3-10-2022 4:00 PM

Efficiency Improvements in the Least-Squares Monte Carlo Algorithm

François-Michel Boire, *The University of Western Ontario*

Supervisor: Stentoft, Lars, *The University of Western Ontario*
Co-Supervisor: Reesor, Mark, *Wilfrid Laurier University*
A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Statistics and Actuarial Sciences

Abstract

This thesis presents a collection of four essays dealing with the efficient pricing of American options with Monte Carlo simulation techniques. Specifically, we focus on developing and optimizing techniques to reduce pricing bias and variance in the Least-Squares Monte Carlo (LSM) algorithm of Longstaff and Schwartz (2001).

In the first chapter, we take advantage of the put-call symmetry property to improve the efficiency of variance reduction techniques applied to American call options. Indeed, it is always more efficient to apply variance reduction techniques to corresponding symmetric put options instead.

The second chapter introduces a new importance sampling approach that performs regressions on paths simulated under the importance probability measure directly. This enhances continuation value estimates by performing regressions on samples that have more density near the optimal exercise boundary. Results show that this method successfully reduces the bias plaguing the standard method where regressions are instead performed under the nominal probability measure.

The third chapter derives an approximation of local LSM estimator bias when exercise decisions and option values are mutually dependent. The derivations presented in this paper explicitly deal with foresight bias and hold true for general asset price processes and option payoff structures. We then introduce a new bias-corrected LSM estimator and demonstrate its robustness in the presence of multiple stochastic factors and price discontinuities.

Finally, the fourth chapter studies the effects of stopping time optimality on the variance of the LSM estimator. We show that exercise errors stifle the variance reduction efficiency gains of control variates and preclude an efficient combined implementation of control variates and importance sampling. Results indicate that the corrected strategy of Rasmussen (2005) substantially improves stopping time optimality, virtually eliminating bias and improving the variance reduction efficiency gains by orders of magnitude.

Throughout the four chapters, we provide an in-depth discussion about the causes leading to the LSM estimator bias and variance. Using these insights, we develop extensions of the LSM, detail their implementation, and make numerous recommendations aimed at improving performance and computational efficiency. The numerical studies in this thesis test the robustness of our proposed methods across a wide range of option characteristics.

Keywords: American Options, Computational Finance, Derivatives Pricing, Numerical Methods for Option Pricing, Monte Carlo Methods, Option Pricing via Simulation

Lay Audience Summary

Options are widely traded assets that find many applications in financial risk management. These derivative contracts stipulate that one party can trade an underlying asset, index or security with another party at a fixed price at a future date. If the option owner decides to exercise the option, the seller is then obligated to enter the agreed-upon trade. Alternatively, the owner may decide to forgo exercise if the trade is not profitable. Options thus offer limited-term insurance to the owner by guaranteeing that the underlying asset can be traded at a fixed price in the future.

Depending on contract modalities, the valuation of the options may be more or less difficult. In the simplest cases, options are exercisable solely when the contract expires. These options are termed "European options" and are relatively easy to price. Another class of options called "American options" allows exercise at any time before expiry, providing the owner with more flexible insurance. Contrary to European options, American options prices are difficult to estimate. This is because the European option exercise date is specified by the contract, whereas the American option exercise time is uncertain and must be approximated. In turn, the management and pricing of American options are complicated by the fact that one must continuously consider whether it is best to exercise the option or to keep it alive. Consequently, any actual numerical approach to this "optimal control problem" involves the joint estimation of the option price and the exercise strategy in a "dynamic programming algorithm".

This thesis focuses on the renowned Least-Squares Monte Carlo (LSM) algorithm of Longstaff and Schwartz (2001) for pricing American options. This simulation-based dynamic programming algorithm is flexible, easy to implement, and robust to various option pricing problems. However, implementing the LSM algorithm can be computationally costly, and its optimization remains an active area of research. Throughout the four chapters presented in this thesis, we focus on developing and optimizing techniques to reduce the bias and variance of LSM estimators. Extensive numerical studies test the robustness of our proposed methods across a wide range of option characteristics and underlying asset processes.

Co-Authorship Statement

I declare that the thesis has been composed by myself and that the work has not been submitted for any other degree or professional qualification. This thesis incorporates materials that are results of joint research conducted from September 2017 to present under the supervision of Dr. Lars Stentoft and the co-supervision of Dr. Mark Reesor in the Department of Statistical and Actuarial Sciences at the University of Western Ontario. My contribution and those of co-authors are explicitly indicated below.

The work presented in Chapter 1 was previously published in the Journal of Risk and Financial Management as "Efficient Variance Reduction for American Call Options Using Symmetry Arguments" by François-Michel Boire (student), R. Mark Reesor (co-supervisor), and Lars Stentoft (supervisor).

> Full citation: Boire, François-Michel, R. Mark Reesor, and Lars Stentoft. 2021. "Efficient Variance Reduction for American Call Options Using Symmetry Arguments" *Journal of Risk and Financial Management* 14, no. 11: 504.
> https://doi.org/10.3390/jrfm14110504

The work presented in Chapter 2 was previously published in the Journal of Risk and Financial Management as "American Option Pricing with Importance Sampling and Shifted Regressions" by François-Michel Boire (student), R. Mark Reesor (co-supervisor), and Lars Stentoft (supervisor).

> Full citation: Boire, François-Michel, R. Mark Reesor, and Lars Stentoft. 2021. "American Option Pricing with Importance Sampling and Shifted Regressions" *Journal of Risk and Financial Management* 14, no. 8: 340.
> https://doi.org/10.3390/jrfm14080340

The content of Chapter 3 is based on a paper in preparation titled: "Bias Correction in the Least-Squares Monte Carlo Algorithm" and co-authored by François-Michel Boire (student), R. Mark Reesor (co-supervisor), and Lars Stentoft (supervisor).

The content of Chapter 4 is based on a paper in preparation titled: "Monte Carlo Variance Reduction and American Option Exercise Strategies" and co-authored by François-Michel Boire (student), R. Mark Reesor (co-supervisor), and Lars Stentoft (supervisor).

I certify that this thesis, and the research to which it refers, is the product of my own work. The formulation and conceptualization of key ideas, coding, data curation, formal analysis, and writing were performed by myself. In all cases Lars Stentoft and Mark Reesor contributed feedback on refinement of ideas and editing of the manuscripts. All remaining errors are my own.

François-Michel Boire, Montreal, December 2021

Acknowledgements

I would first like to thank Dr. Marcos Escobar-Anel, Dr. Mike Ludkovski, Dr. Fredrik Ødegaard, and Dr. Hao Yu for agreeing to serve as members of my dissertation committee. I would also like to acknowledge the financial support I have received from the Ontario Graduate Scholarship (Government of Ontario) and the Graduate Fellowship (University of Western Ontario), as well as SHARCNET for providing computational resources.

I am particularly grateful to my supervisors Dr. Mark Reesor and Dr. Lars Stentoft for everything they taught me. The time spent with them was most valued.

Finally, I thank the friends that I met over the course of my doctoral studies, Alexandru Draghici, Elena Draghici, and Cory Walton for the good laughs and many interesting conversations. Special thanks go to my parents and my sister for their unwavering support, and to Catherine for being by my side.

Contents

List of Figures

List of Tables

List of Appendices

Introduction

The seminal papers of Black and Scholes (1973) and Merton (1973) are at the origin of a wide scope of research areas in option pricing theory and financial engineering. Their work gave rise to the flourishing global market of *financial options* (i.e., exchange-traded derivatives written on a traded asset or security) that we see today. Among those, options with early-exercise features, or American-style options, are certainly the most common, yet complicated derivatives to price and manage. Indeed, like European options, American options are ubiquitously used for risk-management and speculation purposes, but the problem of pricing American options belongs to a whole different class called "optimal stochastic control" or "optimal stopping time" problems. These problems cannot be solved easily, and to this day, no general closed-form solution exists to price American options. Therefore, practitioners must resort to numerical methods that approximate the optimal early-exercise strategy that is needed to price the option.

The first numerical approximations to American option prices were based on deterministic approaches. For instance, finite difference solutions to option pricing problems were used as early as in Brennan and Schwartz (1978). Binomial option pricing models were then formalized by Cox et al. (1979). These lattice- and tree-based solutions, though very accurate, can be impractical, as they quickly succumb to the curse of dimensionality when the dynamics of the underlying asset are governed by multiple stochastic factors or when the payoff is path-dependent. Other quasi-analytical approaches were proposed shortly thereafter by Geske and Johnson (1984), followed by MacMillan (1986) and Barone-Adesi and Whaley (1987). Subsequently, Omberg (1987) introduced an exponential approximation of the early-exercise boundary.

Up until this point, Monte Carlo simulation applications to option pricing were still limited to European option valuation problems (see Boyle (1977), for instance), and were deemed unfit to tackle the problem of pricing American options. That is until Tilley (1993) showed how to adapt a dynamic programming algorithm to a simulation framework, giving rise to a rich strand of the literature aimed at developing efficient simulation-based American option pricing algorithms. Notable examples include the stratified state aggregation technique of Barraquand and Martineau (1995) and the stochastic tree of Broadie and Glasserman (1997). We direct our focus to the class of regression-based approaches, initially described by Carriere (1996), and later developed by Tsitsiklis and Van Roy (2001) and Longstaff and Schwartz (2001). This thesis specifically studies the renowned Least-Squares Monte Carlo (LSM) algorithm introduced by Longstaff and Schwartz (2001). Among other regression-based approaches, the LSM algorithm is certainly the most popular because of its robustness and accuracy (Stentoft 2004*b*, 2014). Moreover, in contrast to the aforementioned deterministic solutions, regressions easily scale to multivariate settings and are well-adapted to general option payoff structures. However, implementing the LSM algorithm can still be computationally costly, and its optimization remains an active area of research.

This thesis contains four chapters dealing with the efficient implementation of the LSM algorithm. Throughout the four chapters, we provide an in-depth discussion about the causes leading to the LSM estimator bias and variance. Using these insights, we optimize existing variance reduction techniques and develop extensions of the LSM. We provide detailed methodological guidelines and make numerous recommendations aimed at improving perfor-

mance and computational efficiency. The numerical studies in this thesis test our proposed methods across a wide range of option and asset characteristics, demonstrating their efficiency and robustness.

In the first chapter, we take advantage of the put-call symmetry property to improve the efficiency of variance reduction techniques applied to American call options. Stentoft (2019) has recently shown that the estimated American call prices obtained with regression and simulation-based methods can be significantly improved by using put-call symmetry. This paper extends these results and demonstrates that it is also possible to significantly reduce the variance of the estimated call price by applying variance reduction techniques to corresponding symmetric put options. First, by comparing performance for pairs of call and (symmetric) put options for which the solution coincides, our results show that efficiency gains from variance reduction methods are different for calls and symmetric puts. Second, control variates should always be used and is the most efficient method. Furthermore, since control variates is more effective for puts than calls, and since symmetric pricing already offers some variance reduction, we demonstrate that drastic reductions in the standard deviation of the estimated call price are obtained by combining all three variance reduction techniques in a symmetric pricing approach. This reduces the standard deviation by a factor of over 20 for long-maturity call options on highly volatile assets. Finally, we show that our findings are not particular to using in-sample pricing but also hold when using an out-of-sample pricing approach.

The second chapter proposes a new method for pricing American options that uses importance sampling to reduce estimator bias and variance in simulation-and-regression-based methods. Our suggested method performs regressions under the importance measure directly instead of under the nominal measure as is the standard (Moreni 2003). Our numerical results show that this method successfully reduces the bias plaguing the standard importance sampling method across a wide range of moneyness and maturities, with negligible change to estimator variance. When a low number of paths is used, our method always improves on the standard method and reduces average root mean squared error of estimated option prices by 22.5%. The findings in Chapter 1 point out that the determination of an optimal change of drift comports a number of challenges. This question is more closely examined in the present chapter. We demonstrate that the saddle point approximation of Glasserman et al. (1999) is far from optimal for short-maturity options and for implementations that combine importance sampling with other variance reduction techniques, leaving open the possibility to further maximize the benefits of importance sampling.

An core aspect of variance reduction in the LSM is the mitigation of error propagation in the backward recursion of the dynamic program. To correct pricing errors occurring at different time periods, the third chapter derives an approximation of local LSM estimator bias when exercise decisions and option values are mutually dependent. Furthermore, a new bias-corrected estimator is introduced. In the LSM algorithm, the dependence between continuation values and future cash flows results in potential model overfitting generating positive foresight bias. We find an expression that measures the impact of overfitting on the bias of option values. Next, we assume that continuation values are normally distributed to derive a local foresight bias approximation. Finally, we propose a bias-corrected in-sample LSM estimator that holds true for general asset price processes and option payoffs. Numerical results show that our proposed method reduces overall estimator bias across a wide range of option characteristics. The relative importance of these improvements increases with the frequency of exercise opportuni-

ties, and when continuation value estimates use a small number of sample paths with a large number of basis functions.

Finally, the fourth chapter studies the effects of stopping time optimality on the variance of the LSM estimator. Using the insights provided in previous chapters, we examine how stopping time estimation errors cause bias and alter the variance of the estimator. In particular, this paper discusses how exercise errors stifle the variance reduction efficiency gains of control variates and preclude an efficient combined implementation of control variates and importance sampling. Results first show that the corrected strategy of Rasmussen (2005) substantially improves stopping time optimality, virtually eliminating bias and mitigating the effect of exercise errors on variance. We find that the control variates method reduces standard error by a factor of up to 230 with a corrected strategy compared to 12 for a similar estimator with an uncorrected strategy. Furthermore, adding importance sampling to control variates reduces standard errors by a factor reaching 375 with a corrected strategy instead of 32 with an uncorrected one. Finally, we argue that the efficiencies obtained with a corrected strategy are essentially maximized because they approach those obtained when using an optimal exercise strategy.

Chapter 1

Efficient Variance Reduction for American Call Options Using Symmetry Arguments

1.1 Introduction

Key to valuing American options with a dynamic programming approach is estimating a continuation value that determines the optimal exercise strategy. For this task, simulation and regression-based methods are nowadays often preferred to other deterministic algorithms, like finite differences and multinomial trees, because they are easy to implement and because of their flexibility. The Least-Squares Monte Carlo (LSM) method of Longstaff and Schwartz (2001) is particularly popular, and several arguments have been made in favor of this methodology for valuing American options (Stentoft 2014). However, like other Monte Carlo pricing methods the LSM method is numerically costly and reducing its variance is therefore important.

This paper examines the relationship between the efficiency of variance reduction techniques and option features like moneyness, maturity, and asset volatility when pricing American-style options with the LSM method. Whereas most of the American option pricing literature has focused on either put or call options individually, we employ the symmetry relation of McDonald and Schroder (1998) such that we can readily compare results for pairs of call and put options whose solutions coincide. Three classical variance reduction techniques are studied in the context of LSM pricing: antithetic sampling, control variates, and importance sampling. We also consider implementations where two or more variance reduction techniques are combined. We restrict our attention to these three techniques because of their popularity, and because they do not require simulating additional paths. These estimators retain the flexibility of a crude LSM implementation and can easily be combined with moment-matching simulation (MMS) of Barraquand and Martineau (1995), or empirical martingale simulation (Duan and Simonato 1998). Other notable variance reduction techniques include low-discrepancy sequences (Lemieux and La 2005), latin hypercube (Glasserman 2003), and stratified sampling (Glasserman et al. 1999). If several estimators are readily available, optimal linear combinations of these estimators can also be explored.

Our results first show that efficiency gains from variance reduction may be quite different for calls and symmetric puts. Moreover, control variates is by far the most efficient of the

three methods. We observe that efficiencies, defined as the ratio of the standard deviations of the crude LSM and the LSM with variance reduction, increase (or never worsen) with time to maturity and asset volatility for symmetric put options. Conversely, efficiencies decrease (or never improve) with time to maturity and asset volatility for call options. Next, since control variates is always more effective for puts than calls, and since symmetric pricing already offers some variance reduction, we demonstrate that a drastic reduction of the standard deviation of the call option price is obtained by combining variance reduction techniques with a symmetric pricing approach. This is particularly so for long maturity call options on volatile assets for which the standard deviation can be reduced by a factor of over 20 when combining all three variance reduction techniques. Finally, we show that these results continue to hold when using an out-of-sample pricing approach.

There are other contributions to the literature on variance reduction techniques for the LSM method. For example, antithetic sampling was used by Longstaff and Schwartz (2001). Control variates was used with the LSM method as early as in Tian and Burrage (2002), and an optimal implementation was suggested in Rasmussen (2005). Importance sampling techniques were discussed in the context of the LSM method in Moreni (2003), and the selection of an optimal importance density was analyzed further in Bolia et al. (2004), Juneja and Kalra (2009), and Morales (2006).[1] However, to our knowledge, the present paper is the first to thoroughly study the simultaneous combination of variance reduction techniques for the Monte Carlo valuation of American-style options with the LSM method and to compare the results for pairs of call and symmetric put options whose solutions coincide such that relative efficiencies can be readily compared.

Our findings have important implications for the potentially complicated problem of efficiently pricing American-style call options. For the 40 American call options considered herein, the implementation of standard variance reduction techniques together with the symmetric pricing approach results in a drastic reduction of the variance over the crude LSM estimator. In particular, combining the symmetric pricing approach with combinations of variance reduction techniques with the control variates method largely facilitates the valuation of long term American call options written on volatile assets.

Our suggested approach is therefore of particular relevance for the valuation of real options, which often take the form of costly investment opportunities for projects whose life may span several decades, and underlying assets are difficult to model at long horizons. Chapter 22 of Brealey et al. (2018), for example, presents several examples of such real call options. In all these cases, the adequate valuation of real options is of paramount importance to give full financial flexibility to a firm's operations. Indeed, the value of the early-exercise feature of such investment opportunities can be seen as the "American premium" of the option, which is most prominent when the option is deep out-of-the-money (OTM), has long maturity, and when the underlying asset processes are volatile or difficult to predict reliably over the life of the option. Moreover, it is worth noting that when the symmetry property holds, a bijective relationship between the exercise boundaries of call and symmetric put options (Detemple 2001) informs practitioners not only about option prices, but also about optimal stochastic control, which is key to the management of real options.

