low estimators against sample size (n) for $N = 18$ exercise opportunities. Left and right panels correspond to initial stock price $S_0 = \$90$ and $S_0 = \$110$, respectively.

In what follows, we will solely focus on the discussion of the numerical results for the low estimator, which does not bear the same drawback as the high estimator when applying the bias reduction method.

In Figure 3.2, we plot the AREs of the low estimators and their corresponding 95% confidence intervals against sample size, where the left and right panels correspond to $S_0 = \$90$ and $S_0 = \$110$, respectively. It appears that the bias reduction method works well regardless of option moneyness — the ARE of the corrected estimator is close to zero for both in-the-money (ITM) and out-of-the-money (OTM) options. On the other hand, sample size has an unobvious impact on the corrected estimator for the ITM option, but a larger sample size reduces the ARE of the corrected estimator for the OTM option so that the corrected estimator becomes essentially unbiased. The bias-corrected estimator is much better than its uncorrected counterpart at smaller sample sizes. This superiority decreases with sample size, as the uncorrected estimator is consistent. This effect is evident across all metrics (i.e., average relative error and root mean square error) used here to evaluate estimators (Figures 3.2 – 3.3).
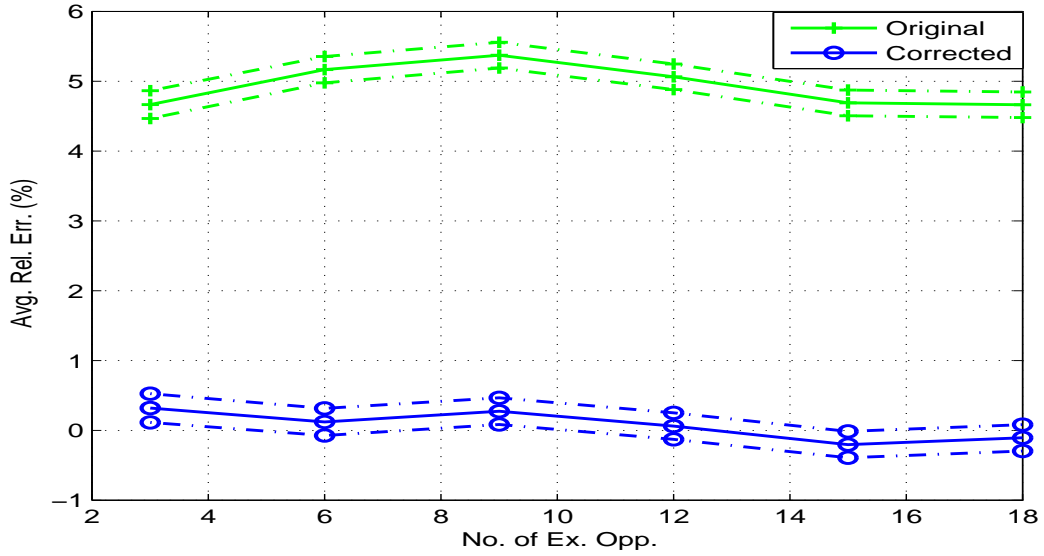
Figure 3.4: Average relative errors (%) of the low estimators versus number of exercise opportunities (N) for initial stock price $S_0 = \$100$ and sample size $n = 100$.

The root mean square error (RMSE) is another useful metric for evaluating estimators. The mean square error (MSE) of an estimator $\tilde{B}_0$ is

$$\text{MSE}(\tilde{B}_0) = \text{Var}(\tilde{B}_0) + [\text{bias}(\tilde{B}_0)]^2,$$

where $\text{bias}(\tilde{B}_0) = \mathbb{E}[\tilde{B}_0] - B_0$. The RMSE is a metric that is sensitive to contributions from both estimator variance and bias. Here, we estimate the bias of an estimator by multiplying the ARE by m, the number of repeated valuations. The estimate of variance is simply the square of the estimator standard error.

In Figure 3.3, we fix $N = 18$ and plot the RMSEs of the price estimators against sample size, where the left and right panels correspond to $S_0 = \$90$ and $S_0 = \$110$, respectively. For both ITM and OTM options, the corrected estimator has a smaller RMSE compared with the uncorrected estimator. The same thing generally holds for all sample sizes, although statistical errors make the RMSEs of the two estimators indistinguishable at large sample sizes. This is due to the fact that the correction terms significantly reduce the bias with little impact on the
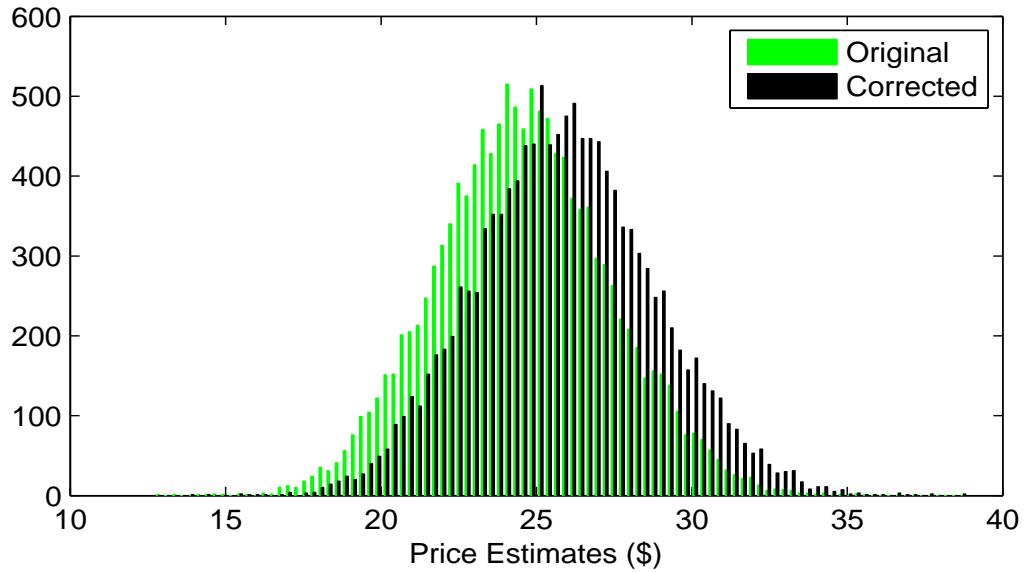
Figure 3.5: A histogram of $m = 12800$ estimates of the low estimators for $N = 9$ exercise opportunities, initial stock price $S_0 = \$100$, and sample size $n = 100$.

estimator variance.

Figure 3.4 plots the AREs of the price estimators and their corresponding 95% confidence intervals against number of exercise opportunities for an initial stock price of \$100 and a sample size of 100. It shows that the bias of both corrected and uncorrected estimators is unaffected by N. In particular, the corrected estimator is close to unbiased across N. This is a very attractive feature of the bias reduction method as a Bermudan option with more exercise opportunities approximates an American option better.

We plot a histogram of the price estimators in Figure 3.5 to illustrate the effects of the bias reduction method on the estimators. Specifically, there are $m = 12800$ estimates for each of the corrected and uncorrected estimators, and they correspond to an option with $N = 9$ exercise opportunities and an initial stock price $S_0 = \$100$. The sample size used to generate each estimate is $n = 100$. From Figure 3.5, it is clear that the distribution of the estimates of the corrected estimator is shifted toward the right hand side (the larger values) while the dispersion of them is apparently preserved compared to the uncorrected estimator. This is consistent with the above findings of reduced bias and unchanged variance for the corrected estimator.

3.5.2 Parallel Computing

The reduction in bias is not totally free. The cost comes from the additional effort required to compute the correction term. The bias-correction algorithm was implemented using Matlab on one processor (i.e., serial computing). We found that the computing time for the corrected estimator was roughly twice that of the uncorrected one for a fixed sample size. However, the ARE of the corrected estimator falls well within 1% for a sample size as small as 100, whereas the uncorrected estimator needs at least 3200 sample paths to achieve a comparable ARE (c.f. Figure 3.2). This presents an opportunity to take advantage of parallel computing to improve computational efficiencies.

To illustrate, denote the number of repeated valuations, the sample size for the corrected and uncorrected estimators, and the number of processors available by m_c, m_u, n_c, n_u, and N_p, respectively. The notation $\lceil x \rceil$ represents the least integer that is greater than or equal to a real number x. Then, the efficiency ratio is

$$\frac{\lceil m_u/N_p \rceil * n_u}{\lceil m_c/N_p \rceil * n_c * 2},\tag{3.67}$$

where 2 is the penalty factor for the additional computational time for the bias-correction term (see Tables 3.5 and 3.6). To attain a similar ARE for the corrected and uncorrected estimators, we take $m_c = 12800$, $m_u = 400$, $n_c = 100$, $n_u = 3200$. When $N_p = 3200$, the efficiency ratio is 4. In other words, implementing four repeated valuations with sample size 100 on each of the 3200 processors for the corrected estimator reduces the computational time by a factor of four compared to one valuation with sample size 3200 on each of the 400 processors for the uncorrected estimator. In the best case when $N_p \geq 12800$, the efficiency ratio becomes 16.

Note that we assume only the outer loops (the independent, repeated valuations) are parallelized. A parallel implementation of a single valuation requires a break-down of the cross-sectional regressions into several pieces. This would force different processors to communicate with each other during each valuation, which is very time-consuming and difficult to imple-

ment efficiently. Furthermore, the communication costs increase linearly with the number of exercise opportunities. These issues can offset the benefits of parallelizing.

3.6 Conclusions

We developed a bias-reduction technique for the high- and low-biased least-squares Monte Carlo estimators of American option values. This method is applicable to virtually all underlying processes and payoff functions in practice. Our numerical results show that this technique is effective in reducing estimator bias across all combinations of sample size, moneyness and number of exercise opportunities. However, it only has significant improvement for the low-biased estimator when computing time is taken into account.

In addition to the directions of further study discussed above, we are currently working on a rigorous justification of the bias-corrected LSM estimators. Furthermore, we will explore the applicability of this technique to optimal-switching time problems, multiple-exercise options and estimating sensitivities for American-style options.

Table 3.3: (High-biased Estimators) Original and corrected estimates and their respective standard errors (below the estimates). S_0 is the initial stock price, N is the number of exercise opportunities, m is the number of repeated valuations, and n is the sample size. The other parameters are strike price $K(= \$100)$, risk-free interest rate $r(= 5\%)$, volatility $\sigma(= 20\%)$, dividend yield $\delta(= 10\%)$ and maturity $T(= 3$ years).