The paper is organized as follows: Section 1.2 outlines the American option pricing prob-

[1] For additional references and examples see Chapter 8 of Glasserman (2003).

lem, provides details on the implementation of the LSM method, and presents the put-call symmetry relation. Section 1.3 presents the results of our numerical experiments, compares variance reduction for comparable call and put options and discusses the efficiency gains made possible by the put-call symmetry. Finally, Section 1.4 concludes. Appendix 1.A describes the variance reduction techniques and discusses some numerical issues that may arise when implementing them. Some additional figures are presented in Appendix 1.B.

1.2 Pricing derivatives with early-exercise features

In this section we first state the valuation problem associated with pricing American options. Next, we demonstrate how the price can be approximated using the Least Squares Monte Carlo method of Longstaff and Schwartz (2001). Finally, we review the put-call symmetry property and present some initial numerical results.

1.2.1 The valuation problem

Consider an American option written on an underlying asset $S(t) : t \in [0, T]$ defined on a continuous filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$ equipped with a filtration $\mathbb{F} = \{\mathcal{F}_t\}_{t=0}^{T}$, where $\mathbb{P}$ is a risk-neutral or pricing probability measure. Consistent with the notation of Glasserman (2003), the objective of this problem is to maximize the discounted payoff $U(t) : t \in [0, T]$ with respect to an $\mathbb{F}$-adapted class of stopping times $\mathcal{T} : \mathcal{T} \subseteq [0, T]$. With underlying asset value $S(t)$, necessarily adapted to $\mathbb{F}$ and constant continuously-compounded interest rate r, we pose the optimal stopping time problem as

$$\begin{aligned} U(0) &= \sup_{\tau \in \mathcal{T}} \mathbb{E}\left[U(\tau) \mid \mathcal{F}_0\right] \\ &= \sup_{\tau \in \mathcal{T}} \mathbb{E}\left[e^{-r\tau} h(S(\tau)) \mid \mathcal{F}_0\right] \\ &= \mathbb{E}\left[U(\tau^*) \mid \mathcal{F}_0\right], \end{aligned} \tag{1.1}$$

where $h(\cdot) \geq 0$ is an $\mathbb{F}$-adapted payoff function, and τ^* is the optimal $\mathbb{F}$-adapted stopping time. Note that for notational simplicity, we have expressed the payoff function as depending only on $S(\tau)$. For instance, the value of an American put option with strike K and payoff $h(S(t)) = (K - S(t))^+$ is given by

$$P(0) = \sup_{\tau \in \mathcal{T}} \mathbb{E}\left[e^{-r\tau}(K - S(\tau))^+ \mid \mathcal{F}_0\right]. \tag{1.2}$$

In general, the payoff need only be adapted to the filtration.

For a put option, the problem of maximizing the expected discounted payoff with respect to a stopping time τ can be seen as the dual problem to the minimization (primal) problem given by

$$\tau^* = \inf\{\tau \in \mathcal{T} : S(\tau) \leq b^*(\tau)\}, \tag{1.3}$$

where $b^*(t)$ represents the optimal exercise boundary at time t. That is, whenever the underlying asset price goes below the threshold, immediate exercise is optimal. Otherwise, the option should be held. Similarly, a call option should be exercised when the underlying asset price

is above its optimal exercise boundary. Thus, each exercise strategy corresponds to a stopping time and determines the American option price via Equation (1.1).

Let us consider a discrete-time formulation of the problem. For an option with a maturity of T years, we consider an evenly spaced partition with J time steps of length $\Delta t = T/J$. The discretized price process of the underlying asset $\{S_j : j = 0, \ldots, J\}$ is defined on a complete risk-neutral probability space $(\Omega, \mathcal{G}, \mathbb{P})$ and adapted to the discrete filtration $\mathbb{G} = \{\mathcal{G}_j\}_{j=0}^{J}$. For simplicity, we use time j to refer to time $\tau_j = j\Delta t$. Option exercise is allowed only at the points defining the time discretization, hence the set of admissible exercise opportunities is $\{\tau_j = j\Delta t : j = 0, \ldots, J\}$, where $0 = \tau_0 < \tau_1 < \cdots < \tau_J = T$. Let $V_j(x)$ denote the time-j value of an unexercised option with underlying asset value $S_j = x$. The option price can then be written as $V_0(S_0)$, which solves the dynamic programming recursion

$$\begin{cases} V_J(S_J) = h(S_J) \\ V_j(S_j) = \max\left(h(S_j), \mathbb{E}\left[e^{-r\Delta t} V_{j+1}(S_{j+1}) \mid \mathcal{G}_j\right]\right), j = J-1, \ldots, 0. \end{cases} \tag{1.4}$$

Options with a discrete early exercise feature of this sort are termed Bermudan options. The continuously exercisable American option price is approximated by letting J tend to infinity (Glasserman 2003). In this paper we work with the time-discretized problem, and describe the options as American rather than Bermudan.

1.2.2 Least-Squares Monte Carlo

Most American option valuation methods rely on the dynamic programming representation in Equation (1.4) to obtain price estimates. The problem remains to estimate the time-j continuation value of the option, i.e., the quantity $\mathbb{E}\left[e^{-r\Delta t} V_{j+1}(S_{j+1}) \mid \mathcal{G}_j\right], j = 0, \ldots, (J-1)$. Since the conditional expectation of a square-integrable function relative to a sigma algebra $\mathcal{G}_j$ can be represented as a countable linear combination of $\mathcal{G}_j$-measurable basis functions $\{\psi_l(\cdot)\}_{l=0}^{\infty}$ (see Royden (1988)), the time-j continuation value can be written as

$$\mathbb{E}\left[e^{-r\Delta t} V_{j+1}(S_{j+1}) \mid \mathcal{G}_j\right] = \sum_{l=0}^{\infty} \psi_l(S_j)\gamma_{j,l}, \tag{1.5}$$

with associated real coefficients $\{\gamma_{j,l}\}_{l=0}^{\infty}$.

However, in practise this quantity cannot be computed and a finite number of basis functions is used. Given a set of N simulated paths $\{S_{n,j} : n = 1, \ldots, N; j = 0, \ldots, J\}$, the LSM method resorts to a parametric approximation of the continuation value with a regression model of the form

$$e^{-r\Delta t} V_{j+1}(S_{n,j+1}) = \sum_{l=0}^{L} \psi_l(S_{n,j})\beta_{j,l} + \varepsilon_{n,j+1}. \tag{1.6}$$

The $(L+1) \times 1$ vector of coefficient estimates $\hat{\beta}_j = \{\hat{\beta}_{j,0}, \ldots, \hat{\beta}_{j,L}\}'$ are obtained by regressing disounted cashflows $e^{-r\Delta t} V_{j+1}(S_{n,j+1})$ against the cross-section of basis functions that relate to time-j in-the-money (ITM) asset paths, denoted by the $1 \times (L+1)$ vector $\psi(S_{i,j}) = \{\psi_0(S_{n,j}), \ldots, \psi_L(S_{n,j})\}$, with ψ_0 as a constant. The time-j path-n fitted continuation value then takes the form

$$\hat{C}_{n,j} = \psi(S_{n,j})\hat{\beta}_j, \tag{1.7}$$

and determines an exercise strategy in which the option is exercised if the payoff is positive and greater than the fitted continuation value. The continuation value approximation of order L provides an estimate of the exercise strategy $b^L(t) : t \in \mathcal{T}$ with a corresponding stopping time τ^L determined by the choice of $L < \infty$. To simplify notation, we hereafter refer to $\tau(n)$ as the path-n LSM stopping time estimate.

The LSM estimator of Longstaff and Schwartz (2001) computes discounted cashflows along each path with the regression approach outlined above. To reduce the notation, we write $\hat{V}_j(S_{n,j})$ as $\hat{V}_{n,j}$ to obtain the *approximate* dynamic program

$$\begin{cases} \hat{V}_{n,J} = h(S_{n,J}) \\ \hat{V}_{n,j} = \begin{cases} h(S_{n,j}) & \text{if } (h(S_{n,j}) \geq \hat{C}_{n,j}) \text{ and } (h(S_{n,j}) > 0) \\ e^{-r\Delta t}\hat{V}_{n,j+1} & \text{if } (h(S_{n,j}) < \hat{C}_{n,j}) \text{ or } (h(S_{n,j}) = 0) \end{cases} \quad j = J-1, \ldots, 0. \end{cases} \tag{1.8}$$

In this dynamic programming representation, the terminal option value is again set to the option payoff at maturity, specifying a continuation value for the next iteration of the backwards-in-time recursion. At the penultimate time step, the criterion for the path-n exercise decision is that the payoff $h(S_{n,J-1})$ is positive and greater than the continuation value $\hat{C}_{n,J-1}$. If the option is exercised, the option value is the immediate payoff. Otherwise, the pathwise option cashflow is discounted back one time step. The algorithm then moves backwards in time, computing option values along all paths at each time, and updating the exercise decisions. If the exercise value never exceeds the continuation value along a simulated path, the option is unexercised, its payoff null, and hence the option value for that path is zero. If the payoff is positive and exceeds the continuation value at least once, the $\mathbb{G}$-adapted stopping time determined by the LSM is the first exercise time.

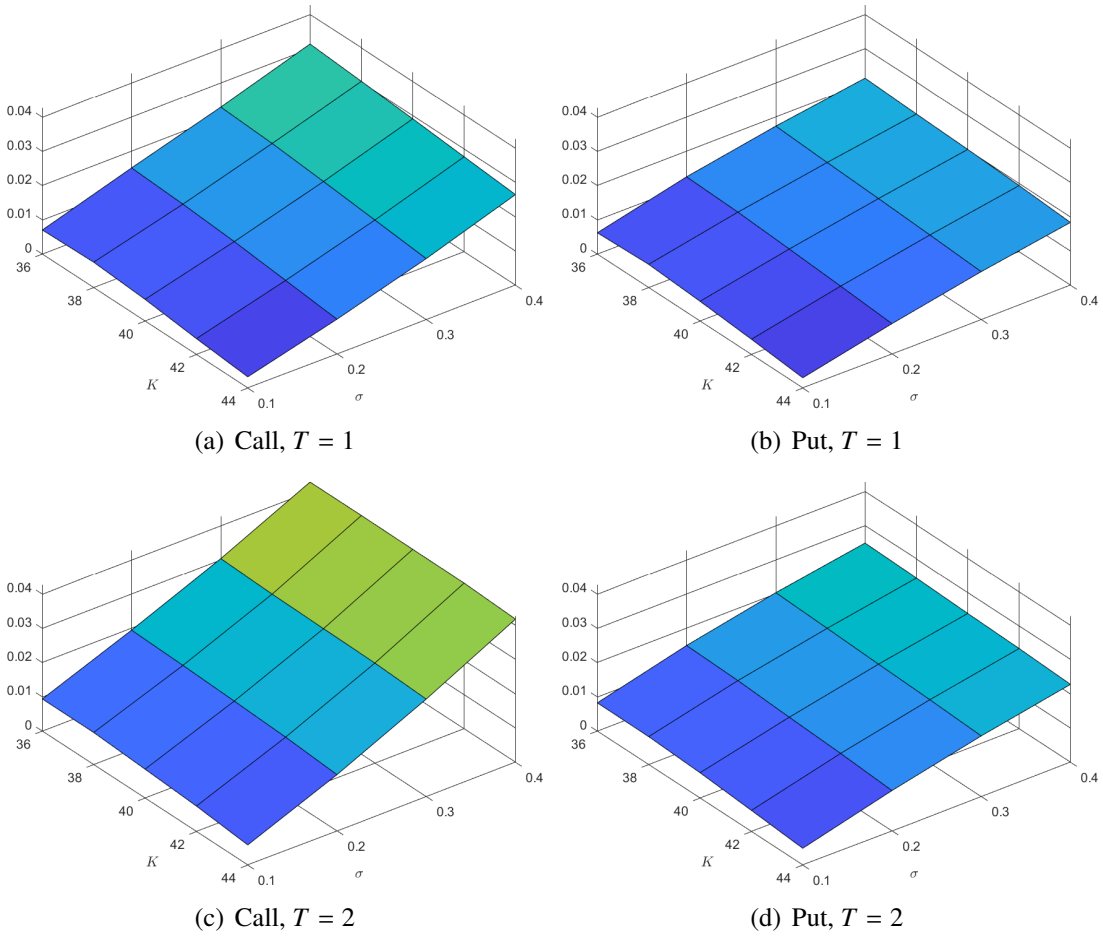
1.2.3 Put-call symmetry

Throughout this paper, results are presented for pairs of American call and put options linked by a put-call symmetry result. Consider an American option with strike price K and maturity T, written on an underlying asset governed by a Geometric Brownian Motion (GBM) and with initial price S_0, volatility of returns σ, and continuous interest rate r and dividend yield q. Denoting put and call option prices by $P(S_0, K, r, q, \sigma, T)$ and $C(S_0, K, r, q, \sigma, T)$, respectively, put-call symmetry states that the following equality holds

$$C(S_0, K, r, q, \sigma, T) = P(K, S_0, q, r, \sigma, T). \tag{1.9}$$

This result was first presented by McDonald and Schroder (1998). By systematically comparing symmetric put and call prices over a wide range of time to maturity, moneyness, and asset volatility, we can readily interpret and compare the efficiencies of variance reduction tools for pairs of problems whose solutions coincide.

Recent literature demonstrates that the symmetry relation is useful because the valuation of American call options poses a number of numerical problems which are otherwise not encountered for put options. In particular, because the payoff of a call option is unbounded, the presence of highly volatile paths will result in a higher frequency of deep moneyness in the cross-section of asset paths. This can impede the approximation of a decision rule enough to



(a) Call, $T = 1$ (b) Put, $T = 1$

(c) Call, $T = 2$ (d) Put, $T = 2$

Figure 1.1: Standard deviation for 1- and 2-year call and symmetric put options.

Results are calculated from $M = 1,000$ replications of the LSM configuration with $N = 100,000$ paths and a cubic approximation ($L = 3$) of the decision rule. The left column illustrates standard deviation for call options, and the right column for symmetric put options. The top and bottom rows present results for one- and two-year options, respectively.

exacerbate variance and create biased price estimates. Symmetric pricing stands as a costless and effective way to value call options and the technique proves particularly effective in cases of long maturity options written on volatile assets.

As an illustration, consider call and (symmetric) put options with 50 exercise opportunities per annum, an initial underlying price $S_0 = 40$, risk-free rate $r = 0.06$, dividend yield $q = 0.06$, and maturities of 1 and 2 years. Figure 1.1 shows the standard deviation calculated from $M = 1,000$ replications of an LSM configuration with $N = 100,000$ paths and a cubic approximation ($L = 3$) of the decision rule. The figure shows that the standard deviation is indeed reduced when using symmetric pricing and this is particularly so for longer maturity options. For shorter maturities and less volatile assets, the symmetric pricing approach has subdued efficiencies in terms of standard deviations, but never produces significantly worse price estimates.

1.3 Empirical results

The Least-Squares Monte Carlo approach for pricing American derivatives has been shown to converge to the true price when the number of exactly simulated GBM paths, N, and the order of the polynomial expansion for the basis functions, L, tend to infinity (Stentoft 2004*a*). In any actual implementation though, a finite number for both is used. In our numerical implementation of the LSM we use $L = 3$ basis functions and $N = 100,000$ sample paths in total to estimate the decision boundary.[2] We price options with $J = 50T$ (i.e., 50 exercise opportunities per annum), $r = q = 0.06$, and $S_0 = 40$. The numerical experiment carried out here consists in pricing 40 call options with varying levels of volatility, $\sigma = \{0.1, 0.2, 0.3, 0.4\}$, different strike prices, $K = \{36, 38, 40, 42, 44\}$, and maturities $T = \{1, 2\}$, as well as their 40 symmetric put counterparts. Results are based on $M = 1,000$ independent replications.

We first compare the variance reduction that can be achieved for call and (symmetric) put options, respectively, and demonstrate that (1) efficiency gains are different for calls and symmetric puts, and (2) control variates should always be used. Second, we note that when using control variates, either alone or in combination with other techniques, efficiencies are always larger for put options than for the corresponding call options and we demonstrate that the joint or total effect of using (combinations of) variance reduction techniques together with symmetric pricing for call options can lead to price estimates with substantially lower variance. We demonstrate that these results are qualitatively identical when using an out-of-sample LSM approach in which the approximated decision rule is applied to a second independent sample of simulated paths. Finally, we provide some intuition for why variance reduction techniques work better than others when it comes to pricing American call and put options.

1.3.1 Variance reduction for call and put options

We start by examining the variance reduction that can be achieved by implementing three stand-alone techniques and combinations thereof for call and (symmetric) put options. Using the symmetric put options allows a direct comparison of the variance reduction for call and put options with similar characteristics and prices. The methods considered are antithetic sampling, denoted with an "A", control variates, denoted with a "C", and importance sampling, denoted with an "I". To compare performance, we consider the standard deviation efficiencies defined as

$$\text{Eff}_{VR} = \sqrt{\frac{\text{Var}\left[\bar{V}^{(N)}\right]}{\text{Var}\left[\bar{V}_{VR}^{(N)}\right]}} - 1, \tag{1.10}$$

where $\bar{V}^{(N)}$ is the crude LSM estimator of an American option using a sample of N paths. Similarly $\bar{V}_{VR}^{(N)}$ is the estimator supplemented with a (combination of) variance reduction technique(s).[3] For example, $\bar{V}_{ACI}^{(N)}$ is the price estimator that combines all three variance reduction

[2] In our implementation we simply set $\psi_\ell(S_{n,j}) = (S_{n,j}/K)^\ell$ such that the continuation value is the fitted value of a polynomial regression of order $L < \infty$. This approach has been shown to be reliable with the LSM, though other orthogonal bases like Laguerre, Legendre, Hermite, or Chebyshev polynomials may be considered as well. For more details about orthogonal bases, refer to Abramowitz and Stegun (1948).

[3] Appendix 1.A discusses the variance reduction techniques and their implementation for American options.

techniques, and Eff_{ACI} denotes its efficiency. The variances are estimated using the $M = 1,000$ replications. Efficiencies greater than zero indicate improvement over crude Monte Carlo.

Table 1.1 reports the results for each of the variance reduction methods with efficiencies for call options in columns 4–10, and symmetric put options in columns 11–17. Option characteristics are shown in the first three columns of Table 1.1. The last row in the table reports the average efficiency for a given variance reduction technique across the call and put options, respectively. The table first demonstrates that efficiency gains from variance reduction techniques are different for call and put options across option characteristics. For example, and we highlight this in Figure 1.2, when using only antithetic sampling the efficiency increases with moneyness and decreases with volatility for call options, whereas the efficiency is highest for at-the-money (ATM) options and increases with the volatility of the underlying asset for put options. Across the methods we observe that efficiencies increase (or never worsen) with time to maturity and asset volatility for symmetric put options whereas efficiencies decrease (or never improve) with time to maturity and asset volatility for call options. For all the individual options as well as on average, some efficiency gain is achieved as all efficiencies are positive.

Secondly, the table shows that the best performing variance reduction method, i.e., the method with the highest efficiency, always involves the use of control variates. As a result the average efficiencies obtained with control variates are much larger than those obtained with antithetic or importance sampling, and in the case of put options this is so by an order of magnitude. The ACI-LSM approach, which combines all three stand-alone variance techniques, is the most efficient approach for most of the options and for 36 out of the 40 put options. ACI-LSM is also the most efficient method for 17 out of the 20 long maturity call options (i.e., options with a maturity of 2 years) and when it is not the best perfoming method it has efficiencies that are very close to the best estimators. In fact, the methods that involve control variates all perform very similarly for call options with very small relative differences and close to optimal variance reductions can be achieved with the computationally simple stand-alone control variate method. For put options however, adding antithetic and importance sampling does further reduce the variance by $(14.1 - 12.4)/12.4 = 13.7\%$ demonstrating that combining all three variance reduction techniques does improve on the efficiency for these options. Figures 1.B.1 and 1.B.2 in Appendix 1.B provide a visual representation of this.

1.3.2 Variance reduction and symmetric pricing

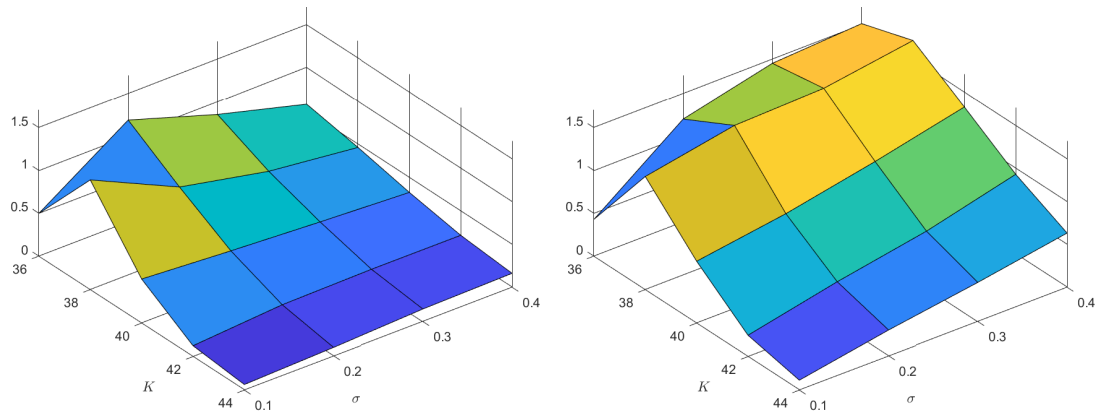
While the efficiency gains from variance reduction techniques are different for call and (symmetric) put options, once a method that involves control variates is considered Table 1.1 shows that a given technique (or combination of techniques) works better for put options than for the corresponding call options. For example, when using control variates the efficiency factor is between 5 and 10 for call options, while for the corresponding puts, the efficiency is larger, and nearly always between 10 and 15. This performance is highlighted in Figure 1.3. The relative improvement is largest when combining all three variance reduction techniques which leads to a $(14.1 - 7.5)/7.5 = 88.0\%$ larger variance reduction for the put than for the call options. Since the variance is always lower for the symmetric put than for the corresponding call option in Figure 1.1 it follows that the variance reduction that can be achieved for call options could be improved upon by considering variance reduction techniques together with symmetrical pricing.

Table 1.1: Efficiency results for various (combinations of) variance reduction techniques.

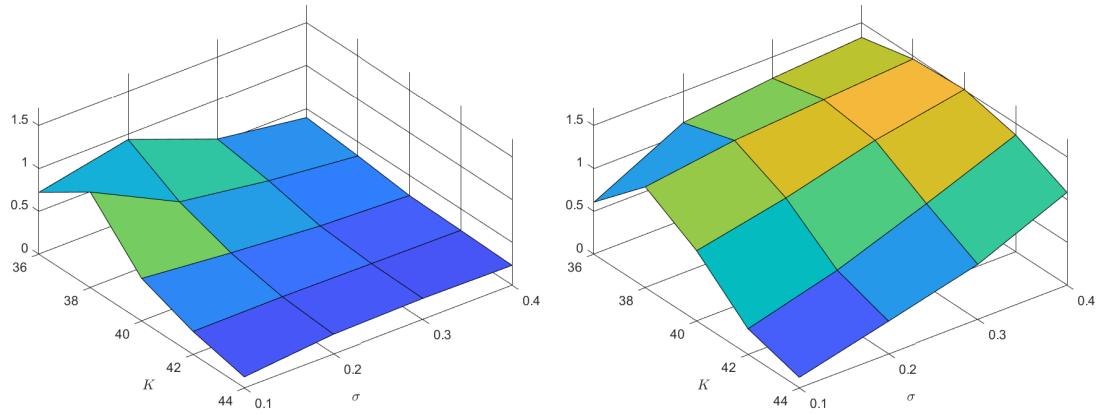
			Call							**Symmetric Put**						
T	K	σ	A	C	I	AC	AI	CI	ACI	A	C	I	AC	AI	CI	ACI
1	36	.1	0.5	8.5	1.1	**9.0**	1.3	7.9	8.9	0.4	10.0	0.9	10.5	1.1	9.5	**10.6**
1	36	.2	1.2	**7.6**	1.4	7.2	1.6	7.4	7.4	1.2	11.2	1.1	11.6	1.3	11.1	**12.8**
1	36	.3	0.9	**6.8**	1.6	6.6	1.9	6.8	6.7	1.5	11.4	1.1	12.8	1.3	11.4	**14.4**
1	36	.4	0.6	6.1	1.8	5.7	2.2	**6.2**	5.8	1.5	12.5	1.1	13.4	1.2	12.6	**15.1**
1	38	.1	1.3	8.7	1.3	8.7	1.6	8.5	**9.1**	1.3	10.5	1.2	11.1	1.4	10.3	**12.1**
1	38	.2	0.8	**8.1**	1.5	7.9	1.8	8.1	7.9	1.5	11.1	1.3	12.3	1.4	11.1	**13.2**
1	38	.3	0.6	7.2	1.7	7.2	2.1	**7.3**	7.3	1.5	11.9	1.2	13.2	1.3	12.0	**14.1**
1	38	.4	0.5	6.0	1.8	5.9	2.2	**6.1**	6.0	1.7	12.2	1.2	13.6	1.3	12.4	**14.8**
1	40	.1	0.5	9.2	1.6	**9.4**	1.9	9.2	9.4	0.7	10.8	1.5	11.8	1.7	10.8	**11.8**
1	40	.2	0.4	8.1	1.7	8.5	2.1	8.1	**8.6**	0.9	11.8	1.4	13.0	1.6	11.9	**13.1**
1	40	.3	0.4	7.2	1.8	7.3	2.3	7.3	**7.3**	1.1	12.2	1.3	13.4	1.5	12.4	**13.7**
1	40	.4	0.3	6.0	1.9	5.6	2.3	**6.0**	5.6	1.3	12.7	1.3	14.0	1.4	12.9	**14.5**
1	42	.1	0.1	8.8	2.0	8.8	2.3	**8.9**	8.8	0.2	11.8	2.1	11.3	2.2	**11.9**	11.3
1	42	.2	0.2	7.8	1.8	7.9	2.3	7.9	**8.0**	0.5	12.5	1.7	12.6	1.8	12.5	**12.6**
1	42	.3	0.2	7.0	1.9	6.8	2.4	**7.1**	6.8	0.7	12.8	1.5	13.2	1.7	12.9	**13.3**
1	42	.4	0.2	5.9	1.9	5.5	2.4	**6.0**	5.5	0.9	13.3	1.4	14.0	1.5	13.6	**14.2**
1	44	.1	0.1	10.3	2.9	10.0	2.8	**10.4**	10.2	0.1	11.4	2.8	12.2	2.6	11.6	**12.3**
1	44	.2	0.1	8.2	2.2	7.8	2.4	**8.3**	7.9	0.3	**13.0**	1.9	12.3	2.1	13.0	12.3
1	44	.3	0.1	6.9	2.0	6.7	2.5	**7.0**	6.7	0.5	13.8	1.6	13.4	1.8	**13.8**	13.5
1	44	.4	0.2	5.9	2.0	5.6	2.5	**6.0**	5.7	0.6	13.7	1.5	13.9	1.6	13.9	**14.1**
2	36	.1	0.7	9.6	1.8	9.6	1.8	9.8	**11.3**	0.6	10.0	1.4	9.8	1.4	10.2	**11.6**
2	36	.2	0.9	8.1	2.0	7.7	1.9	8.5	**8.7**	1.1	11.3	1.4	11.2	1.4	12.5	**16.1**
2	36	.3	0.5	6.4	2.2	6.4	2.2	6.5	**6.8**	1.3	11.9	1.3	11.6	1.3	13.5	**16.5**
2	36	.4	0.4	4.0	2.1	4.0	2.1	4.1	**4.1**	1.3	12.6	1.2	11.8	1.2	14.3	**16.3**
2	38	.1	1.1	9.4	1.9	8.8	1.8	9.9	**10.6**	1.2	11.4	1.5	10.6	1.5	12.4	**14.8**
2	38	.2	0.6	7.9	2.1	8.1	2.1	8.3	**8.7**	1.3	12.9	1.5	11.5	1.5	14.1	**14.6**
2	38	.3	0.4	6.2	2.2	6.4	2.3	6.3	**6.6**	1.4	13.3	1.3	11.8	1.3	14.5	**15.2**
2	38	.4	0.3	3.9	2.1	4.1	2.2	4.0	**4.2**	1.5	13.7	1.2	11.9	1.2	15.0	**15.6**
2	40	.1	0.5	9.3	2.1	9.9	2.2	9.7	**10.4**	0.8	11.9	1.7	12.7	1.8	12.3	**13.8**
2	40	.2	0.4	7.9	2.3	8.7	2.4	8.1	**9.0**	1.0	12.9	1.5	12.8	1.5	13.5	**14.4**
2	40	.3	0.3	5.9	2.3	6.4	2.5	6.1	**6.6**	1.3	13.4	1.3	12.4	1.4	14.3	**14.6**
2	40	.4	0.3	4.0	2.2	4.1	2.3	4.1	**4.2**	1.5	13.9	1.2	12.8	1.2	15.1	**15.9**
2	42	.1	0.3	10.2	2.4	10.3	2.8	10.4	**10.4**	0.3	12.4	1.9	13.2	2.1	13.0	**13.6**
2	42	.2	0.3	7.9	2.4	8.8	2.8	8.1	**9.1**	0.6	13.4	1.6	14.0	1.7	14.3	**15.3**
2	42	.3	0.3	6.0	2.5	6.7	2.8	6.2	**6.8**	1.0	13.4	1.4	14.2	1.4	14.2	**15.9**
2	42	.4	0.3	4.2	2.3	4.2	2.5	4.2	**4.3**	1.4	14.0	1.3	13.4	1.3	14.8	**15.7**
2	44	.1	0.1	9.8	2.6	9.4	3.2	**10.1**	9.6	0.1	11.9	2.2	11.8	2.5	**12.4**	12.1
2	44	.2	0.2	8.3	2.6	8.5	3.0	8.4	**8.6**	0.4	13.5	1.7	14.6	1.8	14.5	**15.3**
2	44	.3	0.2	6.3	2.6	6.1	2.9	**6.4**	6.2	0.6	13.7	1.5	14.3	1.5	14.8	**15.6**
2	44	.4	0.2	4.3	2.4	4.1	2.6	**4.3**	4.2	1.1	14.0	1.3	14.6	1.3	14.9	**16.8**
Average			0.4	7.2	2.0	7.3	2.3	7.3	**7.5**	1.0	12.4	1.5	12.6	1.5	12.9	**14.1**

This table shows the efficiencies for each (combination of) variance reduction technique, with "A" denoting antithetic sampling, "C" denoting control variates and "I" denoting importance sampling, respectively. For each option, the highest efficiencies for call and symmetric put option are indicated in boldface.

To analyse the joint effect of variance reduction and symmetric pricing, we now consider



(a) Call, $T = 1$ (b) Put, $T = 1$



(c) Call, $T = 2$ (d) Put, $T = 2$

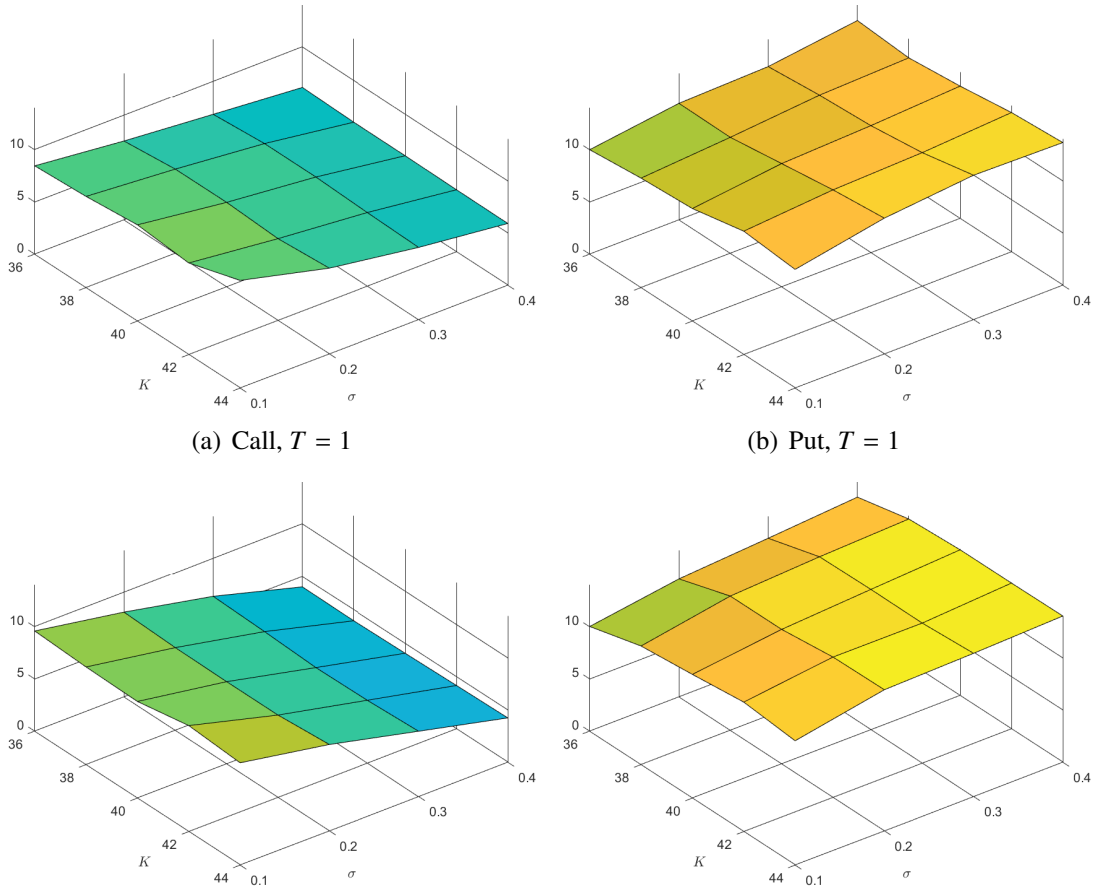
Figure 1.2: Efficiency of pricing with antithetic variates.

Results are calculated from $M = 1,000$ replications of LSM and A-LSM configurations with $N = 100,000$ paths and a cubic approximation ($L = 3$) of the decision rule. The left column illustrates efficiencies for call options, and the right column for symmetric put options. The top and bottom rows present results for one- and two-year options, respectively.

the *total* efficiency for call options of a joint implementation of symmetric pricing and various (combinations of) variance reduction techniques. We define the total efficiency as a slight modification of the efficiency measure in Equation (1.10) given by

$$\text{Tot Eff}_{VR} = \sqrt{\frac{\text{Var}\left[\bar{V}^{(N)}_{call}\right]}{\text{Var}\left[\bar{V}^{(N)}_{put,VR}\right]}} - 1, \tag{1.11}$$

where $\bar{V}^{(N)}_{call}$ is a crude Monte Carlo estimator for a call option using a sample of N paths, and $\bar{V}^{(N)}_{put,VR}$ is the symmetric put estimator using the same sample supplemented with a (combination of) variance reduction technique(s). Again, the variances are calculated across the $M = 1000$ replications. Figure 1.4 depicts the total efficiency for all the possible combinations of variance reduction techniques considered that involve control variates for the 1-year and 2-year call



(a) Call, $T = 1$ (b) Put, $T = 1$

(c) Call, $T = 2$ (d) Put, $T = 2$

Figure 1.3: Efficiency of pricing with control variates.

Results are calculated from $M = 1,000$ replications of LSM and C-LSM configurations with $N = 100,000$ paths and a cubic approximation ($L = 3$) of the decision rule. The left column illustrates efficiencies for call options, and the right column for symmetric put options. The top and bottom rows present results for one- and two-year options, respectively.

options in the top row and bottom row, respectively. This figure can be compared to Figures 1.B.1 and 1.B.2 in Appendix 1.B which shows the efficiency values for call and (symmetric) puts, respectively.

In Table 1.2 we report the results when using an out-of-sample pricing approach instead. Here the optimal stopping time is estimated from the standard LSM method, i.e., without variance reduction, and the same strategy is used for all the out-of-sample pricing with (combinations of) variance reduction techniques. The first thing to notice is that the results are very similar to those reported in Table 1.1 demonstrating that our findings are not particular to using in-sample pricing. Specifically, the effect of variance reduction techniques differ between call and put options, but control variates should always be used.