N	m	n	$S_0 = \$90$		$S_0 = \$100$		$S_0 = \$110$	
			Original	Corrected	Original	Corrected	Original	Corrected
3	12800	100	16.755	16.337	26.193	25.667	36.707	36.080
			0.009	0.009	0.012	0.011	0.013	0.013
	6400	200	16.396	16.128	25.741	25.404	36.176	35.777
			0.009	0.009	0.011	0.011	0.013	0.013
	3200	400	16.197	16.034	25.495	25.292	35.878	35.640
			0.008	0.008	0.011	0.011	0.013	0.012
	1600	800	16.095	16.002	25.361	25.246	35.725	35.592
			0.008	0.008	0.011	0.010	0.012	0.012
	800	1600	16.040	15.990	25.294	25.232	35.644	35.573
			0.008	0.008	0.010	0.010	0.012	0.012
	400	3200	16.013	15.986	25.261	25.229	35.606	35.568
			0.008	0.008	0.010	0.010	0.012	0.012
	200	6400	16.001	15.987	25.245	25.228	35.585	35.565
			0.008	0.008	0.010	0.010	0.012	0.012
	100	12800	15.996	15.989	25.239	25.230	35.574	35.564
			0.008	0.008	0.010	0.010	0.012	0.012
6	12800	100	17.648	17.120	27.300	26.650	37.974	37.205
			0.006	0.006	0.007	0.007	0.009	0.008
	6400	200	17.180	16.844	26.721	26.311	37.290	36.804
			0.005	0.005	0.007	0.007	0.008	0.008
	3200	400	16.920	16.716	26.405	26.158	36.905	36.613
			0.005	0.005	0.007	0.007	0.008	0.008
	1600	800	16.790	16.672	26.242	26.099	36.698	36.531
			0.005	0.005	0.006	0.006	0.008	0.008
	800	1600	16.729	16.663	26.162	26.083	36.594	36.502
			0.005	0.005	0.006	0.006	0.008	0.007
	400	3200	16.705	16.669	26.125	26.083	36.542	36.493
			0.005	0.005	0.006	0.006	0.007	0.007
	200	6400	16.698	16.679	26.110	26.088	36.515	36.489
			0.005	0.005	0.006	0.006	0.007	0.007
	100	12800	16.700	16.690	26.106	26.094	36.504	36.490
			0.005	0.005	0.006	0.006	0.007	0.007

N	m	n	$S_0 = \$90$ Original	$S_0 = \$90$ Corrected	$S_0 = \$100$ Original	$S_0 = \$100$ Corrected	$S_0 = \$110$ Original	$S_0 = \$110$ Corrected
9	12800	100	18.051	17.497	27.785	27.107	38.583	37.782
			0.005	0.004	0.006	0.006	0.007	0.007
	6400	200	17.544	17.196	27.168	26.742	37.837	37.335
			0.004	0.004	0.005	0.005	0.006	0.006
	3200	400	17.270	17.060	26.835	26.580	37.424	37.122
			0.004	0.004	0.005	0.005	0.006	0.006
	1600	800	17.135	17.013	26.662	26.515	37.204	37.030
			0.004	0.004	0.005	0.005	0.006	0.006
	800	1600	17.077	17.010	26.580	26.498	37.091	36.994
			0.004	0.004	0.005	0.005	0.006	0.006
	400	3200	17.053	17.016	26.545	26.501	37.035	36.983
			0.004	0.004	0.005	0.005	0.006	0.006
	200	6400	17.046	17.027	26.530	26.506	37.009	36.981
			0.004	0.004	0.005	0.005	0.006	0.006
	100	12800	17.053	17.042	26.526	26.514	36.997	36.983
			0.004	0.004	0.005	0.005	0.006	0.006
12	12800	100	18.309	17.749	28.095	27.416	38.958	38.155
			0.004	0.004	0.005	0.005	0.006	0.005
	6400	200	17.782	17.431	27.454	27.030	38.197	37.695
			0.003	0.003	0.004	0.004	0.005	0.005
	3200	400	17.500	17.289	27.114	26.860	37.764	37.463
			0.003	0.003	0.004	0.004	0.005	0.005
	1600	800	17.364	17.243	26.944	26.798	37.540	37.366
			0.003	0.003	0.004	0.004	0.005	0.005
	800	1600	17.310	17.242	26.863	26.781	37.424	37.326
			0.003	0.003	0.004	0.004	0.005	0.005
	400	3200	17.292	17.255	26.829	26.785	37.370	37.317
			0.003	0.003	0.004	0.004	0.005	0.005
	200	6400	17.297	17.277	26.819	26.796	37.343	37.315
			0.003	0.003	0.004	0.004	0.005	0.005
	100	12800	17.309	17.298	26.821	26.809	37.331	37.316
			0.003	0.003	0.004	0.004	0.005	0.005

N	m	n	$S_0 = \$90$		$S_0 = \$100$		$S_0 = \$110$	
			Original	Corrected	Original	Corrected	Original	Corrected
15	12800	100	18.510	17.950	28.335	27.658	39.160	38.362
			0.003	0.003	0.004	0.004	0.005	0.005
	6400	200	17.971	17.622	27.688	27.269	38.377	37.880
			0.003	0.003	0.004	0.004	0.004	0.004
	3200	400	17.688	17.480	27.343	27.093	37.936	37.638
			0.003	0.003	0.004	0.004	0.004	0.004
	1600	800	17.555	17.436	27.172	27.028	37.709	37.536
			0.003	0.003	0.004	0.003	0.004	0.004
	800	1600	17.507	17.441	27.092	27.013	37.592	37.494
			0.003	0.003	0.003	0.003	0.004	0.004
	400	3200	17.497	17.461	27.063	27.020	37.535	37.482
			0.003	0.003	0.003	0.003	0.004	0.004
	200	6400	17.504	17.485	27.056	27.033	37.508	37.480
			0.003	0.003	0.003	0.003	0.004	0.004
	100	12800	17.522	17.511	27.059	27.047	37.498	37.482
			0.003	0.003	0.003	0.003	0.004	0.004
18	12800	100	18.657	18.105	28.552	27.884	39.376	38.589
			0.003	0.003	0.004	0.004	0.004	0.004
	6400	200	18.112	17.769	27.898	27.484	38.587	38.097
			0.003	0.003	0.003	0.003	0.004	0.004
	3200	400	17.827	17.623	27.553	27.308	38.151	37.858
			0.003	0.003	0.003	0.003	0.004	0.004
	1600	800	17.696	17.579	27.386	27.245	37.921	37.751
			0.002	0.002	0.003	0.003	0.004	0.004
	800	1600	17.649	17.584	27.316	27.238	37.807	37.711
			0.002	0.002	0.003	0.003	0.004	0.004
	400	3200	17.640	17.606	27.290	27.248	37.752	37.699
			0.002	0.002	0.003	0.003	0.004	0.004
	200	6400	17.653	17.635	27.289	27.267	37.725	37.697
			0.002	0.002	0.003	0.003	0.004	0.004
	100	12800	17.671	17.661	27.295	27.283	37.715	37.700
			0.002	0.002	0.003	0.003	0.004	0.004

Table 3.4: (Low-biased Estimators) Original and corrected estimates and their respective standard errors (below the estimates). S_0 is the initial stock price, N is the number of exercise opportunities, m is the number of repeated valuations, and n is the sample size. The other parameters are strike price $K(=\$100)$, risk-free interest rate $r(=5\%)$, volatility $\sigma(=20\%)$, dividend yield $\delta(=10\%)$ and maturity $T(=3$ years).

N	m	n	$S_0 = \$90$		$S_0 = \$100$		$S_0 = \$110$	
			Original	Corrected	Original	Corrected	Original	Corrected
3	12800	100	15.003	15.861	24.006	25.100	34.168	35.491
			0.021	0.022	0.026	0.027	0.029	0.030
	6400	200	15.339	15.871	24.390	25.065	34.698	35.503
			0.021	0.022	0.026	0.026	0.029	0.030
	3200	400	15.581	15.899	24.719	25.122	35.028	35.505
			0.022	0.022	0.026	0.026	0.029	0.030
	1600	800	15.699	15.883	24.925	25.154	35.285	35.552
			0.022	0.022	0.026	0.026	0.029	0.030
	800	1600	15.864	15.965	25.075	25.197	35.412	35.556
			0.022	0.022	0.026	0.026	0.029	0.030
	400	3200	15.883	15.937	25.124	25.189	35.526	35.602
			0.022	0.022	0.026	0.026	0.029	0.030
	200	6400	15.912	15.941	25.128	25.162	35.566	35.605
			0.022	0.022	0.026	0.026	0.030	0.030
	100	12800	15.931	15.945	25.180	25.197	35.598	35.618
			0.022	0.022	0.026	0.026	0.030	0.030
6	12800	100	15.287	16.328	24.424	25.724	34.706	36.246
			0.021	0.021	0.025	0.026	0.028	0.029
	6400	200	15.704	16.354	24.890	25.684	35.250	36.197
			0.021	0.021	0.025	0.026	0.029	0.029
	3200	400	15.954	16.347	25.251	25.733	35.643	36.211
			0.021	0.021	0.025	0.026	0.029	0.029
	1600	800	16.154	16.383	25.494	25.771	35.916	36.243
			0.021	0.021	0.025	0.025	0.029	0.029
	800	1600	16.301	16.429	25.647	25.802	36.081	36.260
			0.021	0.021	0.025	0.025	0.029	0.029
	400	3200	16.307	16.377	25.698	25.780	36.221	36.318
			0.021	0.021	0.025	0.025	0.029	0.029
	200	6400	16.381	16.419	25.767	25.810	36.298	36.348
			0.021	0.021	0.025	0.025	0.029	0.029
	100	12800	16.406	16.426	25.755	25.777	36.298	36.324
			0.021	0.021	0.025	0.025	0.029	0.029

N	m	n	$S_0 = \$90$		$S_0 = \$100$		$S_0 = \$110$	
			Original	Corrected	Original	Corrected	Original	Corrected
9	12800	100	15.422	16.493	24.571	25.895	34.932	36.490
			0.020	0.021	0.025	0.025	0.028	0.029
	6400	200	15.824	16.479	25.103	25.920	35.497	36.455
			0.021	0.021	0.025	0.025	0.028	0.029
	3200	400	16.111	16.514	25.438	25.930	35.923	36.502
			0.021	0.021	0.025	0.025	0.028	0.029
	1600	800	16.312	16.549	25.638	25.925	36.204	36.542
			0.021	0.021	0.025	0.025	0.028	0.029
	800	1600	16.417	16.551	25.803	25.964	36.334	36.523
			0.021	0.021	0.025	0.025	0.029	0.029
	400	3200	16.484	16.557	25.873	25.961	36.455	36.557
			0.021	0.021	0.025	0.025	0.029	0.029
	200	6400	16.498	16.537	25.940	25.986	36.518	36.572
			0.021	0.021	0.025	0.025	0.029	0.029
	100	12800	16.522	16.542	25.966	25.990	36.548	36.576
			0.021	0.021	0.025	0.025	0.029	0.029
12	12800	100	15.514	16.573	24.691	25.992	35.090	36.621
			0.020	0.021	0.024	0.025	0.028	0.029
	6400	200	15.885	16.538	25.157	25.951	35.634	36.570
			0.020	0.021	0.025	0.025	0.028	0.029
	3200	400	16.169	16.568	25.495	25.978	36.025	36.595
			0.021	0.021	0.025	0.025	0.028	0.029
	1600	800	16.369	16.607	25.742	26.024	36.305	36.642
			0.021	0.021	0.025	0.025	0.029	0.029
	800	1600	16.449	16.585	25.871	26.030	36.439	36.629
			0.021	0.021	0.025	0.025	0.029	0.029
	400	3200	16.526	16.600	25.922	26.010	36.534	36.639
			0.021	0.021	0.025	0.025	0.029	0.029
	200	6400	16.538	16.577	25.990	26.036	36.623	36.679
			0.021	0.021	0.025	0.025	0.029	0.029
	100	12800	16.590	16.611	26.008	26.032	36.650	36.679
			0.021	0.021	0.025	0.025	0.029	0.029