Second, Table 1.2 also shows that when a control variate is used, either by itself or in combination with other techniques, variance reduction is much more efficient for put options

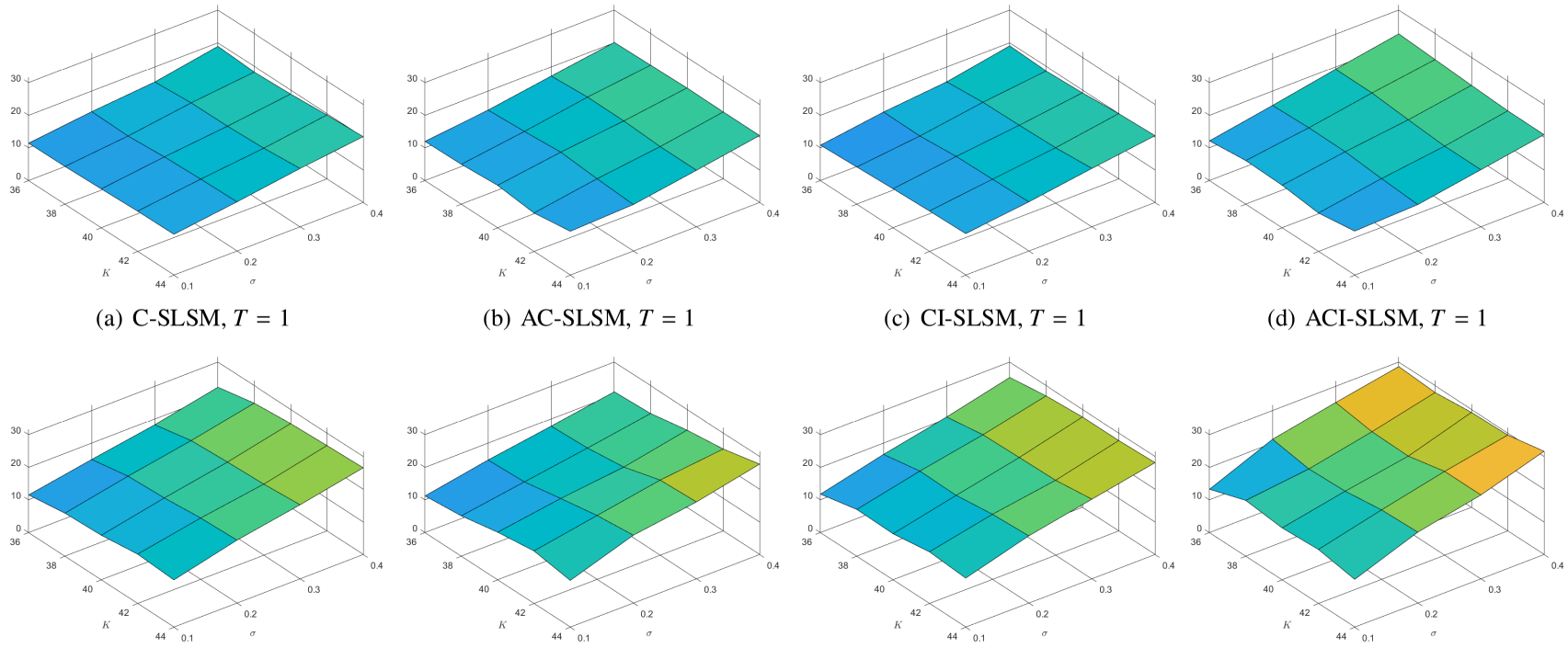
than for call options. One minor difference is that for out-of-sample pricing, ACI-LSM improves somewhat less on the other strategies on average, compared to when using in-sample pricing. However, on average, combining all three variance reduction techniques still offers the highest efficiency, particularly when applied to put options. As a result, combining variance reduction using control variates with symmetric pricing offers huge benefits also when doing out-of-sample call option pricing.[4]

[4]The out-of-sample plots corresponding to Figure 1.4 that shows this are available on request.

Table 1.2: Efficiency results for out-of-sample pricing.

			Call							**Symmetric Put**						
T	K	σ	A	C	I	AC	AI	CI	ACI	A	C	I	AC	AI	CI	ACI
1	36	.1	0.5	7.6	1.1	7.6	1.1	7.6	**7.7**	0.4	9.2	0.9	9.3	1.0	9.2	**9.4**
1	36	.2	1.1	7.1	1.4	7.1	1.4	7.1	**7.1**	1.1	10.3	1.0	10.3	1.1	10.4	**10.6**
1	36	.3	0.9	6.2	1.5	6.2	1.6	6.2	**6.2**	1.4	11.7	1.0	11.7	1.1	11.8	**12.0**
1	36	.4	0.6	5.6	1.7	5.6	1.8	5.6	**5.6**	1.6	12.3	1.0	12.2	1.1	12.4	**12.6**
1	38	.1	1.2	8.1	1.3	8.1	1.4	8.1	**8.1**	1.2	9.8	1.1	9.8	1.2	9.8	**10.0**
1	38	.2	0.9	7.5	1.5	7.5	1.6	7.6	**7.6**	1.4	11.4	1.2	11.5	1.2	11.4	**11.6**
1	38	.3	0.6	6.6	1.6	6.6	1.7	6.7	**6.7**	1.5	12.0	1.1	12.2	1.2	12.1	**12.3**
1	38	.4	0.4	5.6	1.8	5.6	1.8	5.6	**5.6**	1.6	12.7	1.0	12.8	1.1	12.7	**12.9**
1	40	.1	0.5	8.6	1.6	8.6	1.6	8.6	**8.7**	0.8	10.7	1.4	10.8	1.4	10.8	**10.8**
1	40	.2	0.5	7.8	1.7	7.8	1.7	**7.9**	7.8	0.9	11.9	1.3	**11.9**	1.3	11.9	11.9
1	40	.3	0.4	7.0	1.8	7.0	1.8	7.0	**7.0**	1.1	12.7	1.2	**12.8**	1.2	12.7	12.8
1	40	.4	0.3	5.8	1.9	5.8	1.9	5.8	**5.8**	1.3	13.1	1.1	13.2	1.2	13.2	**13.3**
1	42	.1	0.2	9.3	2.2	9.3	2.1	**9.3**	9.3	0.2	11.3	1.8	11.3	1.8	11.4	**11.4**
1	42	.2	0.3	8.1	2.0	8.0	2.0	**8.1**	8.1	0.4	11.8	1.4	11.7	1.5	11.9	**11.9**
1	42	.3	0.3	7.0	2.0	7.0	2.0	**7.0**	7.0	0.7	12.9	1.3	12.8	1.4	12.9	**13.0**
1	42	.4	0.3	5.9	2.0	5.9	2.0	**5.9**	5.9	0.9	13.4	1.2	13.4	1.2	13.4	**13.5**
1	44	.1	0.1	10.4	2.9	10.4	2.8	10.5	**10.5**	0.0	**10.8**	2.4	10.8	2.3	10.8	10.8
1	44	.2	0.2	8.4	2.2	8.4	2.2	8.4	**8.4**	0.2	12.1	1.7	12.1	1.7	12.3	**12.3**
1	44	.3	0.2	7.3	2.2	7.3	2.1	**7.3**	7.3	0.4	12.7	1.4	12.7	1.5	12.8	**12.8**
1	44	.4	0.2	6.1	2.1	6.1	2.1	**6.1**	6.1	0.6	13.3	1.3	13.2	1.3	13.3	**13.4**
2	36	.1	0.7	9.7	1.8	9.8	1.8	9.6	**10.1**	0.7	10.5	1.6	11.0	1.5	10.4	**11.2**
2	36	.2	1.1	8.4	2.3	8.5	2.2	8.6	**8.7**	1.2	12.2	1.5	12.1	1.5	12.4	**13.3**
2	36	.3	0.7	6.5	2.3	6.6	2.3	6.6	**6.6**	1.4	12.8	1.4	12.7	1.4	13.1	**14.0**
2	36	.4	0.5	4.5	2.2	4.5	2.3	4.5	**4.5**	1.4	13.0	1.3	12.8	1.2	13.4	**14.4**
2	38	.1	1.3	10.1	2.2	10.2	2.1	10.5	**10.7**	1.3	12.2	1.7	12.2	1.6	12.4	**13.0**
2	38	.2	0.7	8.2	2.2	8.4	2.2	8.4	**8.5**	1.4	12.4	1.5	12.7	1.5	13.1	**13.7**
2	38	.3	0.6	6.5	2.4	6.6	2.4	6.6	**6.6**	1.5	12.7	1.4	12.9	1.4	13.3	**14.0**
2	38	.4	0.4	4.2	2.1	4.2	2.1	4.2	**4.2**	1.5	13.4	1.3	13.6	1.3	14.0	**14.9**
2	40	.1	0.6	10.1	2.2	10.2	2.1	10.3	**10.5**	0.9	12.4	1.7	12.9	1.7	13.3	**13.6**
2	40	.2	0.5	8.3	2.3	8.3	2.2	8.4	**8.5**	1.1	13.0	1.6	13.6	1.5	13.9	**14.4**
2	40	.3	0.4	6.2	2.3	6.2	2.3	6.2	**6.2**	1.3	13.5	1.5	14.0	1.4	14.5	**15.0**
2	40	.4	0.3	4.1	2.1	4.1	2.1	4.2	**4.2**	1.6	13.6	1.4	13.9	1.3	14.6	**15.3**
2	42	.1	0.3	10.2	2.3	10.2	2.3	10.3	**10.4**	0.4	11.9	1.9	11.9	1.9	12.4	**12.5**
2	42	.2	0.4	8.2	2.4	8.1	2.4	8.2	**8.3**	0.7	13.4	1.7	13.8	1.7	14.3	**14.6**
2	42	.3	0.3	6.2	2.3	6.1	2.3	6.2	**6.3**	1.0	14.1	1.6	14.7	1.5	15.2	**15.7**
2	42	.4	0.3	4.2	2.1	4.2	2.2	4.2	**4.3**	1.4	14.5	1.5	15.3	1.3	15.7	**16.6**
2	44	.1	0.1	9.5	2.5	9.7	2.7	9.7	**9.8**	0.2	12.2	2.1	12.1	2.2	12.6	**12.6**
2	44	.2	0.3	7.9	2.4	7.9	2.4	8.0	**8.0**	0.5	12.9	1.8	13.0	1.8	13.6	**13.7**
2	44	.3	0.3	5.9	2.3	5.9	2.4	5.9	**5.9**	0.8	14.2	1.7	14.7	1.6	15.3	**15.7**
2	44	.4	0.3	4.4	2.2	4.4	2.2	4.4	**4.4**	1.2	14.7	1.5	15.4	1.4	15.9	**16.5**
Average			0.5	7.2	2.0	7.2	2.0	7.3	**7.3**	1.0	12.4	1.4	12.5	1.4	12.8	**13.1**

This table shows the efficiencies for each (combination of) variance reduction technique, with "A" denoting antithetic sampling, "C" denoting control variates and "I" denoting importance sampling, respectively. For each option, the highest efficiencies for call and symmetric put option are indicated in boldface.



(a) C-SLSM, $T = 1$ (b) AC-SLSM, $T = 1$ (c) CI-SLSM, $T = 1$ (d) ACI-SLSM, $T = 1$

(e) C-SLSM, $T = 2$ (f) AC-SLSM, $T = 2$ (g) CI-SLSM, $T = 2$ (h) ACI-SLSM, $T = 2$

Figure 1.4: Total efficiency of symmetric pricing with (combinations of) control variates.

Results are calculated from $M = 1,000$ replications of LSM configurations with $N = 100,000$ paths and a cubic approximation ($L = 3$) of the decision rule. The first column illustrates total efficiencies of C-SLSM prices, the second column for AC-SLSM, the third column for CI-SLSM, and the fourth column for ACI-SLSM. The top and bottom rows present results for one- and two-year options, respectively.

1.3.3 Discussion

In the following, we provide some further intuition for why some variance reduction techniques are more efficient than others. Before doing so, it is important to mention that these comparisons only concern vanilla put and call options, and our conclusions should therefore be interpreted in the specific context of pricing American options with simple payoff functions. There is generally no best-performing variance reduction technique as efficiencies are largely context-specific. A good rule of thumb, as pointed out by Glasserman (2003), is that the more information leveraged about the option properties, the larger the reduction in variance. In practise, however, leveraging this information, and optimizing the variance reduction tools for a specific problem can be so computationally taxing that it would be ill-advised to do so, even if the efficiency gains are large. The success of variance reduction techniques therefore rests on striking a balance between efficiency, and ease of implementation.

On one end of the spectrum, the simplest tool is antithetic sampling, as it requires no knowledge of the option whatsoever, and, as the rule suggests, generally offers lower efficiencies than the two others. Our results show that this intuition applies to our setting in which antithetic sampling has the lowest efficiencies. More generally, we conjecture that the efficiency of antithetic sampling in the context of LSM is subdued in comparison to an European option estimator. Indeed, the efficiency gains permitted by antithetic sampling are larger as the correlation between antithetically simulated option payoffs becomes negative. Supposing that a particular pair of antithetic paths $\{S_{n,j}, \tilde{S}_{n,h},\, j = 0, \ldots, J\}$ is obtained from the normal increments $\{Z_n, -Z_n\}$, respectively, the LSM will estimate stopping times $\{k_n, \tilde{k}_n\}$ corresponding to each path in the antithetic pair, and generate a pair of discounted payoffs $\{e^{-r\tau_{k_n}} h(S_{n,k_n}), e^{-r\tau_{\tilde{k}_n}} h(\tilde{S}_{n,\tilde{k}_n})\}$. The correlation between antithetic payoffs is necessarily negative when stopping times coincide (i.e., $k_n = \tilde{k}_n$), as $h(\cdot)$ is a monotone increasing function in the asset price, and antithetic pairs of asset prices are negatively correlated at a given time. For European options, all exercise times occur at maturity (i.e., $k_n = \tilde{k}_n = J$), and it immediately follows that antithetic sampling necessarily provides an improvement over the standard Monte Carlo approach. This is not generally the case for American options, though, because the antithetically estimated stopping times rarely coincide. Furthermore, if both paths within an antithetic pair are exercised, this will occur at different times and the discounted payoffs will be more positively correlated than pairs of paths having exactly one positive payoff. The resulting effect is a weaker overall correlation, and a subdued efficiency of antithetic sampling.

On the other end of the spectrum, control variates are the most efficient across all option properties, and, as expected, require extensive knowledge about the characteristics of the option. Our experiments consider an ideal case in which the option that serves as a control variate can be valued easily, and replicates the American option payoff exactly at maturity.[5] These ideal control variates are nearly perfectly correlated with the option values, yielding very large efficiencies across all levels of maturity, asset volatility, and moneyness. In essence, the effi-

[5]When the price of a European option replicating the payoff of the American option at maturity is not readily available, one has to resort to using different European options with simpler characteristics. For instance, there is no closed-form formula for the price of an arithmetic Asian option, so the price of a geometric Asian option may be used as a control variate. Another case is if the computation of a European option prices is not feasible for a given underlying diffusion process, where the European price assuming a different process can be used. Although in both cases the European option does not replicate the American option's payoff at maturity, their values are correlated enough to serve as adequate control variates.

ciency of control variates rests on the knowledge of the price of "nearby" options. However, it is worth noting that the implementation of control variates is more challenging and more computationally taxing than the other variance reduction techniques we consider. In particular, the large efficiency gains permitted by control variates largely depend on the availability of payoff-replicating European option prices and on optimization of control parameters.

In the middle of the spectrum, we can place our implementation of importance sampling with a uniform change of drift. This simple parameterization of the importance probability measure uses only limited knowledge of the option characteristics and leads therefore to modest efficiency gains. This variance reduction technique depends on the optimization of the drift parameter, and an ill-chosen drift may very well lead to the estimator having larger variance. If we decided to optimize the importance measure over a larger parametric space (e.g., if we decided to optimize for the drift and the volatility jointly), further gains in efficiency may be obtained, and perhaps an optimal implementation of importance sampling would outperform control variates. However, as discussed in Appendix 1.A.2, the optimization of a drift term is already challenging task and, as detailed, is a delicate process when importance sampling is combined with other variance reduction techniques. That being said, the optimization of importance probability measures with more general parameterizations is still an interesting problem for American option pricing, but one we leave for future research.[6]

1.4 Conclusion

Efficient pricing of American options remains an active area of research. Among the many numerical methods that exist, regression based simulation methods have become popular due to their flexibility. However, as is the case with other Monte Carlo based methods they are numerically costly and reducing the variance of such methods is therefore important. This paper conducts a thorough study of the effect of using various standard variance reduction methods and combinations thereof for a sample of call and (symmetric) put options whose solutions coincide.

Our results first show that the efficiency gains from variance reduction techniques are different for call and put options, and methods that involve control variates work the best. Moreover, any combination of variance reduction methods that include control variates works better for the (symmetric) put option than for the corresponding call option. Finally, since symmetric pricing already offers some variance reduction, our results show that one may obtain reductions in the standard deviation by a factor of more than 20 of a call option when combining control variates with symmetric pricing.

The improvement is largest for long maturity options on high volatility assets. Our suggested approach is therefore of particular relevance to the problem of valuing real options, as they typically take the form of deep OTM, long maturity call options on volatile assets. In vari-

[6]Other interesting sampling strategies for European option pricing could potentially be extended to the LSM algorithm. Chapter 4 of Glasserman (2003) presents several cases of European-style derivatives with rare and path-dependent payoffs where the application of importance sampling and stratification techniques prove extremely efficient. For instance, in the case of a deep OTM knock-in option, the exponential change of measure of the option can be dynamically adjusted such that asset prices are first directed toward the barrier, and once the barrier has been hit, a new change of measure directs asset prices toward the money.

ous sectors, the valuation of real options is of paramount importance for the financial flexibility of the firm, because development projects can be extremely costly, have very long maturities, and the resulting cashflows are difficult to predict. Indeed, in the energy sector, for instance, the proper management of such options adds tremendous value to the firms operations. Because our proposed method is limited to cases where the symmetry property holds, the applicability of the symmetric pricing to real option pricing is left for future research. Extensions to multiple exercise options and alternative stochastic processes where symmetry applies is also left for future work.

Although our general recommendation is to use control variates and symmetric pricing when pricing call options, there is some potential for improving efficiency further for high volatility options by combining this with antithetic and importance sampling. In this case there does not appear to be any offsetting effects of the variance reduction techniques, though implementing the combined techniques is much more challenging. We also leave the question of whether our recommended approach is "optimal" when considering the trade off between computational costs and precision for future research.

Appendix

1.A Variance reduction for American option prices

This appendix first introduces the three variance reduction methods we consider in this paper. Then it provides a discussion of some numerical issues that arises when implementing these techniques. Finally, it provides some evidence on the challenge that arises when choosing the optimal importance sampling measure when combining this variance reduction technique with other techniques, in particular.