N	m	n	$S_0 = \$90$ Original	$S_0 = \$90$ Corrected	$S_0 = \$100$ Original	$S_0 = \$100$ Corrected	$S_0 = \$110$ Original	$S_0 = \$110$ Corrected
15	12800	100	15.553	16.591	24.759	26.030	35.122	36.613
			0.020	0.021	0.024	0.025	0.028	0.029
	6400	200	15.970	16.609	25.226	26.008	35.638	36.551
			0.021	0.021	0.025	0.025	0.028	0.029
	3200	400	16.235	16.628	25.577	26.051	36.004	36.563
			0.021	0.021	0.025	0.025	0.028	0.029
	1600	800	16.389	16.623	25.754	26.033	36.255	36.586
			0.021	0.021	0.025	0.025	0.029	0.029
	800	1600	16.506	16.640	25.881	26.041	36.409	36.600
			0.021	0.021	0.025	0.025	0.029	0.029
	400	3200	16.556	16.631	25.966	26.053	36.548	36.653
			0.021	0.021	0.025	0.025	0.029	0.029
	200	6400	16.603	16.643	25.990	26.036	36.619	36.676
			0.021	0.021	0.025	0.025	0.029	0.029
	100	12800	16.578	16.599	25.977	26.001	36.608	36.637
			0.021	0.021	0.025	0.025	0.029	0.029
18	12800	100	15.581	16.592	24.850	26.094	35.160	36.611
			0.020	0.021	0.024	0.025	0.028	0.029
	6400	200	15.975	16.597	25.345	26.105	35.693	36.584
			0.020	0.021	0.025	0.025	0.028	0.029
	3200	400	16.256	16.638	25.652	26.117	36.091	36.635
			0.021	0.021	0.025	0.025	0.028	0.029
	1600	800	16.445	16.672	25.812	26.088	36.331	36.656
			0.021	0.021	0.025	0.025	0.029	0.029
	800	1600	16.518	16.649	25.964	26.120	36.434	36.621
			0.021	0.021	0.025	0.025	0.029	0.029
	400	3200	16.576	16.648	26.029	26.115	36.572	36.676
			0.021	0.021	0.025	0.025	0.029	0.029
	200	6400	16.580	16.619	26.082	26.128	36.548	36.604
			0.021	0.021	0.025	0.025	0.029	0.029
	100	12800	16.640	16.661	26.066	26.090	36.598	36.627
			0.021	0.021	0.025	0.025	0.029	0.029

Table 3.5: (High-biased Estimators) Computing time.

N	m	n	$S_0 = \$90$ Original	Corrected	$S_0 = \$100$ Original	Corrected	$S_0 = \$110$ Original	Corrected
3	12800	100	20.428	41.312	22.991	46.399	19.536	39.254
	6400	200	19.483	39.804	19.524	39.615	19.419	39.559
	3200	400	19.593	39.889	19.918	40.323	19.500	39.457
	1600	800	19.745	39.326	19.545	39.917	19.551	39.095
	800	1600	19.482	39.626	19.410	39.440	19.198	39.036
	400	3200	19.890	39.745	19.197	39.205	19.182	39.451
	200	6400	19.374	39.605	19.597	39.489	19.677	39.796
	100	12800	19.314	39.467	19.641	39.336	19.380	39.675
6	12800	100	38.088	88.626	38.087	89.392	44.360	104.110
	6400	200	37.950	89.103	39.045	89.369	38.468	89.589
	3200	400	38.356	88.875	38.026	89.463	39.652	90.119
	1600	800	38.298	88.970	38.672	89.033	38.483	90.858
	800	1600	38.424	89.860	38.370	90.003	39.605	90.486
	400	3200	37.841	89.742	38.816	89.822	39.084	89.259
	200	6400	38.278	89.346	38.510	89.571	38.501	89.612
	100	12800	43.938	105.450	39.011	88.888	38.082	88.770
9	12800	100	61.438	151.074	58.462	144.659	59.396	144.208
	6400	200	59.809	144.662	58.859	144.270	59.001	144.672
	3200	400	59.433	143.875	59.103	144.975	58.750	143.329
	1600	800	66.340	157.374	59.484	144.642	59.427	143.854
	800	1600	62.923	154.803	58.298	143.955	58.563	143.974
	400	3200	59.468	143.688	65.325	161.229	58.730	145.174
	200	6400	61.440	149.609	58.706	142.814	59.746	145.012
	100	12800	59.667	146.078	59.053	143.993	60.196	143.408

N	m	n	$S_0 = \$90$ Original	$S_0 = \$90$ Corrected	$S_0 = \$100$ Original	$S_0 = \$100$ Corrected	$S_0 = \$110$ Original	$S_0 = \$110$ Corrected
12	12800	100	79.358	196.081	78.616	195.841	84.347	207.965
	6400	200	78.249	195.787	78.682	196.201	81.939	203.675
	3200	400	78.202	196.049	78.955	195.338	80.010	195.623
	1600	800	78.671	196.338	78.556	197.296	79.493	197.765
	800	1600	86.698	212.317	78.368	195.956	77.957	195.556
	400	3200	79.024	195.126	85.092	210.836	77.875	194.479
	200	6400	79.461	197.742	79.693	196.949	78.674	195.867
	100	12800	78.906	195.076	82.403	203.230	85.131	213.160
15	12800	100	98.843	251.510	99.894	248.997	99.848	248.906
	6400	200	97.938	245.586	98.534	248.210	99.067	250.462
	3200	400	100.199	260.430	96.596	248.522	96.989	248.659
	1600	800	105.705	264.325	97.770	245.596	98.076	249.995
	800	1600	100.889	245.963	105.426	262.567	96.827	245.489
	400	3200	99.206	251.388	97.769	245.360	108.590	268.902
	200	6400	101.437	252.542	99.020	252.274	96.757	247.403
	100	12800	101.819	250.222	98.750	251.696	101.798	258.553
18	12800	100	124.146	319.230	118.117	297.996	115.732	299.456
	6400	200	118.279	300.921	120.130	307.678	117.071	299.649
	3200	400	116.867	302.079	121.913	310.373	116.243	300.470
	1600	800	117.251	298.664	120.649	306.439	124.897	313.566
	800	1600	116.784	299.768	116.470	300.003	117.324	301.786
	400	3200	125.721	313.665	118.728	300.000	117.811	297.906
	200	6400	120.486	306.003	119.138	309.168	116.603	300.992
	100	12800	119.601	303.720	122.759	309.822	120.939	299.882

Table 3.6: (Low-biased Estimators) Computing time.

N	m	n	$S_0 = \$90$		$S_0 = \$100$		$S_0 = \$110$	
			Original	Corrected	Original	Corrected	Original	Corrected
3	12800	100	37.967	64.143	42.948	72.857	36.722	63.208
	6400	200	37.069	62.765	36.671	62.524	37.093	62.575
	3200	400	37.003	62.492	37.080	63.400	37.224	62.887
	1600	800	36.461	62.352	36.807	62.592	36.590	62.986
	800	1600	36.467	62.547	36.283	62.289	36.783	62.465
	400	3200	36.264	62.248	36.972	63.211	36.685	63.015
	200	6400	36.804	62.648	37.128	62.554	37.120	62.939
	100	12800	37.247	62.213	36.732	61.970	36.583	62.527
6	12800	100	75.476	142.840	75.410	142.460	91.051	170.828
	6400	200	76.282	142.091	75.570	142.653	76.139	142.003
	3200	400	75.703	143.828	76.389	142.055	76.372	146.169
	1600	800	75.226	141.622	76.178	141.107	78.052	145.337
	800	1600	75.660	142.186	75.700	141.005	76.486	144.850
	400	3200	76.140	142.787	75.404	141.471	75.837	142.814
	200	6400	75.703	142.020	75.371	142.182	75.990	142.125
	100	12800	89.829	165.839	76.050	140.709	74.316	144.005
9	12800	100	124.198	237.381	117.263	230.360	118.663	230.224
	6400	200	119.065	231.423	119.182	231.135	118.637	230.347
	3200	400	118.978	231.244	119.511	230.203	118.534	230.455
	1600	800	133.672	263.208	117.642	230.150	119.608	231.271
	800	1600	125.725	242.707	118.362	230.352	118.416	231.559
	400	3200	118.902	231.658	131.761	253.789	118.922	229.694
	200	6400	124.273	240.614	118.859	231.250	118.609	231.310
	100	12800	120.550	236.395	119.332	230.291	117.266	230.894

N	m	n	$S_0 = \$90$		$S_0 = \$100$		$S_0 = \$110$	
			Original	Corrected	Original	Corrected	Original	Corrected
12	12800	100	159.736	312.427	159.759	313.859	173.696	339.196
	6400	200	160.803	314.053	158.246	314.264	166.038	324.236
	3200	400	157.984	314.062	159.102	312.107	158.673	312.721
	1600	800	158.371	312.697	158.223	311.523	161.787	318.510
	800	1600	175.717	341.104	158.242	312.825	158.978	312.574
	400	3200	158.858	321.361	172.151	341.307	157.948	313.049
	200	6400	161.429	315.243	160.595	314.001	159.883	315.331
	100	12800	158.461	313.959	167.717	328.052	168.754	334.826
15	12800	100	196.761	397.130	199.890	398.409	203.155	398.300
	6400	200	200.030	399.722	200.987	395.702	202.045	396.765
	3200	400	211.877	411.753	200.404	394.858	196.090	399.365
	1600	800	211.551	420.633	201.021	396.307	199.465	405.413
	800	1600	195.952	398.082	211.700	425.664	204.477	394.845
	400	3200	202.723	401.004	202.658	394.681	211.772	417.592
	200	6400	197.450	389.254	201.648	399.017	201.495	396.147
	100	12800	197.046	392.004	199.300	395.488	205.832	407.409
18	12800	100	252.205	508.979	238.559	478.312	238.365	480.117
	6400	200	239.852	480.265	245.378	494.471	236.854	480.096
	3200	400	239.117	478.623	250.533	496.318	240.435	479.509
	1600	800	240.046	479.235	245.973	487.578	250.740	505.154
	800	1600	239.203	478.819	237.515	478.930	241.316	478.071
	400	3200	252.761	506.816	238.358	480.883	238.945	476.829
	200	6400	247.719	494.894	244.741	492.721	238.743	478.983
	100	12800	242.489	484.816	246.952	494.995	238.537	479.420

Bibliography

[1] Andersen, L., and Broadie, M. (2004) *A primal-dual simulation algorithm for pricing multi-dimensional American options*. Management Science **50** (9): 1222-1234.

[2] Avramidis, A.N., and Hyden, P. (1999) *Efficiency improvements for pricing American options with a stochastic mesh*. Proceedings of the 1999 Winter Simulation Conference, 344-350.

[3] Barraquand, J., and Martineau, D. (1995) *Numerical valuation of high-dimensional multivariate American securities*. Journal of Financial and Quantitative Analysis **30** (3): 383-405.

[4] Broadie, M., and Glasserman, P. (1995) *A pruned and bootstrapped American option simulation*. Proceedings of the 1995 Winter Simulation Conference, 229-235.

[5] Broadie, M., and Glasserman, P. (1997) *Pricing American-style securities using simulation*. Journal of Economic Dynamics and Control **21** (8-9): 1323-1352.

[6] Broadie, M., and Glasserman, P. (2004) *A stochastic mesh method for pricing high-dimensional American options*. The Journal of Computational Finance **7** (4): 35-72.

[7] Carrière, J.F. (1996) *Valuation of the early-exercise price for options using simulations and nonparametric regression*. Insurance: Mathematics and Economics **19** (1): 19-30.

[8] Clèment, E., Lamberton, D., and Protter, P. (2002) *An analysis of a least squares regression method for American option pricing*. Finance and Stochastics **6** (4): 449-471.

[9] Glasserman, P. (2004) *Monte Carlo Methods in Financial Engineering*. Springer-Verlag, New York.

[10] Glasserman, P., and Yu, B. (2004) *Number of paths versus number of basis functions in American option pricing*. The Annals of Applied Probability **14** (4): 2090-2119.

[11] Haugh, M., and Kogan, L. (2004) *Pricing American options: A duality approach*. Operations Research **52** (2): 258-270.

[12] Juneja, S., and Kalra, H. (2009) *Variance reduction techniques for pricing American options using function approximations*. The Journal of Computational Finance **12** (3): 79-102.

[13] Kan, K.H.F., and Reesor, R.M. (2010) *Bias reduction for pricing American options by least-squares Monte Carlo*. Submitted.

[14] Kan, K.H.F., Reesor, R.M., Whitehead, T., and Davison, M. (2009) *Correcting the bias in Monte Carlo estimators of American-style option values*. Monte Carlo and Quasi-Monte Carlo Methods 2008, ed. P. L'Ecuyer and A. B. Owen, Springer-Verlag, 439-454.

[15] Lemieux, C., and La, J. (2005) *A study of variance reduction techniques for American option pricing*. Proceedings of the 37th conference on Winter simulation, 1884-1891.

[16] Longstaff, F.A., and Schwartz, E.S. (2001) *Valuing American options by simulation: A simple least-squares approach*. Review of Financial Studies **14** (1): 113-147.

[17] Moreni, N. (2004) *A variance reduction technique for American option pricing*. Physica A: Statistical Mechanics and its Applications **338** (1-2): 292-295.

[18] Rogers, L.C.G. (2002) *Monte Carlo valuation of American options*. Mathematical Finance **12** (3): 271-286.

[19] Staum, J. (2009) *Monte Carlo Computation in Finance*. Monte Carlo and Quasi-Monte Carlo Methods 2008, ed. P. L'Ecuyer and A. B. Owen, Springer-Verlag, 19-42.

[20] Stentoft, L. (2004) *Convergence of the least squares Monte Carlo approach to American option valuation*. Management Science **50** (9): 1193-1203.

[21] Tilley, J.A. (1993) *Valuing American options in a path simulation model*. Transactions of the Society of Actuaries **45**: 83-104.

[22] Tsitsiklis, J., and Van Roy, B. (2001) *Regression methods for pricing complex American-style options*. IEEE Transactions on Neural Networks **12** (4): 694-703.

[23] White, H. (2001) *Asymptotic Theory for Econometricians: Revised Edition*. Academic Press.

[24] Whitehead, T. (2008) *Correcting the Monte Carlo optimal-stopping bias*. PhD Thesis. University of Western Ontario.

[25] Whitehead, T., Reesor, R.M., and Davison, M. (2008a) *A bias-reduction technique for Monte Carlo pricing of early-exercise options*. Submitted.

[26] Whitehead, T., Reesor, R.M., and Davison, M. (2008b) *Correcting the Monte Carlo optimal-stopping bias*. Submitted.

Chapter 4

Optimized least-squares Monte Carlo for measuring counterparty credit exposure of American-style options

4.1 Introduction

In this chapter, we would like to consider the risk management aspects of American options, as opposed to valuation of these contracts in the previous two chapters. Particularly, we investigate the counterparty risk evolved in trading OTC derivatives, which is carried out directly between two parties without going through an exchange or other intermediaries. Counterparty risk is the risk that a party to an OTC derivatives contract may fail to perform on its obligations, causing losses to the other party. This chapter focuses on the methodology for estimating the counterparty exposure of American options. The counterparty exposure here means the potential loss to the buyer of an American option resulting from a naked position in the option. To be more precise, it is the larger of zero and the market value of the option if the counterparty were to default and there were zero recovery.

---

[5]A version of this chapter has been accepted for publication [3].

Because our goal is to develop an efficient algorithm that is applicable to most financial instruments, including those that have many underlying risk factors, we concentrate exclusively on the Monte Carlo approach as it can overcome the curse of dimensionality. In order to find the quantile of the exposures at a future time point, we need to estimate the market value of the option at every simulated state at that specific time point. Since the valuation of American options is done by simulation, a straightforward implementation of the Monte Carlo approach for estimating counterparty exposures requires nested simulations, which is computationally formidable, especially when the risk horizon is long. Thus, a more efficient simulation algorithm for estimating counterparty exposures is in demand.

There is a rich literature in pricing American options by simulation. In particular, Longstaff and Schwartz [6] propose the least-squares Monte Carlo (LSM) method which is very popular in practice since it is very easy to implement and readily applicable to various options that contain an early-exercise feature. Thus, LSM serves as a natural starting point to measure the counterparty exposure of American options.

There are two main components in calculating the credit exposure of an American option: continuation value function (CVF) estimation and exposure valuation. Suppose we are interested in calculating the credit exposures at each of the possible exercise times. Then, we need to know the risk-neutral option values across a wide range of states of the underlying risk factors at each possible exercise time. Putting this into the LSM framework, we obtain the risk-neutral values of the option by estimating the CVFs through a cross-sectional linear regression at each possible exercise time. The sample paths for this estimation purpose are generated using evolution models of the underlying risk factors under the risk-neutral measure. For valuing credit exposure, however, we simulate the sample paths under the physical measure, since our concern is on the potential future market value of the option. The continuation value is estimated by inserting the realized values of the risk factors into the estimated CVF. The exposure is given by the maximum of the continuation value and the immediate exercise value. The future exposures are set to zero once the option is exercised. We term this version

of LSM that is tailored to calculate credit exposure the optimized least-squares Monte Carlo (OLSM) method.

While LSM usually yields a reasonably accurate estimate for the option price today (using a moderate sample size, say 10000), the estimated CVFs at future exercise opportunities are generally less accurate, resulting in possible incorrect exercise decisions, which in turn produces misleading credit exposures. For instance, the future exposures are set to be zero once the option is determined to be exercised by the LSM algorithm. If the option actually should not be exercised, that implies its exposure is grossly underestimated. To minimize such error, we apply variance reduction techniques, initial state dispersion and multiple bucketing (piecewise linear regression) to OLSM.

We test the OLSM method on a simple American call option as there exists an analytical approximation for it in Ju and Zhong [2]. It is assumed that the underlying stock does not distribute dividends. By no arbitrage theory, it is never optimal to exercise this kind of option early. Thus, its exposure is very sensitive to incorrect exercise decisions as a result of simulation error or regression model error in the stopping time estimators. We do not address the important issue of suboptimal exercise decisions due to other factors such as pricing model error or investor behaviour. Exposure estimates due to these other factors can be above or below the actual exposure. The main reason for this choice is that the exposures are often underestimated in case the option is mistakenly exercised early, which means that the average exposures likely fall under the graph of the analytical option prices. This allows for easy observation of the bias in exposures. On the contrary, for other American options whose optimal stopping times are before maturity, incorrect exercise decisions could lead to over- or under-estimations, causing the average exposures to appear above or below the analytical option prices. Moreover, the positive and negative biases due to incorrect exercise decisions could offset each other. These make it hard to observe the significance of the bias. Therefore, we pick the simple American call option as a benchmark instrument to evaluate the accuracy of the continuation value function approximations. Numerical results show that OLSM works well for the call option even

when the underlying's volatility is as high as 80%, but further research on multiple bucketing is required in order to generalize this method to handle various financial instruments.

The rest of this chapter is organized as follows. The next section sets up the OLSM framework for calculating credit exposures, entailing the two main components aforementioned. Section 4.3 is the heart of this chapter, which develops several techniques to improve the performance of OLSM. We provide numerical results for an American call option to demonstrate the effectiveness of OLSM in Section 4.4. Section 4.5 concludes and suggests future research directions.

4.2   OLSM Framework

OLSM consists of two phases. Phase one estimates the CVF at each exercise opportunity, whereas phase two makes use of the estimated CVFs to calculate credit exposure to the counterparty. Assume that we are working under the constraint that the sample paths are generated externally, and all the sample paths are used to maximize the performance of OLSM. Specifically, $M_1 = 10000$ paths are simulated under the risk-neutral measure and $M_2 = 10000$ paths are simulated under the physical measure. This fixed sample size will affect our selection of techniques that accelerate the convergence of the continuation value estimators. The risk-neutral prices and physical prices are used to estimate the CVFs and the credit exposures, respectively.

In this chapter, we assume the lognormal model for the underlying physical prices

$$S(t) = S_0 \exp\left[(\mu - \frac{1}{2}\sigma^2)t + \sigma W(t)\right] \tag{4.1}$$

and the underlying risk-neutral prices

$$S(t) = S_0 \exp\left[(r - \frac{1}{2}\sigma^2)t + \sigma W^*(t)\right] \tag{4.2}$$

where $W(t)$ is a standard Brownian motion under the physical measure, $W^*(t)$ is a standard

Brownian motion under the risk-neutral measure, S_0 is the initial stock price, μ is the drift, r is the risk-free interest rate, and σ is the volatility. Note that the S_0 in (4.1) could be different from that in (4.2) as we disperse the initial states in Section 4.3.2. The ideas developed in this chapter are applicable to more general price process models (e.g., stochastic volatility, jump diffusions).

4.2.1 CVF Estimation

In [6], Longstaff and Schwartz develop an LSM interleaving estimator, where the determiner (the estimator used to make the exercise/hold decision) and the propagator (option value passed on to preceding exercise opportunity) are dependent through one overlapping path. This overlap causes the estimator to be biased high for a relatively small sample size, since (a) there exists a Jensen's inequality effect, and (b) it makes use of future information to make exercise decisions, violating the definition of a stopping time. However, it converges to a lower bound for the true option value as the sample size gets large in which case the determiner tends to a constant and defines a valid stopping time.