1.A.1 Variance reduction techniques

Let $\hat{g}(Z_n)$ be the path-n discounted payoff of an American option resulting from an approximate solution to the dynamic program. The sample size-N crude estimator of the American option is then

$$\bar{V}^{(N)} = N^{-1} \sum_{n=1}^{N} \hat{g}(Z_n), \tag{1.12}$$

where the sample paths are generated with independent random normal increments $Z_n = \{z_{n,1}, \ldots, z_{n,J}\}$, where $z_{n,j} \overset{IID}{\sim} \mathcal{N}(0,1) : n = 1, \ldots, N; j = 1, \ldots, J$. Antithetic sampling, control variates, and importance sampling are widely used to improve on such estimators. We restrict our focus to these three techniques because they require no additional simulated paths. In this way, the estimated exercise strategies are derived from similar samples, allowing for a fair comparison of the efficiency of each technique. Other variance reduction techniques that are not examined here include latin hypercube sampling (Broadie et al. 1997), local policy enhancement (Broadie and Cao 2008), and low discrepancy sequences (Lemieux and La 2005). Initial state dispersion (Rasmussen 2005) may also indirectly function as a variance reduction technique.

The idea behind antithetic sampling is that one can construct a new estimator by combining $\hat{g}(Z_n)$ and $\hat{g}(-Z_n)$, the latter of which is based on the antithetic sample of random variates. For a fair assessment of the efficiency of antithetic sampling, we say that an antithetic estimator with N sample paths is the average of $N/2$ antithetic pairs of regular estimators such that

$$\bar{V}_A^{(N)} = N^{-1} \sum_{n=1}^{\frac{N}{2}} \left(\hat{g}(Z_n) + \hat{g}(-Z_n) \right). \tag{1.13}$$

Antithetic sampling was used with the LSM method in the original paper by Longstaff and Schwartz (2001).

The idea behind control variates is instead to correct (at least partially) for the sample variance by using an available unbiased estimator $\hat{f}_N = N^{-1}\sum_{n=1}^{N}\hat{f}(Z_n)$ of a known quantity f. Specifically, a controlled estimator is obtained from

$$\bar{V}_C^{(N)} = N^{-1}\sum_{n=1}^{N}\left(\hat{g}(Z_n) + \theta\left(\hat{f}(Z_n) - f\right)\right), \tag{1.14}$$

where θ is the *control coefficient*. A natural candidate for a control variate is the price of the corresponding European price or the price of a derivative when using simpler dynamics. Control variates was used with the LSM method as early as in Tian and Burrage (2002).

The idea behind importance sampling is to adjust the paths in such a way that more of them have non-zero payoffs and therefore contain information about the value of the option. In the simplest possible implementation a drift term is added to the normal increments by posing $\tilde{Z}_n \equiv \{z_{n,1} + \sqrt{\Delta t}\lambda, \ldots, z_{n,J} + \sqrt{\Delta t}\lambda\}$. Parameterizing the drifted probability measure as $\mathbb{P}^\lambda$, the importance sampling estimator is then given by

$$\bar{V}_I^{(N)} = N^{-1}\sum_{n=1}^{N}\hat{g}(\tilde{Z}_n)\frac{d\mathbb{P}}{d\mathbb{P}^\lambda}(S_{n,\tau(n)}), \tag{1.15}$$

where the last term is the likelihood ratio. Thus, pathwise discounted cashflows are first computed with paths simulated under $\mathbb{P}^\lambda$ and subsequently multiplied with the corresponding likelihood ratio. Importance sampling was used with the LSM method as early as in Moreni (2003).

1.A.2 Selecting the optimal importance density

The selection of an importance density is a delicate problem where a simple suboptimal approach is often preferred to an optimal one requiring a hefty implementation effort. In our implementation, we simply optimize the importance density for the European case with a uniform drift of the normal increments, keeping the variance unchanged. An important question is then: how much efficiency are we giving up with this suboptimal European solution in comparison to the optimal density for an American option? To answer this question, we study the variance of I-LSM and AI-LSM estimators with respect to the drift of the Brownian increments. As an example, we consider a deep OTM, long maturity, high volatility American call option with $S_0 = 40$, $K = 44$, $\sigma = 0.4$, $r = q = 0.06$, $T = 2$, $J = 50T$. For a range of values of λ, $M = 50$ replications of the LSM are implemented each with $N = 100,000$ paths and a cubic approximation of the continuation value (i.e., $L = 3$). Figures 1.A.1 and 1.A.2 shows the relative variance as a function of the drift and the estimated prices with 95% confidence intervals.

From Figure 1.A.1 the main takeaway is that for stand-alone importance sampling, the Glasserman et al. (1999) (GHS) approximation of the optimal European drift is indeed very close to the American optimum. Moreover, the GHS drift yields gains in efficiency that are similar to what we would otherwise obtain with the true optimum. In such cases, we confirm that it is not worthwhile to undertake computationally taxing gradient descent schemes to reach the true optimum. From Figure 1.A.2, however, the main takeaway is that the convexity of the variance function is not preserved when importance sampling is coupled with antithetic sampling. This is very important for practitioners who intend to compute a gradient-based type of

solution, because a preliminary search of the optimization surface is warranted, thereby adding to the optimizing burden. From Figure 1.A.2 we see that the GHS drift performs slightly worse than the true optimum, but the variance reduction is still substantial. Note that the optimal drift value will also change when importance sampling is combined with control variates.

Finally, one should also note that gradient-based solutions are subject to a number of numerical defects. First, the optimum is sensitive to the number of simulation paths, particularly for small samples sizes. A large number of simulation paths is then required to obtain stable solutions that match those in the literature. Second, numerical convergence issues can arise if the gradient descent does not lead to solutions in a certain neighborhood of the optimum. For instance, if the drift leads all paths to be OTM, the pathwise cashflows will all be zero and the variance is zero. Similarly, if the drift leads all paths to be deep ITM, the likelihood ratio will converge to zero, and the weighted pathwise payoffs will be infinitesimally small. Of course, these issues are easily identifiable as they lead to inordinately biased prices. Still, meticulous care is advised with such delicate algorithms (Moreni 2003), and they might not be useful for valuating a large panel of options in an automated procedure. All things considered, we use the approximate optimal European drift of Glasserman et al. (1999) due to its ease of implementation and realized efficiency.

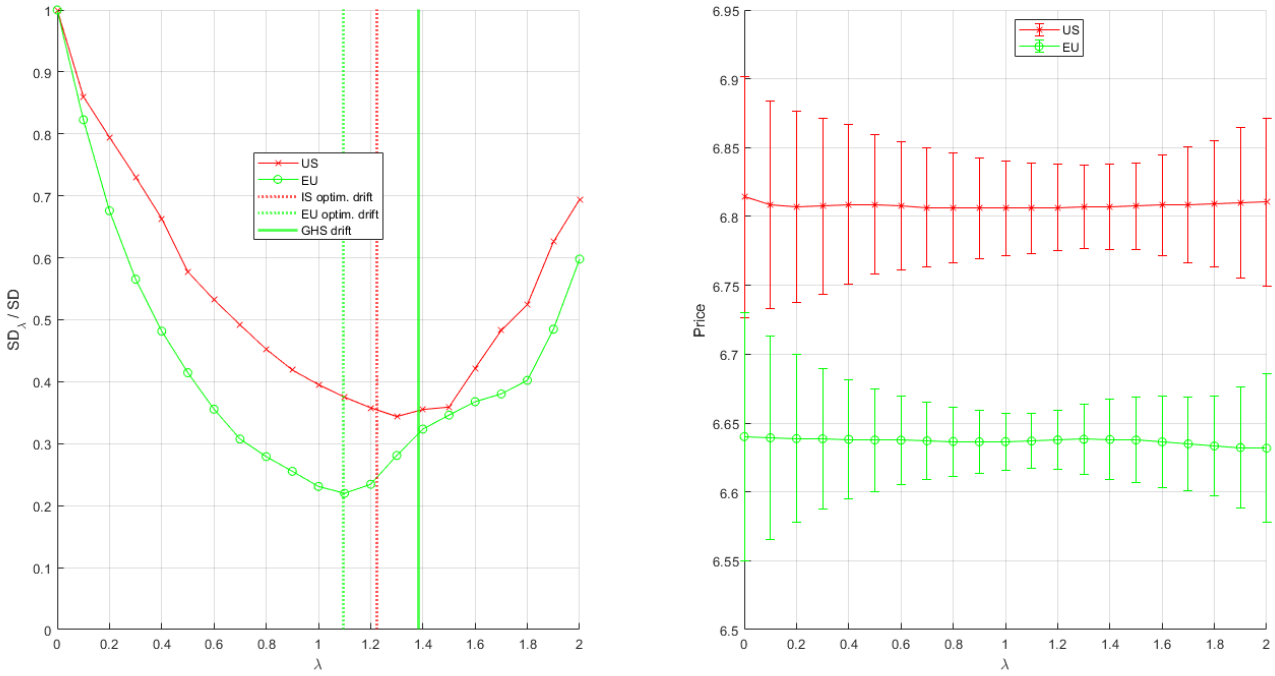


Figure 1.A.1: Variance and price estimates of the I-LSM estimator as a function of λ.

Results for the American call option appear in red, and the results for European call price estimators with importance sampling appear in green. For each value of λ, $M = 50$ replications of the estimators are implemented with $N = 100,000$ paths. LSM and I-LSM price estimates use a cubic approximation of the continuation value (i.e., $L = 3$). The left panel shows the ratio of estimator variances compared to an estimator under the nominal probability measure (i.e. $\lambda = 0$). The optimal values of λ selected by gradient-descent, and that selected by the GHS approach appear as dotted, and full vertical lines, respectively. The right panel shows price estimate with 95% confidence intervals.

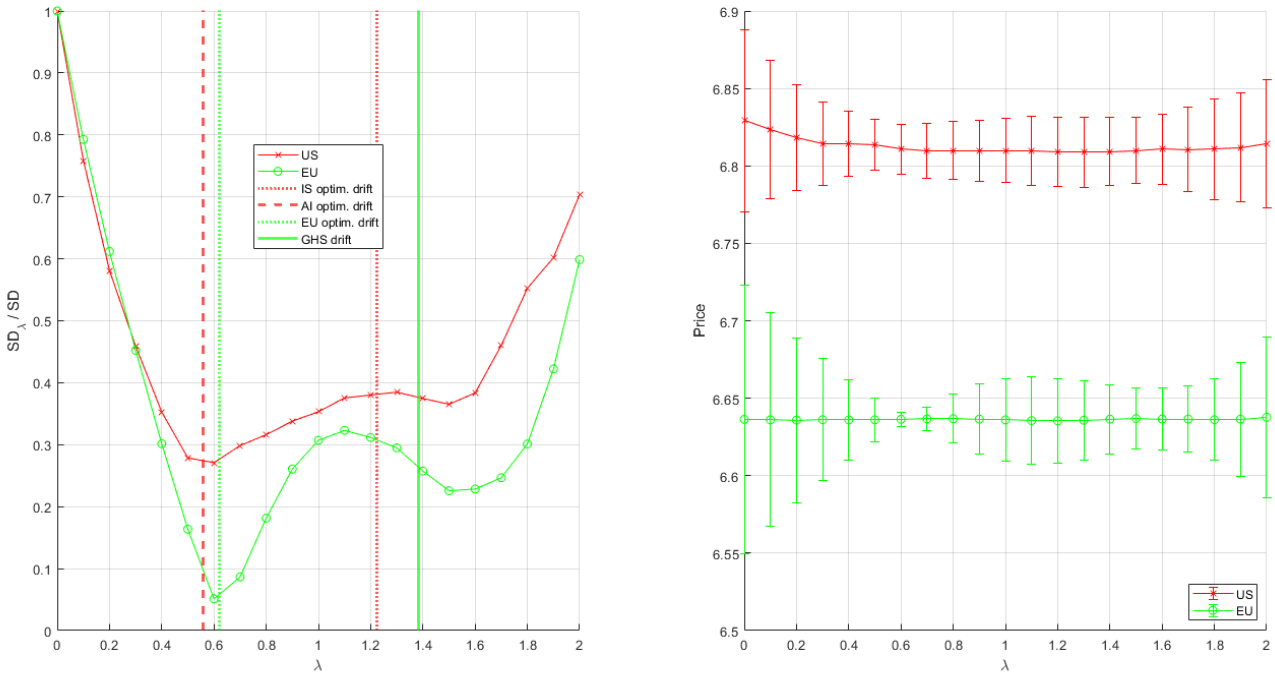
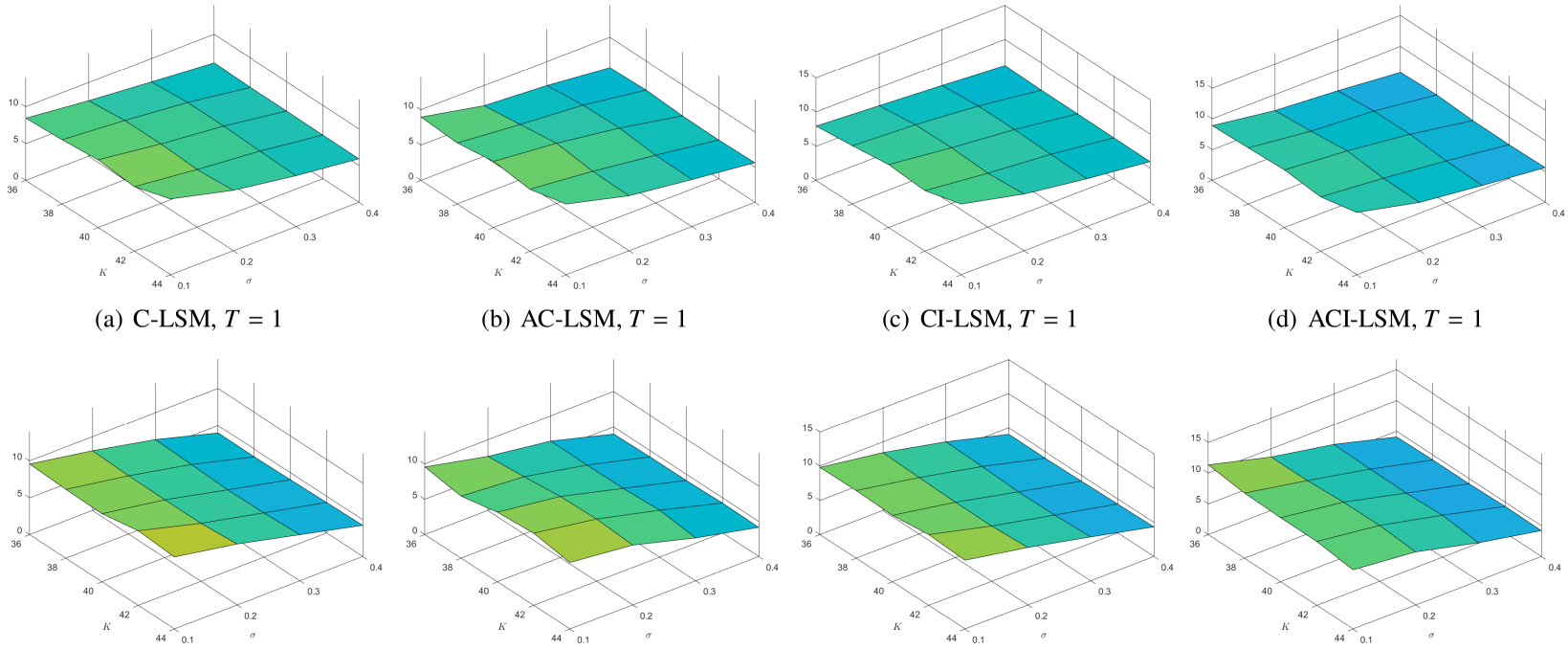


Figure 1.A.2: Variance and price estimates of the AI-LSM estimator as a function of λ.

Results for the American call option appear in red, and the results for European call price estimators that combine antithetic variates and importance sampling appear in green. For each value of λ, $M = 50$ replications of the estimators are implemented with $N = 100,000$ paths. LSM and AI-LSM price estimates use a cubic approximation of the continuation value (i.e., $L = 3$). The left panel shows the ratio of estimator variances compared to an estimator under the nominal probability measure (i.e., $\lambda = 0$). The optimal values of λ selected by gradient-descent for AI- and I-LSM estimators, and that selected by the GHS approach appear as dashed, dotted, and full vertical lines, respectively. The right panel shows price estimate with 95% confidence intervals.

1.B Additional figures

This appendix contains additional figures complementing those presented in the body of the paper. In particular we present variance reduction efficiency surfaces for estimators that use control variates. Namely, we present results for the C-, AC- CI- and ACI-LSM estimators. Results are reported across strike prices $K \in \{36, 38, 40, 42, 44\}$, volatilities $\sigma = \{0.1, 0.2, 0.3, 0.4\}$ and maturities $T = \{1, 2\}$. All results are calculated from $M = 1,000$ replications of LSM configurations with $N = 100,000$ simulated paths, and a cubic approximation ($L = 3$) of the decision rule.