American option pricing is an optimal stopping time problem. Transforming it into a dynamic programming problem and running it on the M_1 risk-neutral price paths, the recursive equations for the interleaving estimator are

$$\tilde{H}_k^i = x_k^i \tilde{\beta}_k \tag{4.3}$$

$$\hat{H}_k^i = e^{-r\Delta T} \tilde{B}_{k+1}^i \tag{4.4}$$

and

$$\tilde{B}_k^i = \begin{cases} \hat{H}_k^i & \text{if } \tilde{H}_k^i > P_k^i, \\ P_k^i & \text{if } \tilde{H}_k^i \leq P_k^i, \end{cases} \tag{4.5}$$

where x_k^i is a $(1 \times p)$ vector of basis functions evaluated at time k for path i, $\tilde{\beta}_k$ is a $(p \times 1)$ vector of least-squares regression coefficients, p is the number of basis functions, $\tilde{H}_k^i$(determiner) and

$\hat{H}_k^i$(propagator) are the time-k, path-i continuation value estimators, $\tilde{B}_k^i$ is the time-k, path-i option value estimator, P_k^i is the time-k, path-i exercise value, $\Delta T = T/N$ is the time interval, T is the maturity, $k \in \{1, \ldots, N\}$ is shorthand for $k\Delta T$ which is a possible exercise time, and the terminal conditions are given by $\tilde{H}_N^i = \hat{H}_N^i = 0$ for all $i \in \{1, \ldots, M_1\}$.

The basis functions that we will use are monomials up to the 3rd degree, that is, $x = \left[1, S, S^2, S^3\right]$. The only results from dynamic programming that we need to calculate the exposures are $\tilde{\beta}_k$, which are equal to $(\mathbf{X_k}'\mathbf{X_k})^{-1}\mathbf{X_k}'\hat{\mathbf{H}}_\mathbf{k}$, where $\mathbf{X_k} = ((x_k^1)', \ldots, (x_k^{M_1})')'$, $\hat{\mathbf{H}}_\mathbf{k} = (\hat{H}_k^1, \ldots, \hat{H}_k^{M_1})'$, $^{-1}$ denotes the matrix inverse and $'$ denotes the matrix transpose.

4.2.2 Exposure Valuation

The exposures are measured based on the simulated physical prices. At each time on each path, we can calculate the immediate exercise value and the value of all the basis functions using the physical prices, where the product of the latter and the previously estimated regression coefficients gives the continuation value. The exposure is equal to the maximum of the immediate exercise value and the continuation value. Once the option is exercised, the exposures at the later time points are all set to be zero. In practice, however, the exposure should only be set to zero after an appropriate close-out period. Therefore, the actual exposure is slightly underestimated by our algorithm.

There are some subtle issues that are not considered in our exposure estimation approach. For instance, when a firm buys an option with no margining, this can bring about additional exposure due to suboptimal decisions by the firm, possibly resulting in higher amounts of exposures for unexercised options. However, where there is a margining agreement, an unexercised sold option can produce additional counterparty (the writer of the option) exposure arising from unreturned margin. It could lead to an increase in bilateral exposures if the derivative is something like a swaption that can generate profit and loss of either sign after exercise.

4.3   Beyond LSM

We would like to highlight the fact that incorrect exercise decisions will result in misleading exposures due to the practice of setting the exposures at the later points to be zero after exercise. Thus, it is highly important to estimate the continuation value functions (characterized by β_k) corresponding to the physical prices accurately. However, the original LSM method does not produce CVFs that are accurate for a wide range of underlying prices due to extrapolation of the fitted regression beyond the data values used to estimate β_k. This might result in incorrect exercise decisions especially at future times in which real-world values of the underlying price are likely to be outside of the range of risk-neutral prices used to estimate β_k. To remedy this problem, several techniques are proposed in this section.

4.3.1   Faster Convergence to the Approximation

Given that the set of basis functions have been fixed, we would like to accelerate the convergence of the continuation value estimators to the true approximation provided by this set of basis functions. In other words, our aim is to reduce the simulation error. This can be achieved by bias and/or variance reduction.

Bias Reduction

The probabilistic nature of simulation makes incorrect exercise decisions (w.r.t. the true approximation) possible, giving rise to estimator bias. Kan et al. [5] develop a bias correction technique for Monte Carlo pricing of early-exercise options. Some numerical results for the LSM high- and low-biased estimators are given in Chapter 3. Kan et al. [4] conduct a detailed analysis of the bias of the LSM estimators and provides comprehensive numerical results on these estimators and their bias-corrected versions. Although all the numerical results show that the bias correction technique is effective in reducing the bias of the LSM estimators, we do not apply this technique to calculate the exposures because we work with a sample size of 10000,

where there is, in general, little simulation bias to be corrected. This is evident in the numerical results for the pricing of a five-dimensional American max-call option [4].

Variance Reduction

Since bias is not a big issue for a sample size of 10000, we focus on increasing the speed of convergence by reducing the variance. We consider two common variance reduction techniques in the following.

(a) Antithetic variates

Antithetic variates can be applied to the sample paths generated under the physical measure. Specifically, here 5000 antithetic pairs of negatively correlated paths are simulated, and exposures for each path are estimated. At each possible exercise time, the average exposure for each antithetic pair is computed, resulting in 5000 values that form an empirical distribution of the exposures. We can then estimate the expected exposure or the value-at-risk (VaR) using this distribution. This method works well when the function to be estimated is monotonic.

(b) Inner control variates

Since we would like to reduce the variance of the continuation value estimators at the intermediate time steps rather than at time zero, we apply control variates to the response variables in the cross-sectional regressions. This kind of control variates was termed inner control variates in Broadie and Glasserman [1]. To simplify notations, we suppress the subscript k. Rewriting the determiner in (4.3) as a weighted average of the propagators gives

$$\tilde{H}^i = x^i \tilde{\beta} = \sum_{j=1}^{M} \omega(i,j)\hat{H}^j, \tag{4.6}$$

where $\omega(i,j)$ denotes the weight connecting path i at time k to path j at time $k+1$. With control variates , (4.6) becomes

$$\tilde{H}^i(\alpha) = \sum_{j=1}^{M} \omega(i,j)\hat{H}^j - \alpha \left(\sum_{j=1}^{M} \omega(i,j)Y^j - \sum_{j=1}^{M} \omega(i,j)\mathbb{E}\left[Y^j\right] \right) \tag{4.7}$$

$$= \sum_{j=1}^{M} \omega(i,j)\left(\hat{H}^j - \alpha\left(Y^j - \mathbb{E}\left[Y^j\right]\right)\right), \tag{4.8}$$

where $\alpha = \text{cov}(Y^j, \hat{H}^j)/\text{var}(Y^j)$. As a result, we adjust the propagators with the control variates before running a cross-sectional regression to obtain the variance-reduced determiners.

In our numerical study, we will use three martingale control variates, namely $e^{-r(\tau-k\Delta T)}S_\tau$, $e^{-(2r+\sigma^2)(\tau-k\Delta T)}S_\tau^2$ and $e^{-(3r+3\sigma^2)(\tau-k\Delta T)}S_\tau^3$, which are sampled at the estimated exercise times to make them have a higher correlation to the propagators. The higher the correlation the more significant the amount of variance reduction.

4.3.2 Improving the Approximation

The methods in Section 4.3.1 can only reduce the number of sample paths required to attain the same accuracy for the crude Monte Carlo estimators. They help the estimators converge faster to the true approximation for a given set of basis functions, but they do not produce a better approximation to the true value. In fact, not much work has been done to address the approximation error associated with using a finite set of basis functions in the literature. We provide part of a solution to this problem in the following.

In OLSM, the CVF is estimated by a simulated regression. Initial state dispersion deals with the simulation part, generating a wider support for regression and hence reducing extrapolation error when computing credit exposures. Multiple bucketing improves the fit of the regression model.

Initial State Dispersion

Determining the exercise strategy of an American option does not require the risk-neutral sample paths to have the same initial price. Thus, we disperse the initial state so as to improve the stability of the regressions. Now, the question is how to disperse the initial state.

Rasmussen [7] suggests simulation of the state variables from a fictitious initial time point prior to time zero and the original initial state using the discounted underlying asset prices, a martingale under the risk-neutral measure. The distribution at time 0 will reflect the volatility of the underlying assets while being centered at the original initial state. The problem with this method is that it is difficult to determine the fictitious initial time point. Preliminary numerical results showed that Rasmussen's method is not very stable for the purpose of exposure calculation.

Since our goal is to measure the exposure, we have to estimate the risk-neutral value of the option at the physical prices. The drift of the model for the physical prices is usually greater than that for the risk-neutral prices. This implies a demand for an accuracy of the CVF for a wide support of underlying asset prices. Instead of drawing initial states from a sampling distribution, we avoid introducing one more dimension of variation by deterministically allocating the initial states. This is to help ensure that there exist initial states in the target region from which the risk-neutral prices are simulated to cover a wide range of the physical prices at future times. On the contrary, drawing from a sampling distribution might result in initial states concentrated on a small region, which is not a desirable property because that does not make the support wide enough at future times for accurate estimation of the CVF.

We divide the initial states into three regions, [10, 80], [80, 280] and [280, 510]. 4000, 2000 and 4000 initial states are allocated to the three regions, respectively. Within each region, the initial states are chosen uniformly. The ratios 4:2:4 are chosen since we want a very accurate estimated CVF at the beginning and more simulated data points are necessary to capture the shape of the CVF for the widely distributed physical prices near the maturity. The region boundaries are determined by matching the mean of the risk-neutral prices with the mean plus

six standard deviations of the physical prices at T/N, $T/2$ and T, i.e,

$$S_0^Q e^{rk\Delta T} = S_0 e^{\mu k\Delta T} + 6S_0 e^{\mu k\Delta T} \sqrt{e^{\sigma^2 k\Delta T} - 1}, \qquad (4.9)$$

where S_0^Q is an initial state for the risk-neutral prices, k is equal to 1, 20 and 40, respectively. With the parameter values in our numerical study, the S_0^Q's are rounded off to 80, 280 and 510, respectively.

There are several remarks on our proposed method of initial state dispersion:

- There is no need to determine a fictitious initial time point.

- The way to choose the initial state regions is not very rigorous, but it yields satisfactory results.

- The ratios 4:2:4 are more like guidelines than rules.

- This method is independent of the financial instrument type. All the dispersion parameter values remain the same if various financial instruments share the same risk factors.

Multiple Bucketing

After generating a wide range of values of the regressors, our next target is to improve the fit of the regression model. This could be achieved by multiple bucketing, which is an alias for piecewise linear regression. For instance, the state values are divided into two regions (buckets) so that an independent regression can be run on each bucket. This is on contrary to using a single regression for all the state values. This approach is especially effective when it comes close to the maturity date because the true CVF is not like a polynomial there. On the other hand, the CVF is pretty smooth at the beginning. It is expected that a single regression model would provide a good fit of the simulated data points. Following these principles, the bucket boundaries for an American call option with strike price 40 are chosen as follows:

- In the first quarter of the maturity, the bucket boundary for the underlying stock prices is chosen to be 100 (rounded-off), which is calculated using the sum of the expected value and four times the standard deviation of the underlying stock prices,

$$S_0 e^{rk\Delta T} + 4S_0 e^{r\Delta T} \sqrt{e^{\sigma^2 k\Delta T} - 1}, \qquad (4.10)$$

  where $k = 5$. The interval $(0, 100)$ would cover most simulated stock prices, thus we essentially run a single bucket regression in the first quarter, although technically speaking, that is a two-bucket regression since another regression is performed over the interval $(100, \infty)$.

- After $T/4$, we use in-the-money (ITM) and out-of-the-money (OTM) buckets. In other words, the bucket boundary is the strike price.

Different from the initial state dispersion, the selection of the bucket boundaries is sensitive to the type of the financial instrument. While formula (4.10) still applies in general, it is hard to justify why the bucket boundary should be at the strike price after $T/4$ for instruments other than the simple American call option. One possible solution to this problem is to choose the "best" boundary out of several candidates according to a certain criterion. This is left for future research.

4.4   Numerical Results

The OLSM method is tested on a simple American call option where the underlying stock does not pay dividends. By no arbitrage theory, it can be shown that it is optimal not to exercise this option early. Thus, the estimated exposures of this option are very sensitive to incorrect exercise decisions. If they closely resemble the exposures calculated using the approximate formula given in Ju and Zhong [2], that indicates the estimated exposures are accurate. We consider a call option that has a maturity of two years, and a strike price of 40. The maturity
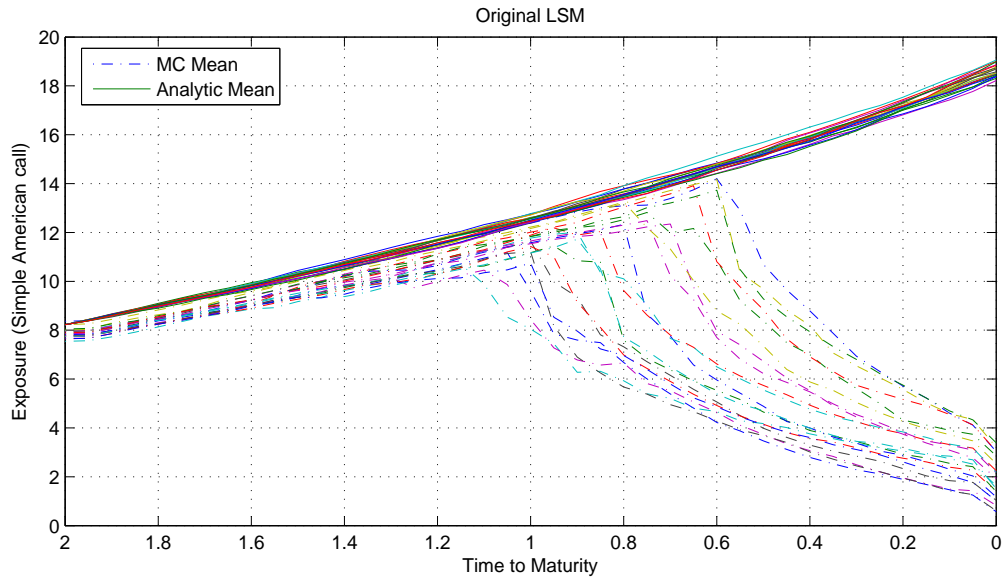
Figure 4.1: 20 average Monte Carlo and 20 average analytic exposures versus time-to-maturity. Volatility = 40%.

is divided into 40 time steps, so the time interval is 0.05 year. The risk-free interest rate is 6%. The initial underlying stock price is 36. The stock prices are governed by geometric Brownian motion with a drift of 20% and volatility of 40% for Figures 4.1-4.8. Figures 4.9-4.11 use a drift of 20% and volatility of 80%. Note that we used the volatility of 80% to derive the initial states regions and the bucket boundaries in Section 4.3.2. Nonetheless, the same results are applied to estimate the exposures for the option with the underlying's volatility equal to 40%.

All figures plot 20 average Monte Carlo (MC) exposures and 20 average analytical exposures, where the average is taken over 10000 paths. Figure 4.1 plots the original LSM exposures. It shows that the original LSM exposures are quite accurate at the beginning, but fall apart from the analytical exposures after half of the maturity. Figures 4.2-4.4 plot the Monte Carlo exposures with the use of antithetic variates, inner control variates, and the combination of both, respectively. It appears that either antithetic variates or inner control variates can reduce the variance of the exposures, but using them together gives the best result. However, the MC exposures still fall off after half of the maturity.

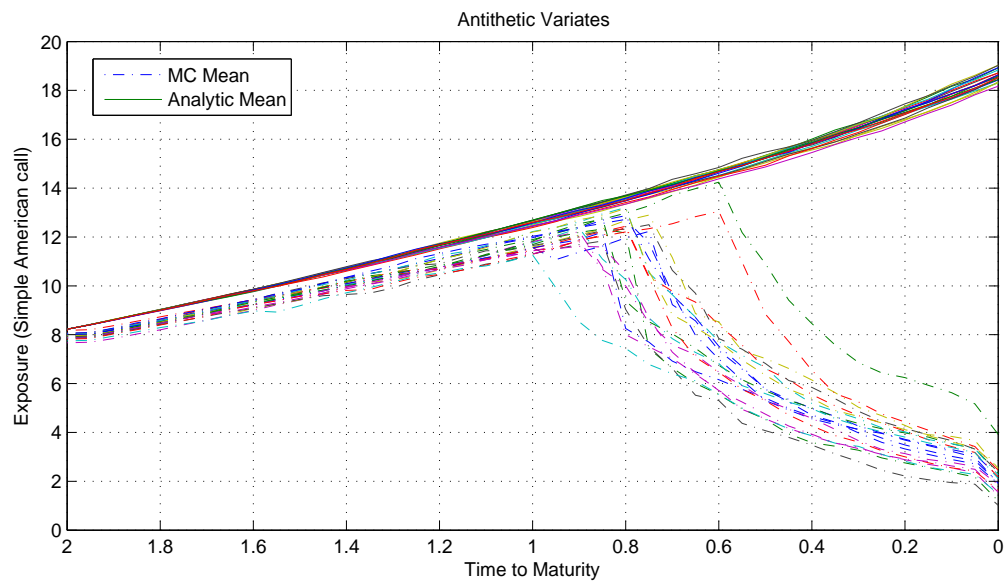In Figure 4.5, we use two buckets for regressions instead of one bucket, where the bucket

Figure 4.2: 20 average Monte Carlo and 20 average analytic exposures versus time-to-maturity. Antithetic variates are used on the "exposure" paths. Volatility = 40%.
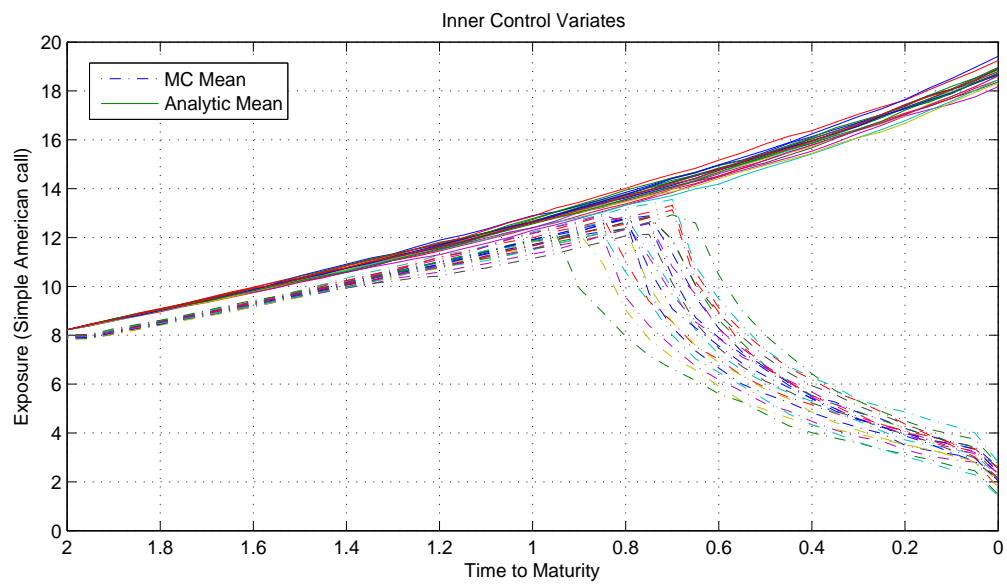

Figure 4.3: 20 average Monte Carlo and 20 average analytic exposures versus time-to-maturity. Inner control variates are used in estimating continuation value functions. Volatility = 40%.
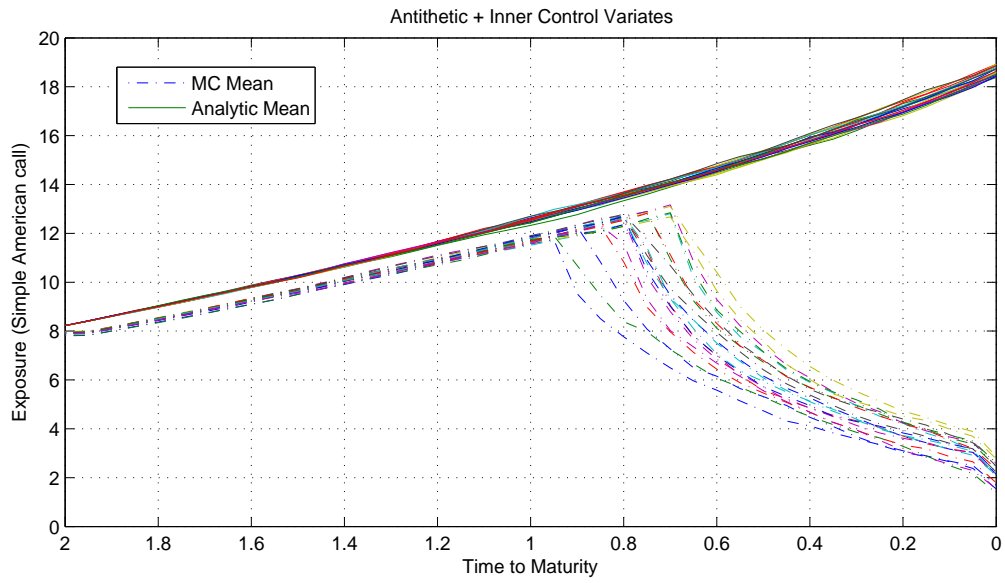
Figure 4.4: 20 average Monte Carlo and 20 average analytic exposures versus time-to-maturity. Both antithetic and inner control variates are used. Volatility = 40%.
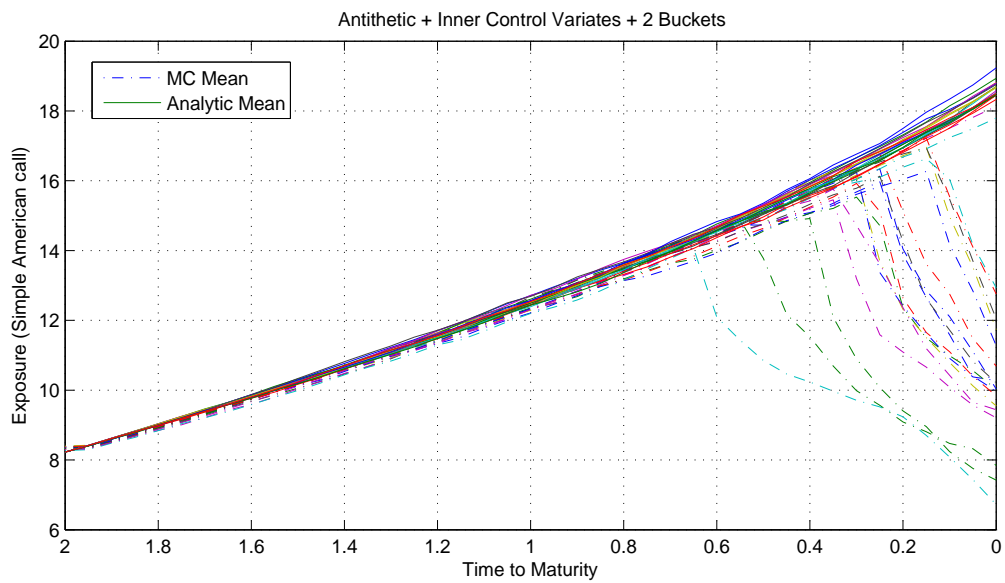