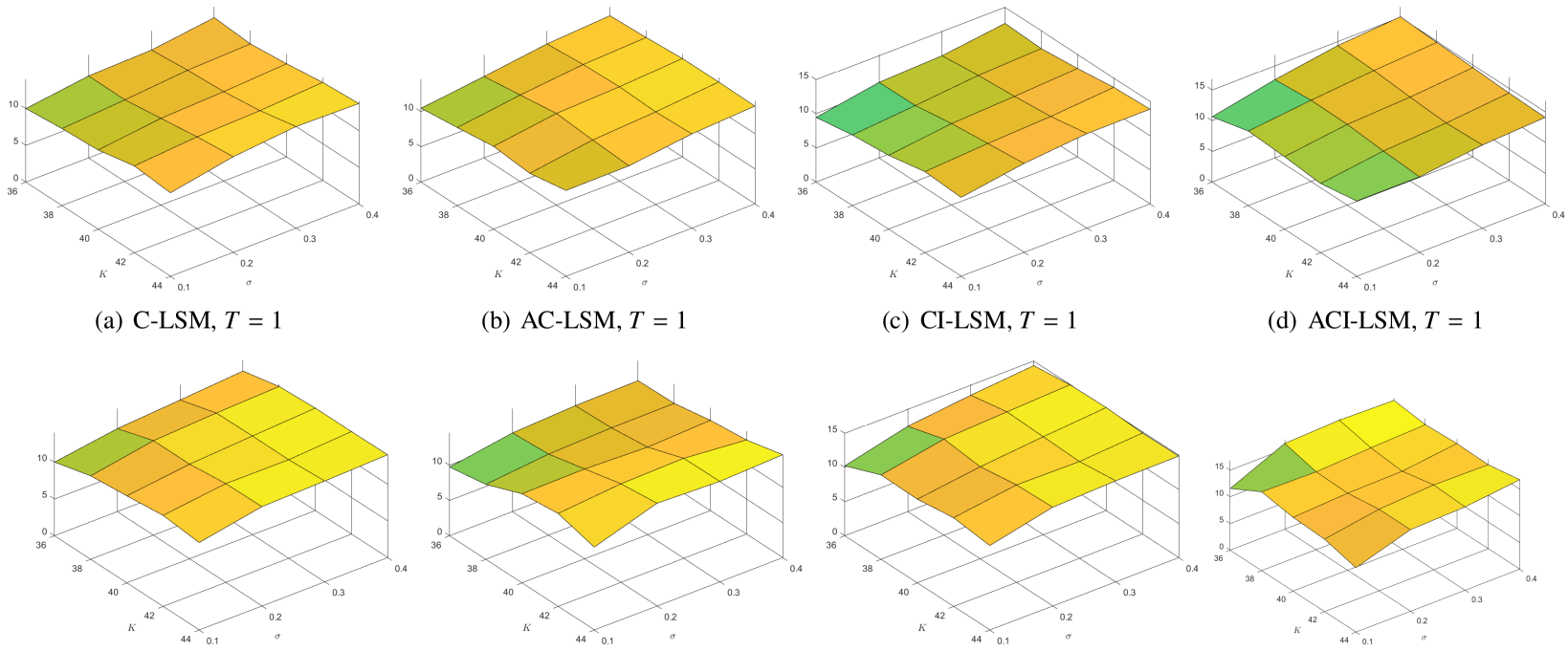


(a) C-LSM, $T = 1$ (b) AC-LSM, $T = 1$ (c) CI-LSM, $T = 1$ (d) ACI-LSM, $T = 1$

(e) C-LSM, $T = 2$ (f) AC-LSM, $T = 2$ (g) CI-LSM, $T = 2$ (h) ACI-SLSM, $T = 2$

Figure 1.B.1: Efficiency for call option pricing with (combinations of) control variates.

Results are calculated from $M = 1,000$ replications of LSM configurations with $N = 100,000$ paths and a cubic approximation ($L = 3$) of the decision rule. The first column illustrates efficiencies of C-LSM call prices, the second column for AC-LSM, the third column for CI-LSM, and the fourth column for ACI-LSM. The top and bottom rows present results for one- and two-year options, respectively.



(a) C-LSM, $T = 1$ (b) AC-LSM, $T = 1$ (c) CI-LSM, $T = 1$ (d) ACI-LSM, $T = 1$

(e) C-LSM, $T = 2$ (f) AC-LSM, $T = 2$ (g) CI-LSM, $T = 2$ (h) ACI-LSM, $T = 2$

Figure 1.B.2: Efficiency for put option pricing with (combinations of) control variates.

Results are calculated from $M = 1,000$ replications of LSM configurations with $N = 100,000$ paths and a cubic approximation ($L = 3$) of the decision rule. The first column illustrates efficiencies of C-LSM put prices, the second column for AC-LSM, the third column for CI-LSM, and the fourth column for ACI-LSM. The top and bottom rows present results for one- and two-year options, respectively.

Chapter 2

American Option Pricing with Importance Sampling and Shifted Regressions

2.1 Introduction

As its name suggests, the least-squares Monte Carlo (LSM) method of Longstaff and Schwartz (2001) employs a least-squares regression approach to estimate an optimal exercise strategy when pricing American style options. Contrary to deterministic solutions like those from multinominal trees or finite differences, regression-based estimators like the LSM are well adapted to multivariate settings and allow for great flexibility in the modeling of a strategy. The estimated exercise strategy is a core element of the LSM procedure, and one of the main obstacles to implementing an accurate LSM estimator is the bias and variance introduced by a sub-optimal exercise strategy. In order to efficiently compute accurate LSM estimators, a simple and effective solution is to employ Monte Carlo variance reduction tools.

One such variance reduction method is importance sampling. The objectives of importance sampling for European option pricing are twofold. First, it reduces the frequency of null payoffs that do not contribute to the estimation of the option premium. The same kind of concept is applied in stratified sampling for instance, where the optimal sampling weights are proportional to the variance of the discounted payoffs within each strata, such that no effort should be spent on sampling unexercised paths. Second, importance sampling reduces the variance of the estimator by generating more payoff events that are otherwise ill-represented in a small sample drawn from the nominal distribution. Importance sampling thereby reduces the number of simulations that are necessary to obtain a balanced sample that includes some of the outliers that play an important role in the computation of an expectation. For the specific problem of pricing American options, a third objective can be formulated. That is, importance sampling should improve the accuracy of an exercise strategy and, hence, reduce estimator bias resulting from incorrect exercise decisions.

This paper demonstrates that the stability of the LSM algorithm is improved when regression coefficients are estimated from the same importance distribution that is used to simulate option prices. The regression of discounted option cashflows on shifted paths improves continuation value predictions where the paucity of data under the nominal measure otherwise impedes the polynomial approximation. Indeed, when too few non-zero cashflows are simu-

lated, the strategy obtained from a standard LSM approach leads to significantly low-biased prices. Our proposed method corrects this bias. The relative importance of the benefits of using our approach is particularly noticeable when a low number of simulation paths is available. Hence, the Monte Carlo valuation of deep out of the money options, knock-in options or barrier options with American-style features may greatly benefit from the approach introduced in this paper.

A crucial aspect of importance sampling lies in the choice of importance measure. For American option pricing with the LSM algorithm, Moreni (2003) considers a uniform change in the drift of the simulated geometric Brownian motion (GBM) paths of the Black and Scholes (1973) model. This method uses the standard continuation values estimated under the nominal measure to determine stopping times for paths simulated under the importance measure using the same normal increments to generate paths under both measures. Another suggestion by Moreni (2004) is simply to implement the LSM algorithm under the nominal measure, and apply a uniform change of drift to the discounted option payoffs at the end of the algorithm after stopping times are estimated using paths under the nominal measure. These American option pricing algorithms successfully achieve the two aforementioned variance reduction objectives. First, they increase the frequency of non-zero payoffs. Second, they generate more rare events by redirecting paths deeper in the money. However, they only achieve modest improvements on the third objective.

It is in this spirit that we introduce the new LSM-s estimator, where the "s" stands for "shifted", in which the shifted paths are used in the regressions. To do so, the LSM-s algorithm employs stepwise likelihood ratios as weights in a weighted least squares approach. The LSM-s comports several advantages over the standard LSM. First, the LSM-s has significantly reduced bias and variance compared to the standard LSM. Second, the LSM-s requires no prior knowledge of regression coefficients since the regressions are directly carried out with the importance paths. Our new approach, therefore, removes the overhead related to estimating an exercise strategy under the nominal measure. Third, the LSM-s enjoys a larger sample of relevant observations in the regression due to a higher incidence of non-zero values in the cross-section of shifted payoffs. Finally, the polynomial approximation of an exercise strategy is enhanced as more paths are simulated in the vicinity of the exercise region. These improvements reduce root mean squared error of the "unshifted" LSM estimator by 22.5% on average. Moreover, the LSM-s is particularly effective at bias reduction when the number of simulated paths is small, since it improves the polynomial approximations in data-poor regions where exercise is optimal.

Several contributions to the literature on variance reduction techniques for Monte Carlo derivative pricing have focused on reducing the variance of discounted payoffs to reduce the variance of the estimator. This approach is perfectly reasonable in the case of European-style options because exercise times are known. Whenever early exercise features are involved however, a significant portion of the error of the estimator is governed by the random errors in the exercise strategy. Consequently applying variance reduction tools to both continuation values and discounted payoffs achieves a significant reduction in both bias and variance. Our results, therefore, contribute to a strand of the literature for the valuation of American options in which bias reduction is achieved by reducing random exercise errors. A similar approach was used in Rasmussen (2005) in the context of variance reduction with control variates, as well as in Létourneau and Stentoft (2019) in the context of bootstrap aggregation of regression coefficients.

See also Kan and Reesor (2012), who derive explicit approximations to the bias of Monte Carlo estimators of American option prices.

The paper is organized as follows: Section 2.2 outlines the American option pricing problem, provides details on the implementation of the LSM method, discusses the use of importance sampling and proposes a new method for using importance sampling with simulation and regression based algorithms. Section 2.3 presents the results of our numerical experiments along with robustness checks demonstrating the benefit of using our proposed method. Next, Section 2.4 explains how our improved approach generalizes to multivariate settings and a wider class of diffusion processes. Finally, Section 2.5 concludes. Appendix 2.A contains further details on the numerical results.

2.2 Pricing derivatives with early-exercise features

In this section, we first state the valuation problem associated with pricing American options. Next, we illustrate how the price can be approximated using the least squares Monte Carlo (LSM) method of Longstaff and Schwartz (2001). Finally, we explain how importance sampling can be used in the context of simulation and regression methods. In this section we also propose our improved method in which regressions are carried out directly under the importance measure instead of under the nominal measure which has been the standard.

2.2.1 The valuation problem

Consider a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ equipped with a continuous filtration $\mathbb{F} = \{\mathcal{F}_t : 0 \leq t \leq T\}$. The valuation of an American option written on the $\mathbb{F}$-adapted underlying asset $S(t) : t \in [0, T]$ is a stochastic optimal control problem with the objective of maximizing discounted payoffs with respect to an $\mathbb{F}$-adapted class of stopping times $\mathcal{T} \subseteq [0, T]$. Assuming a constant continuously-compounded interest rate r and a risk-neutral pricing probability measure $\mathbb{P}$, we denote the price process of the option as $U(t) : t \in [0, T]$ and write

$$U(t) = \sup_{\tau \in \mathcal{T} : t \leq \tau} \mathbb{E}^{\mathbb{P}} \left[e^{-r(\tau - t)} h(S(\tau)) \mid \mathcal{F}_t \right], \tag{2.1}$$

where $h(\cdot) \geq 0$ is an $\mathbb{F}$-adapted payoff function. For instance, the payoff of a put option with strike price K is $h(S(t)) = \max(K - S(t), 0)$.

In a discrete-time formulation of the pricing problem, the complete probability space $(\Omega, \mathcal{G}, \mathbb{P})$ is equipped with a discrete filtration $\mathbb{G} = \{\mathcal{G}_j : j = 0, \ldots, J\}$ to which is adapted the discretized asset price process $\{S_j : j = 0, \ldots, J\}$. Here we assume that the option may be exercised only at evenly spaced time points of length $\Delta t = T/J$ which define the time discretization, such that $\mathcal{T} \subseteq \{\tau_j = j\Delta t : j = 0, \ldots, J\}$. Let $V_j(x) : j = 0, \ldots, J$ denote the time-j value of an unexercised option with underlying asset value $S_j = x$. The option price can then be written as

$$V_j(S_j) = \sup_{k : k \geq j} \mathbb{E}^{\mathbb{P}} \left[e^{-r(k-j)\Delta t} h(S_k) \mid \mathcal{G}_j \right]. \tag{2.2}$$

One can then show that $V_0(S_0)$ solves the dynamic program below

$$\begin{cases} V_J(S_T) = h(S_T) \\ V_j(S_j) = \max\left(h(S_j), \mathbb{E}^{\mathbb{P}} \left[e^{-r\Delta t} V_{j+1}(S_{j+1}) \mid \mathcal{G}_j \right] \right), j = J-1, \ldots, 0. \end{cases} \tag{2.3}$$

Options with such discrete early exercise features are typically called Bermudan options. To value an American option numerically, one is technically required to instead price a Bermudan option with a high number of exercise opportunities, thereby letting $J \to \infty$ (Glasserman 2003). In line with the literature and for the sake of our discussion, we refer to options as American rather than Bermudan in this paper.

2.2.2 Least-Squares Monte Carlo

We can approximate the dynamic program in Equation (2.3) by estimating the conditional expectations $\mathbb{E}^{\mathbb{P}}\left[e^{-r\Delta t}V_{j+1}(S_{j+1}) \mid \mathcal{G}_j\right], j = 0, \ldots, (J-1)$ with ordinary least squares (OLS) regressions. For this task a finite set of $\mathcal{G}_j$-measurable basis functions $\{\psi_\ell(\cdot) : \ell = 0, \ldots, L\}$ is used. Given a set of N simulated paths $\{S_{n,j} : n = 1, \ldots, N; j = 0, \ldots, J\}$, a parametric approximation of the continuation value is obtained with a regression model of the form

$$e^{-r\Delta t}V_{j+1}(S_{n,j+1}) = \sum_{l=0}^{L} \psi_l(S_{n,j})\beta_{j,l} + \varepsilon_{n,j+1}, \tag{2.4}$$

where the path-n, time-j error term $\varepsilon_{n,j}$ satisfies the usual OLS assumptions. Let $\mathcal{N}_j = \{n : h(S_{n,j}) > 0\}$ be the set of time-j in-the-money (ITM) paths and let $N_j = |\mathcal{N}_j|$. In the LSM algorithm, the $(L+1)$ vector of coefficient estimates $\hat{\beta}_j = \{\hat{\beta}_{j,0}, \ldots, \hat{\beta}_{j,L}\}$ are obtained by regressing discounted cashflows $e^{-r\Delta t}\hat{V}_{j+1}(S_{n,j+1})$ against the cross-section of basis functions $\psi(S_{i,j}) = \{\psi_0(S_{n,j}), \ldots, \psi_L(S_{n,j})\}$, with $\psi_0(S_{n,j})$ as a constant, and for $n \in \mathcal{N}_j$. One can then easily show that

$$\hat{\beta}_j = \left(\Psi_j'\Psi_j\right)^{-1}\Psi_j'\hat{\mathbf{V}}_{j+1}, \tag{2.5}$$

where the $N_j \times (L+1)$ matrix Ψ_j denotes the time-j cross-section of basis functions and the $N_j \times 1$ vector $\hat{\mathbf{V}}_{j+1}$ is the sample of discounted option cashflows at time $j+1$. The time-j cross-section of fitted continuation values then takes the form

$$\hat{\mathbf{C}}_j = \Psi_j\hat{\beta}_j = \Psi_j\left(\Psi_j'\Psi_j\right)^{-1}\Psi_j'\hat{\mathbf{V}}_{j+1}, \tag{2.6}$$

which is an unbiased estimator of the conditional expectation of interest $\mathbb{E}^{\mathbb{P}}[\hat{\mathbf{V}}_{j+1}|\Psi_j]$. Posing $\hat{V}_{n,j} \equiv \hat{V}_j(S_{n,j})$ for notational simplicity, the resulting approximate dynamic program computes a sample of discounted cashflows as

$$\begin{cases} \hat{V}_{n,J} = h(S_{n,J}) & \\ \hat{V}_{n,j} = \begin{cases} h(S_{n,j}) & \text{if } h(S_{n,j}) \geq \hat{C}_{n,j} \text{ and } n \in \mathcal{N}_j \\ e^{-r\Delta t}\hat{V}_{n,j+1} & \text{if } \left(h(S_{n,j}) < \hat{C}_{n,j} \text{ and } n \in \mathcal{N}_j\right) \text{ or } n \notin \mathcal{N}_j \end{cases} & j = J-1, \ldots, 0. \end{cases} \tag{2.7}$$

It is easy to see that the estimated continuation values play a critical role in the LSM algorithm because they form a criterion for exercise and, thus, determine stopping time estimates. For each path, $\hat{\tau}_n$ is given by the first time the exercise value is positive and exceeds the estimated continuation value. For unexercised paths, we set $\hat{\tau}_n = J\Delta t = T$.

In this paper, we focus on asset prices governed by a geometric Brownian motion (GBM) as in the Black and Scholes (1973) model with continuous risk-free rate r, dividend yield q,

and volatility of returns σ. The path-n simulation of the asset price is obtained from normal increments $Z_n = \{z_{n,1}, \ldots, z_{n,J}\}$ with

$$S_{n,j} = S_{n,j-1} \exp\left((r - q - \frac{\sigma^2}{2})\Delta t + \sigma\sqrt{\Delta t} z_{n,j}\right), \tag{2.8}$$

where $z_{n,j} \overset{IID}{\sim} \mathcal{N}(0,1) : n = 1, \ldots, N; j = 1, \ldots, J$. The sample-size N LSM estimator of $V_0(S_0)$ is then computed as the average of the pathwise option cashflows

$$\begin{aligned} \hat{g}_N &= N^{-1} \sum_{n=1}^{N} \hat{g}(Z_n) \\ &= N^{-1} \sum_{n=1}^{N} e^{rk_n\Delta t} h(S_{n,k_n}), \end{aligned} \tag{2.9}$$

where $\hat{\tau}_n = k_n \Delta t$ is the path-n estimated stopping time from following the LSM exercise strategy derived from the algorithm in Equation (2.7).

2.2.3 Variance reduction with importance sampling

A potential pitfall with the LSM method arises in the event that only a few of the sample paths meet the optimal exercise threshold. This has the undesirable effect of impeding the reliable approximation of continuation values and of reducing the sample of payoffs that measures the early exercise premium. Importance sampling techniques directly tackle this issue by generating more possible early-exercise instances. In accordance with the literature on importance sampling in a simulation and regression based approach, we select an equivalent importance probability measure among the Gaussian family of densities with identical scale, only allowing for a shift in the location parameter.[1]

To be specific, we consider the situation in which a uniform drift term is added to the normal increments by posing $\tilde{Z}_n \equiv \{z_{n,1} + \sqrt{\Delta t}\lambda, \ldots, z_{n,J} + \sqrt{\Delta t}\lambda\}$ and we denote the "shifted" probability measure as $\tilde{\mathbb{P}}$. The same normal variates are used to simulate the GBM under the importance measure with $\tilde{S}_0 \equiv S_0$ and we have

$$\tilde{S}_{n,j} = \tilde{S}_{n,j-1} \exp\left((r - q - \frac{\sigma^2}{2} + \sigma\lambda)\Delta t + \sigma\sqrt{\Delta t} z_{n,j}\right). \tag{2.10}$$

By virtue of the Girsanov theorem the k-step likelihood ratio for the discretized diffusion process then takes the form

$$\frac{d\mathbb{P}}{d\tilde{\mathbb{P}}}(\tilde{S}_{n,j}, k\Delta t) = \exp\left[\frac{\lambda}{\sigma}\left(r - q - \frac{\sigma^2}{2}\right)k\Delta t - \frac{\lambda}{\sigma}\log\left(\frac{\tilde{S}_{n,j}}{\tilde{S}_{n,j-k}}\right) + \frac{\lambda^2 k\Delta t}{2}\right], \quad k = 0, \ldots, j, \tag{2.11}$$

and the standard importance sampling LSM estimator is given by

$$\begin{aligned} \hat{\tilde{g}}_N &= N^{-1} \sum_{n=1}^{N} \hat{g}(\tilde{Z}_n) \frac{d\mathbb{P}}{d\tilde{\mathbb{P}}}(\tilde{S}_{n,\tilde{k}_n}, \hat{\tilde{\tau}}_n) \\ &= N^{-1} \sum_{n=1}^{N} e^{r\tilde{k}_n\Delta t} h(\tilde{S}_{n,k_n}) \frac{d\mathbb{P}}{d\tilde{\mathbb{P}}}(\tilde{S}_{n,\tilde{k}_n}, \hat{\tilde{\tau}}_n), \end{aligned} \tag{2.12}$$

[1] See Moreni (2003), Lemieux and La (2005), and Morales (2006), for instance.

where $\hat{\tilde{\tau}}_n = \tilde{k}_n \Delta t$ denotes the estimated path-n LSM stopping time.