Figure 4.5: 20 average Monte Carlo and 20 average analytic exposures versus time-to-maturity. Both antithetic and inner control variates are used. Two buckets are used in the regression, where the bucket boundary is the strike price. Volatility = 40%.
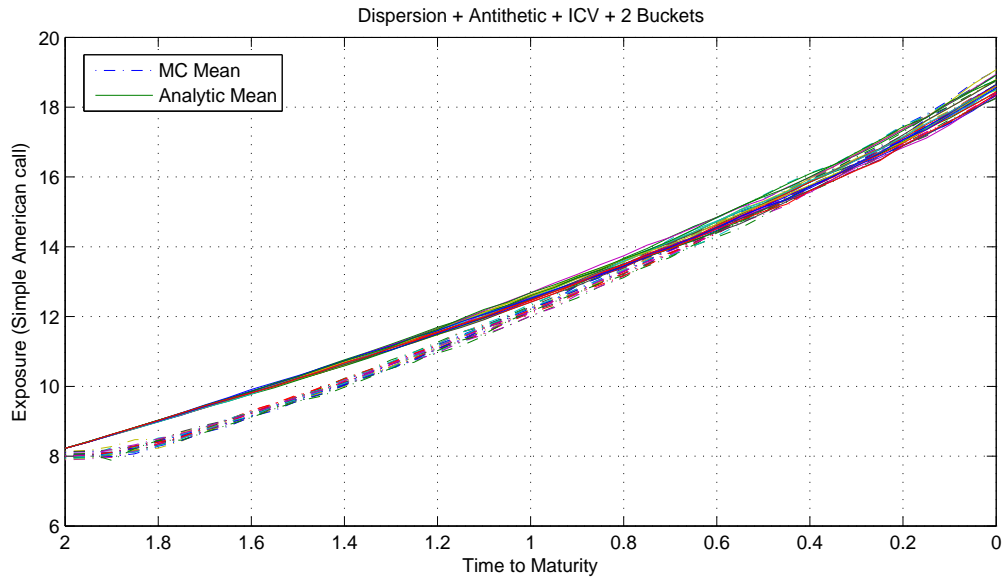
Figure 4.6: 20 average Monte Carlo and 20 average analytic exposures versus time-to-maturity. Both antithetic and inner control variates are used. Two buckets are used in the regression, where the bucket boundary is the strike price. Initial states are dispersed based on the ratio 4:2:4 to (10, 80), (80, 280) and (280, 510). Volatility = 40%.
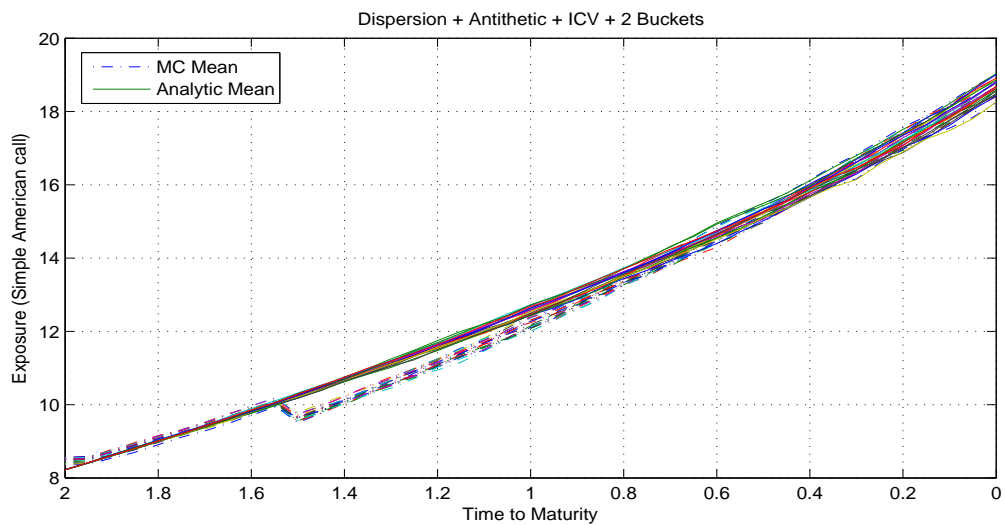

Figure 4.7: 20 average Monte Carlo and 20 average analytic exposures versus time-to-maturity. Both antithetic and inner control variates are used. Two buckets are used in the regression, where the bucket boundary is 100 for the first quarter of the maturity, and the strike price thereafter. Initial states are dispersed based on the ratio 4:2:4 to (10, 80), (80, 280) and (280, 510). Volatility = 40%.
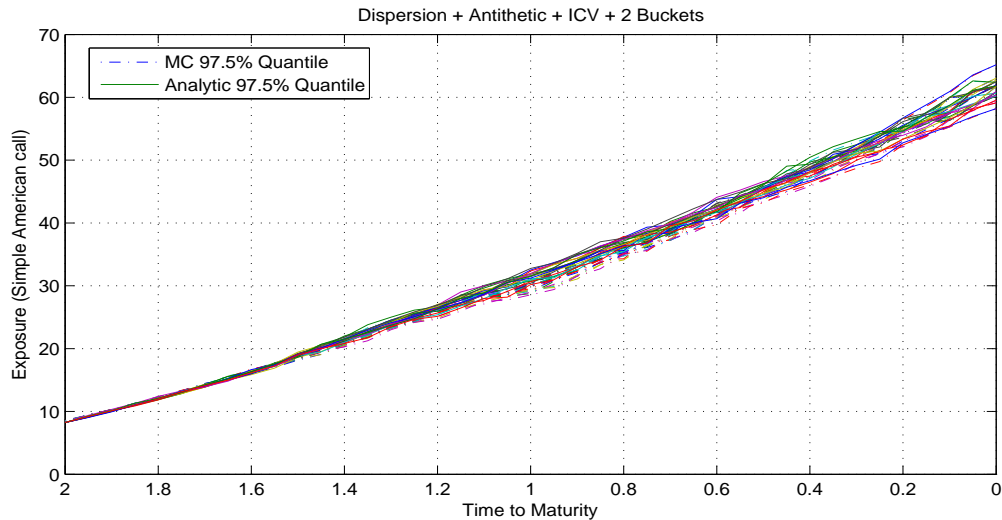
Figure 4.8: 20 97.5% quantile Monte Carlo and 20 97.5% quantile analytic exposures versus time-to-maturity. Both antithetic and inner control variates are used. Two buckets are used in the regression, where the bucket boundary is 100 for the first quarter of the maturity, and the strike price thereafter. Initial states are dispersed based on the ratio 4:2:4 to (10, 80), (80, 280) and (280, 510). Volatility = 40%.

boundary is the strike price. It shows that the accuracy of the exposures has been significantly improved — the MC exposures are "exact" in the first half and fall apart in the last quarter of the maturity. We incorporate the initial state dispersion in Figure 4.6. Amazingly, the MC exposures no longer fall apart from the analytical exposures, which indicates that very few or no incorrect exercise decisions have been made. However, this comes with the cost that the errors of the MC exposures at the beginning become significant. To get around this problem, the bucket boundary in the first quarter is changed to 100. Figure 4.7 shows that the MC exposures are very accurate after the change. Figure 4.8 demonstrates that the OLSM method works even better for the 97.5% quantiles than the average of the MC exposures.

We also plot the estimated exposures corresponding to the volatility of 80% in Figures 4.9-4.11. The results are similar to those with the volatility equal to 40%, qualitatively. The variance of the MC exposures is bigger as expected.

It is worth mentioning that although the initial states regions and the bucket boundaries were not designed for the case where volatility equals 40%, its results are excellent. This
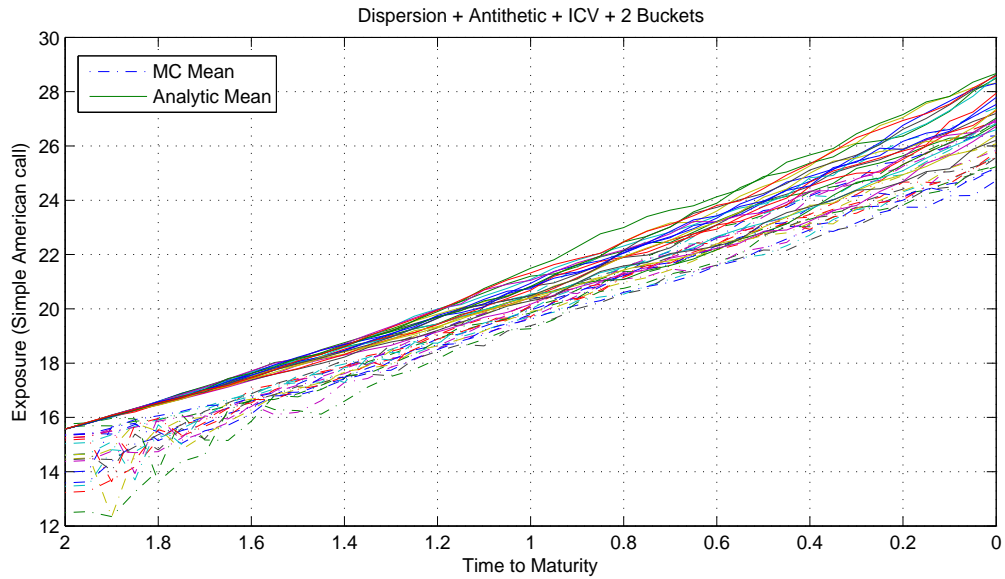
Figure 4.9: 20 average Monte Carlo and 20 average analytic exposures versus time-to-maturity. Both antithetic and inner control variates are used. Two buckets are used in the regression, where the bucket boundary is the strike price. Initial states are dispersed based on the ratio 4:2:4 to (10, 80), (80, 280) and (280, 510). Volatility = 80%.
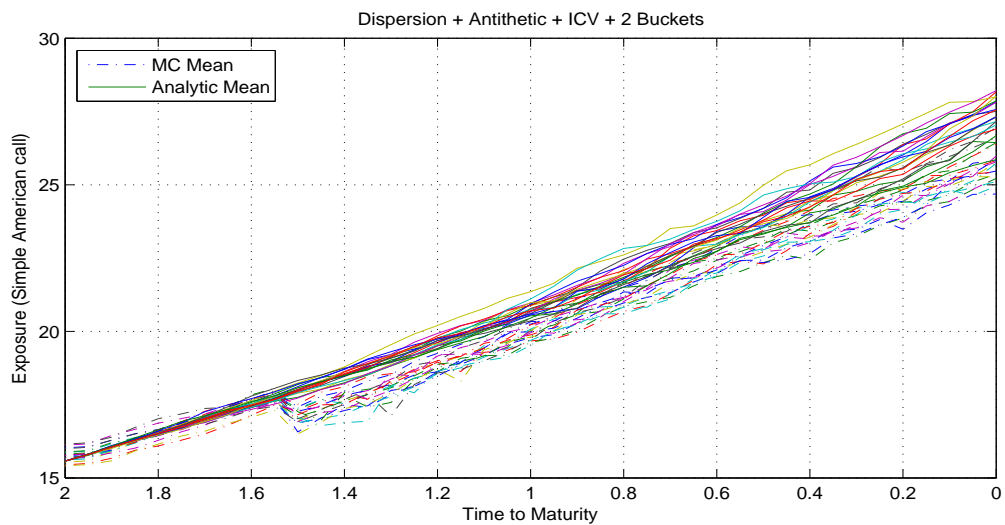