It is apparent from Equation (2.12) that importance sampling simulates option payoffs under the importance measures. However, in the standard implementation of importance sampling the dynamic program being solved is

$$\begin{cases} \hat{\tilde{V}}_{n,J} = h(\tilde{S}_{n,J}) \\ \hat{\tilde{V}}_{n,j} = \begin{cases} h(\tilde{S}_{n,j}) & \text{if } h(\tilde{S}_{n,j}) \geq \hat{\tilde{C}}_{n,j} \text{ and } n \in \tilde{\mathcal{N}}_j \\ e^{-r\Delta t} \hat{\tilde{V}}_{n,j+1} & \text{if } \left(h(\tilde{S}_{n,j}) < \hat{\tilde{C}}_{n,j} \text{ and } n \in \tilde{\mathcal{N}}_j \right) \text{ or } n \notin \tilde{\mathcal{N}}_j \end{cases} \quad j = J-1, \ldots, 0, \end{cases} \tag{2.13}$$

where $\tilde{\mathcal{N}}_j = \{n : h(\tilde{S}_{n,j}) > 0\}$ is the set of time-j ITM paths simulated from the importance measure $\tilde{\mathbb{P}}$ and $\tilde{N}_j = |\tilde{\mathcal{N}}_j|$. In this problem the continuation value predictions used to obtain the stopping time are given by

$$\hat{\tilde{\mathbf{C}}}_j = \tilde{\Psi}_j \hat{\beta}_j = \tilde{\Psi}_j \left(\Psi_j' \Psi_j \right)^{-1} \Psi_j' \hat{\mathbf{V}}_{j+1}, \tag{2.14}$$

where the regression coefficients are determined from Equation (2.5).

Thus, the standard implementation uses a parametrization obtained by projecting discounted payoffs onto the space spanned by the basis functions evaluated under the nominal measures. This method is, therefore, somewhat cumbersome and essentially requires one to store two sets of simulated paths to calculate both $\Psi_j : j = 0, \ldots, J-1$ and $\tilde{\Psi}_j : j = 0, \ldots, J-1$. Moreover, it stands to reason that if the shifted payoffs improve the determination of the option price, essentially a conditional expectation, it seems likely that using the shifted paths could improve on the estimated continuation values, also conditional expectations, used to estimate the optimal stopping time strategy, thus reducing the bias that results from incorrect exercise decisions.

Our proposed improvement to the standard implementation of importance sampling with the LSM method leverages this and calculates the cross sectional regressions using the shifted paths. We call this estimator the LSM-s and write

$$\begin{cases} \hat{\tilde{V}}^{(s)}_{n,J} = h(\tilde{S}_{n,J}) \frac{d\mathbb{P}}{d\tilde{\mathbb{P}}}(\tilde{S}_{n,J}, \Delta t) \\ \hat{\tilde{V}}^{(s)}_{n,j} = \begin{cases} h(\tilde{S}_{n,j}) \frac{d\mathbb{P}}{d\tilde{\mathbb{P}}}(\tilde{S}_{n,j}, \Delta t) & \text{if } h(\tilde{S}_{n,j}) \geq \hat{\tilde{C}}^{(s)}_{n,j} \text{ and } n \in \tilde{\mathcal{N}}_j \\ e^{-r\Delta t} \hat{\tilde{V}}^{(s)}_{n,j+1} \frac{d\mathbb{P}}{d\tilde{\mathbb{P}}}(\tilde{S}_{n,j}, \Delta t) & \text{if } \left(h(\tilde{S}_{n,j}) < \hat{\tilde{C}}^{(s)}_{n,j} \text{ and } n \in \tilde{\mathcal{N}}_j \right) \text{ or } n \notin \tilde{\mathcal{N}}_j \end{cases} \quad j = J-1, \ldots, 0, \end{cases} \tag{2.15}$$

where the continuation values are now calculated from the cross-section of basis functions evaluated under the importance measure. That is, the $(L+1)$ vector of LSM-s coefficient estimates $\hat{\tilde{\beta}}_j = \{\hat{\tilde{\beta}}_{j,0}, \ldots, \hat{\tilde{\beta}}_{j,L}\}$ are obtained by regressing shifted discounted cashflows $e^{-r\Delta t} \hat{\tilde{V}}_{j+1}(S_{n,j+1}) \frac{d\mathbb{P}}{d\tilde{\mathbb{P}}}(\tilde{S}_{n,j}, \Delta t)$ against the cross-section of basis functions $\psi(\tilde{S}_{i,j}) = \{\psi_0(\tilde{S}_{n,j}), \ldots, \psi_L(\tilde{S}_{n,j})\}$, with $\psi_0(\tilde{S}_{n,j})$ as a constant, and for $n \in \tilde{\mathcal{N}}_j$. The resulting LSM-s coefficients are then

$$\hat{\tilde{\beta}}_j = \left(\tilde{\Psi}_j' \tilde{\Psi}_j \right)^{-1} \tilde{\Psi}_j' \hat{\tilde{\mathbf{V}}}^{(s)}_{j+1}, \tag{2.16}$$

where the $\tilde{N}_j \times (L+1)$ matrix $\tilde{\Psi}_j$ denotes the time-j cross-section of basis functions and the $\tilde{N}_j \times 1$ vector $\hat{\tilde{\mathbf{V}}}^{(s)}_{j+1}$ is the sample of shifted discounted option cashflows at time $j+1$ multiplied

with the corresponding $N \times 1$ vector of stepwise likelihood ratio for $n \in \tilde{\mathcal{N}}_j$. The time-$j$ cross-section of fitted continuation values is then written as

$$\hat{\tilde{\mathbf{C}}}_j^{(s)} = \tilde{\Psi}' \hat{\tilde{\beta}}_j = \tilde{\Psi}' \left(\tilde{\Psi}_j' \tilde{\Psi}_j \right)^{-1} \tilde{\Psi}_j' \hat{\tilde{\mathbf{V}}}_{j+1}^{(s)}. \tag{2.17}$$

Note that, instead of multiplying the likelihood ratios with the payoffs only when pricing the option in Equation (2.12), the stepwise likelihood ratios are now multiplied with the discounted payoffs at each iteration. This leads to the LSM-s estimator

$$\begin{aligned} \hat{\tilde{g}}_N^{(s)} &= N^{-1} \sum_{n=1}^{N} \hat{\tilde{g}}^{(s)}(Z_n) \\ &= N^{-1} \sum_{n=1}^{N} e^{r \tilde{k}_n^{(s)} \Delta t} h(\tilde{S}_{n, \tilde{k}_n^{(s)}}) \frac{d\mathbb{P}}{d\tilde{\mathbb{P}}}(\tilde{S}_{n, \tilde{k}_n^{(s)}}, \hat{\tilde{\tau}}_n^{(s)}), \end{aligned} \tag{2.18}$$

where $\hat{\tilde{\tau}}_n^{(s)} = \tilde{k}_n^{(s)} \Delta t$ denotes the path-n LSM-s stopping time.

To see why LSM-s is better adapted to the task of estimating stopping times, we consider the work by Whitehead et al. (2012) and Kan and Reesor (2012) which derive approximations to the bias of Monte Carlo estimators of American option prices. Whitehead et al. (2012) do this in the context of the easy to analyze stochastic tree, while Kan and Reesor (2012) provide analogous derivations for LSM estimators. In both cases, the approximation to estimator bias depends on the ratio Var/N, where N is the sample size used to construct the continuation value estimator and Var is the estimated variance. The bias goes to zero as Var/N goes (monotonically) to zero, which happens as the sample size increases or as the estimated variance decreases. The typical importance sampling technique for American options (e.g., Moreni (2003)) has the effect of reducing the estimated variance (Var) compared to the regular LSM estimator. Our proposed LSM-s estimator enjoys the same reduced estimated variance as the LSM estimator with importance sampling, but also increases the sample size used to construct the continuation value estimator. Hence our proposed estimator leads to a further reduction in bias compared to the LSM estimator with importance sampling.

2.3 Numerical results

In this section we report results from a numerical application of the shifted regression method proposed above and compare the results to what would be obtained with the standard simulation and regression based method. We first consider two techniques that can be used to approximate the variance-minimizing change of measure in the context of importance sampling. The first such technique directly optimizes the drift parameter with a numerical gradient-based method. The second technique uses an approximation of the optimal drift for a European option to price an analogous American option. Although suboptimal, the latter technique is common in the literature.[2] Second, we present numerical results for a sample of options with different moneyness and maturity highlighting the advantages of applying the optimal importance sampling tools in the regression component of the LSM. To do so, we examine the effect of

[2] See, e.g., Moreni (2003) and Lemieux and La (2005).

importance sampling on the standard deviation, bias, and RMSE efficiency of the LSM and the LSM-s methods. Note that although the drift is selected to minimize variance, it has a remarkable effect on the estimator bias, particularly for LSM-s. Finally, we provide evidence on the robustness of these improvements when a sub-optimal change of drift is considered.

2.3.1 Determination of the optimal drift

The optimal parameterization of an importance measure requires prior knowledge about the option price. Consequently practitioners are faced with the difficult problem of designing an automated procedure to select a distribution. That is to say, the chosen importance measure should consistently reduce the variance of a nominal estimator. One common approach is to find an easy solution to an approximating problem. With this in mind, Moreni (2003) simply proposes to approximate the optimal change of measure for a European option with the saddle point approximation of Glasserman et al. (1999), here called the GHS drift, and apply it to the corresponding American option. For put options, the GHS change in drift λ can be computed very quickly as

$$\lambda = \underset{x:x<\xi}{\operatorname{argmax}} \log\left[K - S_0 \exp\left((r - q - \frac{\sigma^2}{2})T + \sigma\sqrt{T}x\right)\right] - \frac{x^2}{2}, \tag{2.19}$$

where $\xi = \left(\log(K/S_0) - (r - q - \sigma^2/2)T\right)/(\sigma\sqrt{T})$. This approach successfully achieves the task of reducing the variance of the nominal estimator, and has the advantage of employing a very cost-effective approximation of the optimal drift.

However, one can easily find instances in which the GHS drift is significantly sub-optimal and where the variance can be further reduced with a more involved optimization scheme. One such example is shown in Figure 2.3.1, which plots the standard deviation efficiency (left hand plot) and the bias with respect to a Cox et al. (1979) binomial tree with a large number of steps (right hand plot), obtained for different values of the change of drift λ for a put option with $r = q = 0.06$, $\sigma = 0.2$, $S_0 = K = 40$, $T = 1$, and $N = 10^4$. The three lines correspond to using an optimal exercise strategy, labeled FDM, and early exercise strategies estimated with the LSM and LSM-s methods, respectively. The FDM simply replaces the exercise criterion in the LSM algorithm in Equation (2.13) by a precisely estimated boundary criterion obtained with a finite difference method with a fine grid. The figure demonstrates that the value of λ obtained with the GHS method, although quite close to the optimal level, is sub-optimal in terms of the variance reduction that could be achieved. Moreover, using the GHS drift could lead to price estimates that are more biased than with the actual optimal variance minimizing drift.

Motivated by the findings in Figure 2.3.1, in Section 2.3.2 we first consider the relative performance when an optimal drift is used. To implement this we use a gradient-based solution, as proposed in Morales (2006), by numerically computing the gradient of the variance as a function of the drift term λ. Since Figure 2.3.1 shows that the variance of the estimators is convex with respect to the drift parameter, numerical convergence is achieved fairly quickly.[3]

[3]Note that Boire et al. (2021*c*) show that this approach is not appropriate when importance sampling is combined with other variance reduction techniques. In this situation the variance is no longer convex in the change of drift and alternative methods like, e.g., a grid search is required.

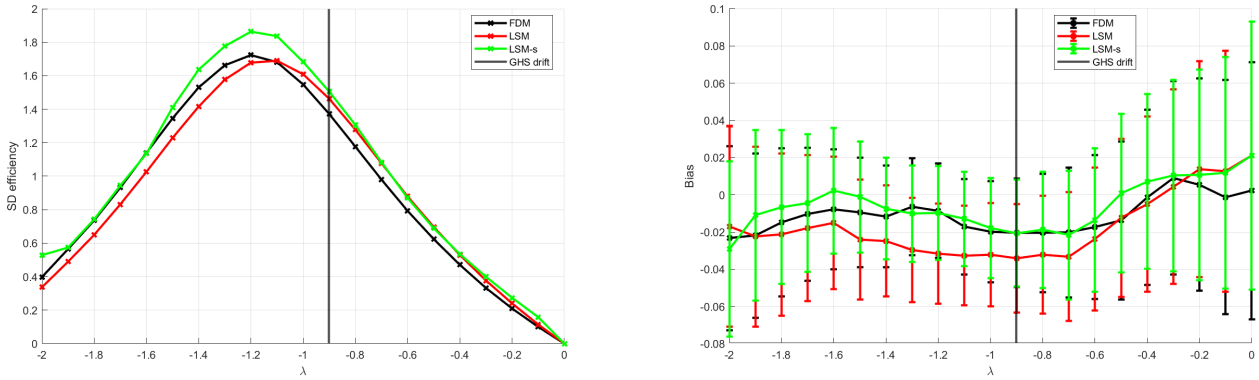


Figure 2.3.1: Standard deviation efficiency and bias as a function of change in drift λ.
Results are for a put option with $r = q = 0.06$, $\sigma = 0.2$, $S_0 = K = 40$, $T = 1$, which is priced using $N = 10^4$ paths. The three lines correspond to optimal, labeled FDM, and LSM and LSM-s estimates obtained with a polynomial of order 3 in the cross-sectional regressions, respectively. The error bars around the bias illustrate 95% confidence intervals.

For a fair comparison of the performance of the LSM and LSM-s under optimal importance sampling, we use the FDM method with $N = 100,000$ paths in the simulation to optimize the drift. In Section 2.3.3, we consider the results obtained when using the easier to calculate but sub-optimal GHS change of measure as a robustness check.

2.3.2 The benefits of using shifted regressions

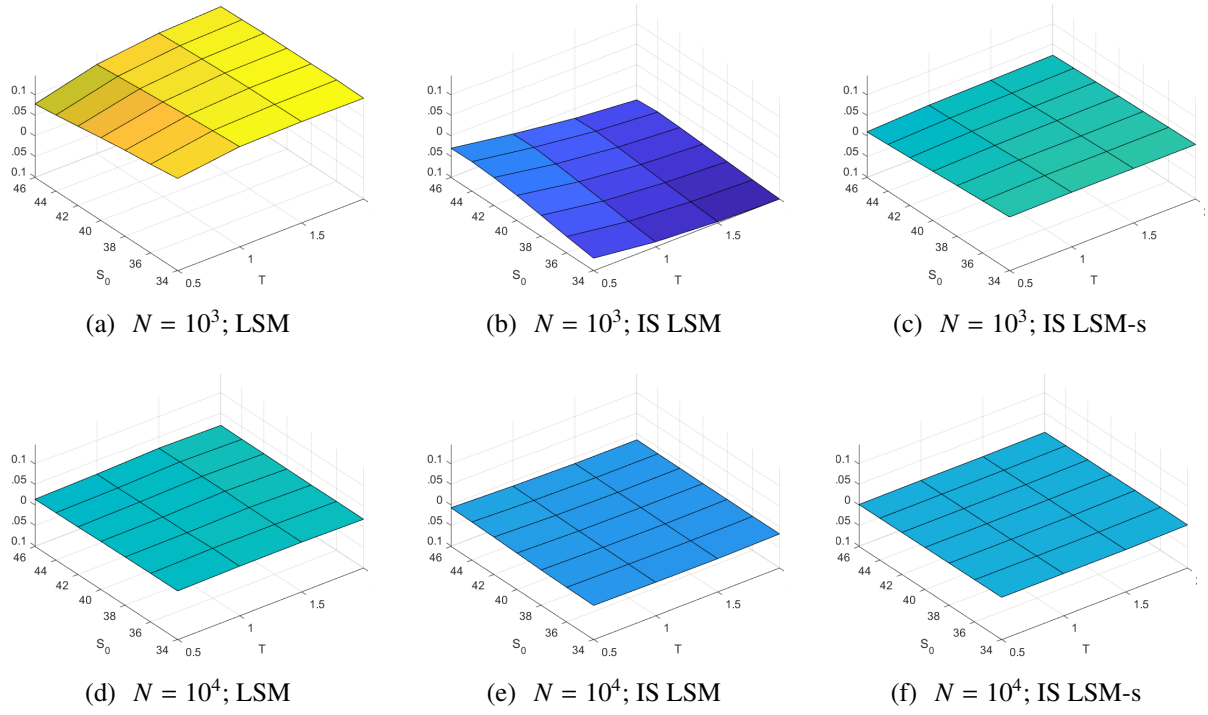
In our numerical results, we price a total of 28 put options with 50 exercise opportunities per year (i.e., $J = 50T$), maturities (in years) $T \in \{0.5, 1, 1.5, 2\}$ and strike prices $K \in \{34, 36, 38, 40, 42, 44, 46\}$, where the underlying asset price process is governed by a GBM with initial asset price $S_0 = 40$, volatility of log-returns $\sigma = 0.4$, and risk free rate r and dividend yield q set at $r = q = 0.06$. Both the LSM and the LSM-s algorithms employ a cubic approximation of the continuation value with the four basis functions $\{\psi_\ell(s) = s^\ell : \ell = 0, 1, 2, 3\}$. We compute Monte Carlo estimators with $N \in \{10^3, 10^4, 10^5\}$ paths and $R = 10^7/N$ replications to illustrate how the relative importance of this method increases for estimators with smaller sample sizes.