Figure 4.10: 20 average Monte Carlo and 20 average analytic exposures versus time-to-maturity. Both antithetic and inner control variates are used. Two buckets are used in the regression, where the bucket boundary is 100 for the first quarter of the maturity, and the strike price thereafter. Initial states are dispersed based on the ratio 4:2:4 to (10, 80), (80, 280) and (280, 510). Volatility = 80%.
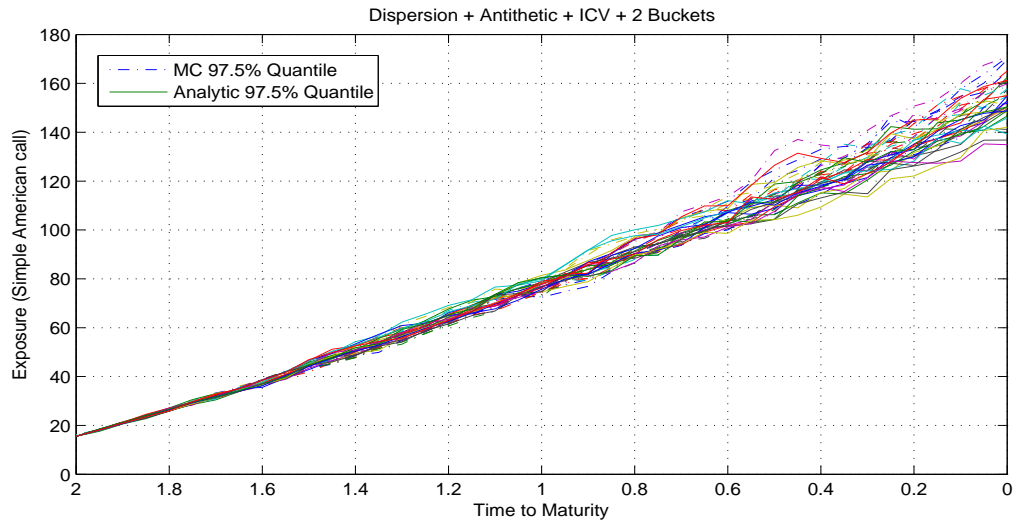
Figure 4.11: 20 97.5% quantile Monte Carlo and 20 97.5% quantile analytic exposures versus time-to-maturity. Both antithetic and inner control variates are used. Two buckets are used in the regression, where the bucket boundary is 100 for the first quarter of the maturity, and the strike price thereafter. Initial states are dispersed based on the ratio 4:2:4 to (10, 80), (80, 280) and (280, 510). Volatility = 80%.

implies that the estimation of the CVF is not very sensitive to the initial states regions and the bucket boundaries when the volatility of the underlying is low. Hence, we should use a high volatility to test the robustness of a new method to measure the exposures.

4.5   Conclusions and Future Work

The popular LSM method is extended for the purpose of measuring counterparty credit exposure of American-style options. We optimized the LSM method using variance reduction techniques, initial state dispersion and multiple bucketing, hence the name OLSM. Numerical results for a simple American call option indicate that OLSM can reduce the absolute relative error of the average or the 97.5% quantile of the exposures to less than 10%. All of the above techniques can be applied to other American options. In particular, the same initial states can be recycled for multiple options as long as they have some common risk factors. However, the way to determine the bucket boundaries is not very general. Thus, one direction of future

work is to develop a systematic approach to pick the bucket boundaries. Furthermore, we are interested in investigating the performance of the OLSM algorithm when the underlying of the option is governed by a volatility model, which allows for time-varying and state-dependent volatilities, including implied volatilities. This can result in additional exposure, or changes in early exercise decisions.

Bibliography

[1] Broadie, M., and Glasserman, P. (2004) *A stochastic mesh method for pricing high-dimensional American options*. The Journal of Computational Finance **7** (4): 35-72.

[2] Ju, N., and Zhong, R. (1999) *An approximate formula for pricing American options*. Journal of Derivatives **7** (2): 31-40.

[3] Kan, K.H.F., Frank, G., Mozgin, V., and Reesor, R.M. (2010) *Optimized least-squares Monte Carlo for measuring counterparty credit exposure of American-style options*. Mathematics-in-Industry Case Studies Journal **2**: 1-21.

[4] Kan, K.H.F., and Reesor, R.M. (2010) *Bias reduction for pricing American options by least-squares Monte Carlo*. Submitted.

[5] Kan, K.H.F., Reesor, R.M., Whitehead, T., and Davison, M. (2009) *Correcting the bias in Monte Carlo estimators of American-style option values*. Monte Carlo and Quasi-Monte Carlo Methods 2008, ed. P. L'Ecuyer and A. B. Owen, Springer-Verlag, 439-454.

[6] Longstaff, F.A., and Schwartz, E.S. (2001) *Valuing American options by simulation: A simple least-squares approach*. Review of Financial Studies **14** (1): 113-147.

[7] Rasmussen, N. (2005) *Control variates for Monte Carlo valuation of American options*. Journal of Computational Finance **9** (1): 84-102.

Chapter 5

Conclusions

5.1 Summary

We constructed a new regression-based Monte Carlo estimator for valuing American options. Our numerical results show that the estimates are typically high-biased and have small relative error. A bias reduction method was developed for the high- and low-biased LSM estimators of American option value. We used a numerical example to demonstrate that this method is effective for the low-biased estimator with the use of parallel implementations. There is not much computational gain for the high-biased estimator, though. Finally, we introduced the OLSM method to estimate the counterparty exposures of American options. We found that OLSM yields reasonably accurate exposure estimates while using substantially less computing time than nested simulations.

5.2 Future Directions

When implementing the alternative regression-based algorithm for American options valuation, we adopted an ad-hoc initial state dispersion scheme. It may be worthwhile conducting a deeper study in this aspect and finding a systematic method to determine the value of the dispersion parameter.

We had some success at the application of a bias reduction method to the LSM estimators. However, the LSM estimators are already quite efficient, leaving little room for improvement with the bias reduction method. It is much more promising to apply this method to the relatively less efficient stochastic mesh estimators, which have an advantage over the LSM estimators in that they are consistent to the true option value, while the LSM estimators converge to an approximation value given a finite set of basis functions. Hence, I would suggest that future theoretical and empirical work of the bias reduction method should be done in the context of the stochastic mesh estimators.

In the OLSM chapter, we focused on one numerical example: an American call option on an non-dividend paying stock. Extensive numerical studies are necessary to test the robustness of OLSM. A benchmark value is required for comparison with the exposure estimate, yet an analytical solution does not exist for most American options. One way around this problem is to test OLSM on a set of European options that can be priced analytically or with a fast numerical scheme. This is because our main concern with OLSM is the accuracy of the function approximations for the continuation values. We conjecture that an effective function approximation for a European option would be effective for the corresponding American option as well.

VITA

Name: Kin Hung (Felix) Kan

Post-secondary   The University of Western Ontario
Education and   London, ON, Canada
Degrees: 2007 – 2010 Ph. D. in Applied Mathematics

                 The University of British Columbia
                 Vancouver, BC, Canada
                 2005 – 2007 M. Sc. in Mathematical Finance

                 Hong Kong University of Science and Technology
                 Hong Kong, China
                 2001 – 2005 B. Sc. in Mathematics

Work   Research Assistant
Experience: Department of Applied Mathematics, UWO
                 London, ON, Canada
                 2007 – 2010.

                 Teaching Assistant
                 Department of Applied Mathematics, UWO
                 London, ON, Canada
                 2007 – 2010.

                 Intern
                 Analytics, Capital Markets Risk Management, CIBC
                 Toronto, ON, Canada
                 Jun – Oct 2009.

                 Intern
                 Energy Planning, BC Hydro
                 Vancouver, BC, Canada
                 May 2006 – Apr 2007.

                 Teaching Assistant
                 Department of Mathematics, UBC
                 Vancouver, BC, Canada
                 2005 – 2006.

Scholarships and Awards:

- *Ontario Graduate Scholarship*, Canada, 2010.

- *Graduate Thesis Research Award*, UWO, London, ON, Canada, 2010.

- *Western Graduate Research Scholarship*, UWO, London, ON, Canada, 2007 – 2010.

- *MITACS Internship Award*, Canada, 2009.

- *MITACS Internship Travel Award*, Canada, 2009.

- *Statistical Society of Canada Student Travel Award*, Canada, 2009.

Publications:

- K.H.F. Kan, G. Frank, V. Mozgin, and R. M. Reesor. *Optimized least-squares Monte Carlo for measuring counterparty credit exposure of American-style options.* Mathematics-in-Industry Case Studies Journal, volume 2, pp. 64-85, 2010.

- K.H.F. Kan, R. M. Reesor, T. Whitehead, and M. Davison. *Correcting the bias in Monte Carlo estimators of American-style option values*; In: P. L'Ecuyer, A.B. Owen (eds.), Monte Carlo and Quasi-Monte Carlo Methods 2008, pp. 439-454. Springer-Verlag, 2009.

Submitted manuscripts:

- K.H.F. Kan, and R. M. Reesor. *An alternative regression-based method for American option valuation*; submitted, 2010.

- K.H.F. Kan, and R. M. Reesor. *Bias reduction for pricing American options by least-squares Monte Carlo*; submitted, 2010.