Bias

We first note from Figure 2.3.2 that the LSM-s estimators always have the smallest bias.[4] Indeed, a significant bias is observed when importance sampling is used with the LSM, IS LSM, but this bias is eliminated with the importance sample LSM-s, IS LSM-s, even when as little as $N = 1,000$ paths are used in the simulation. The LSM-s also has smaller bias than the LSM with and without importance sampling for all 28 options when using $N = 10,000$ paths,

[4]Since there is no closed-form solution to the price of an American option, we use a very precise approximation of the price from a Cox et al. (1979) binomial tree with a large number of time steps as the benchmark

although the magnitude of the bias tends to be smaller. The detailed results in Table 2.A.1 in Appendix 2.A shows that this pattern persists when a large sample of $N = 100,000$ paths is used in the simulation although in this case the size of the bias is essentially negligible. From these results we conclude that the LSM-s offers important reductions in the estimator bias compared to the LSM counterparts with and without importance sampling, and this is particularly so when the number of simulation paths is low.



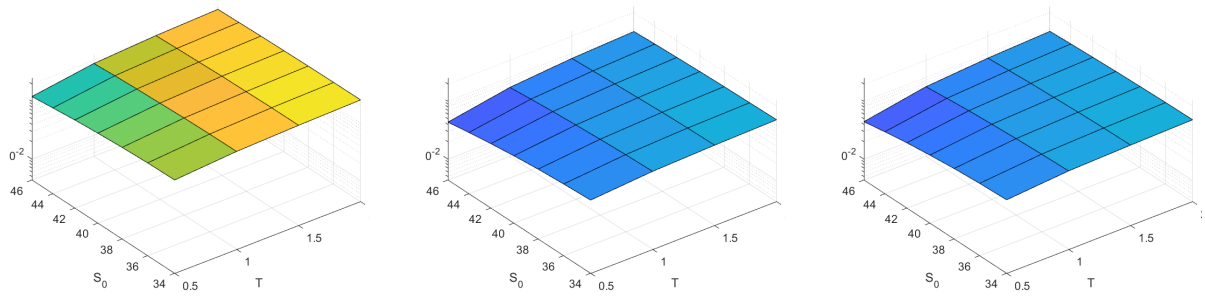
(a) $N = 10^3$; LSM (b) $N = 10^3$; IS LSM (c) $N = 10^3$; IS LSM-s

(d) $N = 10^4$; LSM (e) $N = 10^4$; IS LSM (f) $N = 10^4$; IS LSM-s

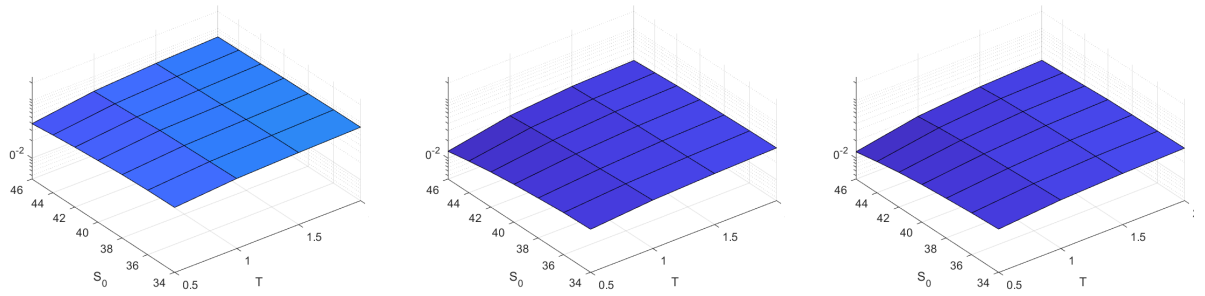
Figure 2.3.2: Bias without and with optimal drift.

Bias results are calculated with respect to a Cox et al. (1979) binomial tree solution for the LSM and the LSM-s estimators. The left column illustrates results for the LSM without importance sampling, the center column LSM with importance sampling, and the right column LSM-s with importance sampling. The top row presents results for estimators using 1000 paths and the bottom row for 10,000 paths.

The results in Figure 2.3.2 demonstrate that fitted regression values tend to be less accurate when regression coefficients estimated under the nominal measure are used to estimate continuation values when conditioning on values simulated from the importance measure. Indeed, continuation values can be difficult to approximate, especially if they take the form of a polynomial extrapolation for deep ITM paths. By regressing on paths from the importance measure the fitted values are more robust for deep ITM paths. Having a larger sample of deep ITM paths also contributes to preserving the convexity of continuation values with respect to the asset price. This effect reduces the frequency and importance of exercise errors and thus corrects the bias.



(a) $N = 10^3$; LSM (b) $N = 10^3$; IS LSM (c) $N = 10^3$; IS LSM-s



(d) $N = 10^4$; LSM (e) $N = 10^4$; IS LSM (f) $N = 10^4$; IS LSM-s

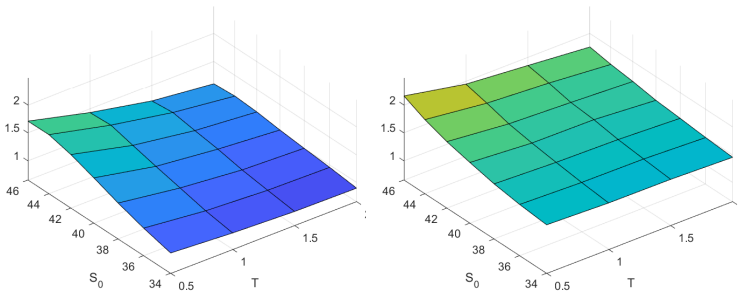
Figure 2.3.3: Standard deviation without and with optimal drift.

Standard deviation results are calculated from a sample of $R = 10^7/N$ replications of the LSM and the LSM-s estimators. The left column illustrates results for the LSM without importance sampling, the center column LSM with importance sampling, and the right column LSM-s with importance sampling. The top row presents results for estimators using 1000 paths and the bottom row for 10,000 paths.

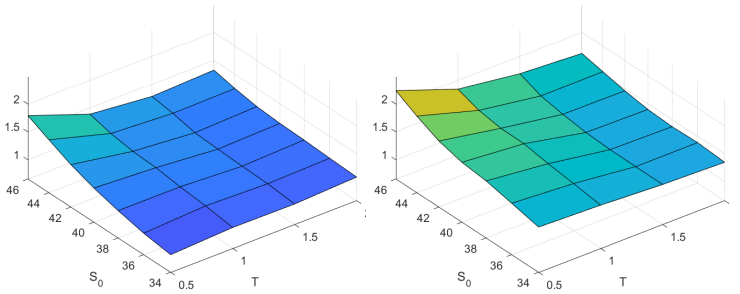
Standard deviation

We next note from Figure 2.3.3 that the importance sampling variance reductions of the LSM and LSM-s estimators are remarkably similar. This is also confirmed by the detailed results shown in Table 2.A.3 in Appendix 2.A. The figure also shows that compared to the variance of the LSM estimator without importance sampling the standard deviations are significantly lower, often around half of what is obtained without importance sampling, and importance sampling is, thus, clearly effective as a variance reduction technique.

The results in Figure 2.3.3 demonstrates that while the estimated early exercise strategy is worse with the IS LSM than when using the IS LSM-s it is "equally" sub-optimal across all repetitions leading to an estimator bias of similar size. Once an early exercise strategy has been estimated, valuing the option amounts to averaging discounted payoffs and the fraction of such payoffs that are non-zero depends much less on the estimated exercise strategy and much more on the change of measure.



(a) $N = 10^3$; IS LSM (b) $N = 10^3$; IS LSM-s



(c) $N = 10^4$; IS LSM (d) $N = 10^4$; IS LSM-s

Figure 2.3.4: RMSE efficiency with optimal drift.

RMSE efficiency results are calculated from a sample of $R = 10^7/N$ replications of the LSM and the LSM-s estimators. The left column illustrates results for the LSM without importance sampling, the center column LSM with importance sampling, and the right column LSM-s with importance sampling. The top row presents results for estimators using 1000 paths and the bottom row for 10,000 paths.

RMSE efficiency

The results above demonstrate the efficiency of importance sampling as a variance reduction technique. They also demonstrate that the standard implementation may introduce a significant bias in the estimator which fortunately can be eliminated by using our proposed shifted LSM regression methodology. Combining these two findings we expect the root mean squared error, RMSE, of the estimated prices to be smaller when using LSM-s than when using LSM with importance sampling. We also expect that both methods have RMSEs that are much smaller than what is obtained with the LSM method implemented without any importance sampling. To illustrate this in a more concise way, we compute the RMSE efficiency of a given estimator $\hat{f}_N = N^{-1} \sum_{n=1}^{N} \hat{f}(Z_n)$ as

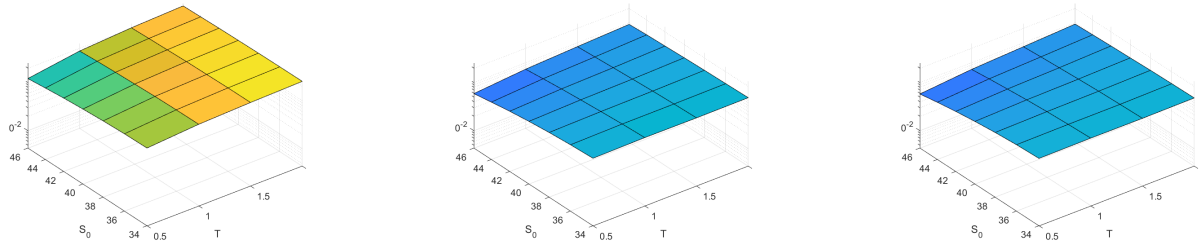
$$\text{Eff}(\hat{f}) = \frac{\text{RMSE}\left[\hat{g}_N)\right]}{\text{RMSE}\left[\hat{f}_N\right]} - 1. \tag{2.20}$$

In other words, the efficiencies of importance sampling for the LSM and the LSM-s are measured relative to the standard LSM estimator with the same number of paths.

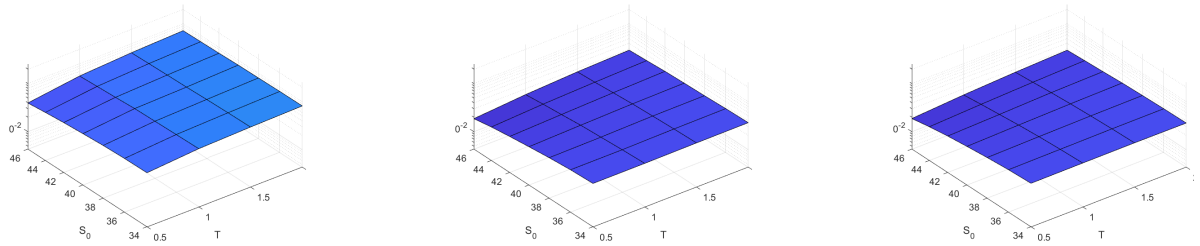
Figure 2.3.4 plots the efficiencies for the two methods and demonstrates that the LSM-s method indeed has the highest RMSE efficiency across the board for all the options, and this

holds irrespective of the number of paths used. When using only $N = 1,000$ paths the average efficiency of the LSM-s is an impressive 1.68. The RMSE of the IS LSM-s is on average 22.5% smaller than that of the IS LSM, at least 8.8% better and could improve on the IS LSM by as much as 31.4%. The detailed results in Table 2.A.5 in Appendix 2.A shows that the LSM-s also exhibits higher efficiencies when a large sample of $N = 100,000$ paths is used in the simulation. However, the relative difference between the efficiencies of importance sampling for LSM and LSM-s methods decreases with the sample size because the bias improvement is negligible.

The results in the above figures therefore provide strong support for our proposed method. In particular, they demonstrate that our proposed method should be particularly useful when one is restricted to using a low number of simulated paths for option pricing. Moreover, because of the reduced bias this method lends itself to simple parallel implementation where many independent jobs (with small sample size) are serially farmed out to processors in parallel. The reduction in bias allows for more accurate estimators when computing price estimators in this fashion as compared to the LSM and IS LSM estimators, respectively.



(a) $N = 10^3$; LSM (b) $N = 10^3$; IS LSM (c) $N = 10^3$; IS LSM-s



(d) $N = 10^4$; LSM (e) $N = 10^4$; IS LSM (f) $N = 10^4$; IS LSM-s

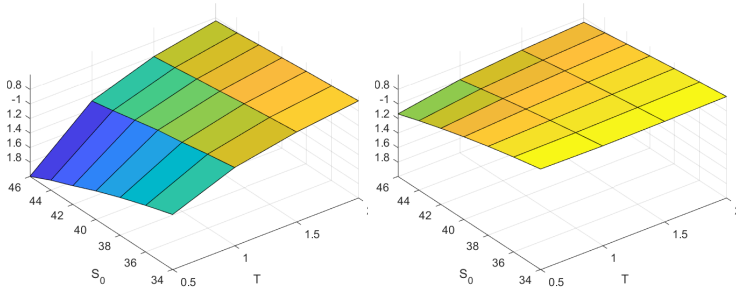
Figure 2.3.5: Standard deviation without and with GHS drift.

Standard deviation results are calculated from a sample of $R = 10^7/N$ replications of the LSM and the LSM-s estimators. The left column illustrates results for the LSM without importance sampling, the center column LSM with importance sampling, and the right column LSM-s with importance sampling. The top row presents results for estimators using 1000 paths and the bottom row for 10,000 paths.

2.3.3 Robustness

The results in Section 2.3.2 are based on the optimal (but potentially infeasible) drift which is difficult (or even impossible) to calculate as it requires knowledge of the optimal early exercise strategy. In this section we, therefore, consider the robustness of our results when the easier

to calculate but sub-optimal GHS change of drift is used. Figure 2.3.6 demonstrates that the changes of drift indeed differ and that it tends to be much smaller for the sub-optimal GHS method and this particularly so for short maturity options.



(a) Optimal drift (b) GHS drift

Figure 2.3.6: Optimal and GHS drift.

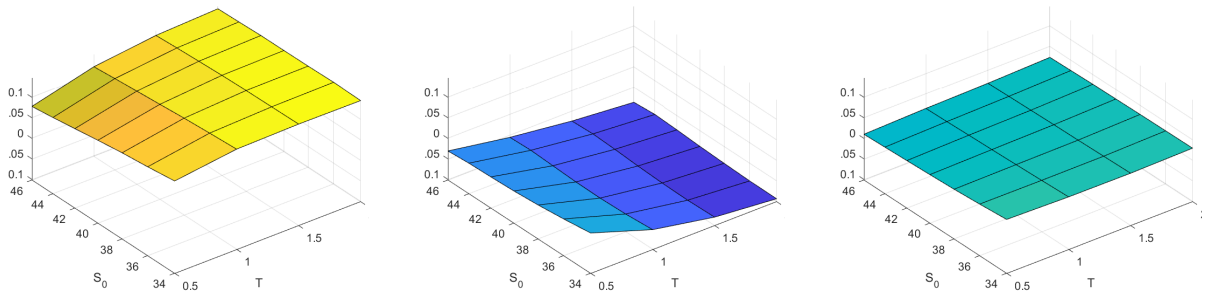
Results from panel (**a**) show the estimated variance minimizing drift parameter for an FDM algorithm with $N = 10^5$ paths obtained from a gradient descent procedure. Results from panel (**b**) show the saddle point approximation of the variance minimizing drift parameter for a European put option following the methodology of Glasserman et al. (1999).

Figures 2.3.7–2.3.8 show the bias, standard deviation and RMSE efficiencies when the GHS drift is used. The corresponding detailed results are found in Tables 2.A.2, 2.A.4, and 2.A.6 in Appendix 2.A. These plots look very similar to what is obtained with an optimal drift. In particular, Figure 2.3.7 shows that the LSM-s has the smallest bias in the vast majority of the cases, although the bias is slightly smaller with the IS LSM for the 3 out of the money short term options when $N = 1,000$ paths are used, Figure 2.3.5 shows that the standard deviations of the IS LSM and IS LSM-s are very close and much smaller than the LSM without importance sampling, and Figure 2.3.8 shows that as a result the IS LSM-s is significantly more efficient than the IS LSM when measured with RMSE.

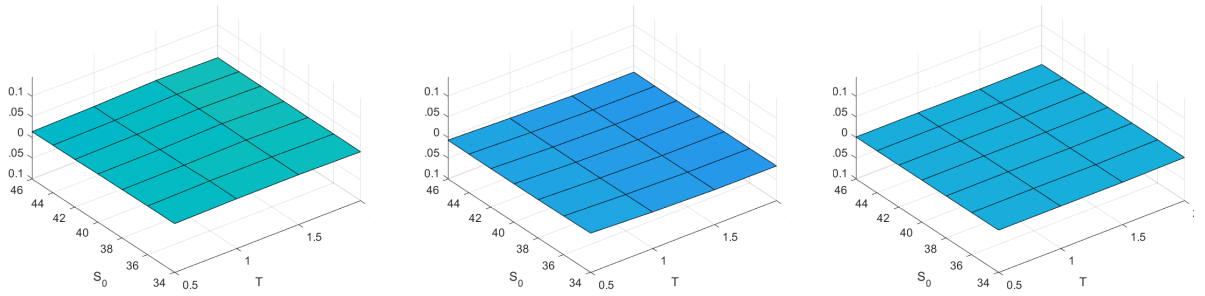
When using $N = 1,000$ paths only, with the GHS drift the average efficiency of the IS LSM-s is still large at 1.48 and the RMSE of the IS LSM-s is on average 18.1% smaller than that of the IS LSM. These results, therefore, demonstrate that our proposed method is indeed robust to the choice of (optimal) change of drift and hold true even when a sub-optimal change of drift is considered. Moreover, the results confirm that our proposed method should be particularly useful when one is restricted to using a low number of simulated paths for option pricing.

2.4 Extensions and future research

Importance sampling can be extended to a multivariate setting as in Moreni (2004). In particular, we can model D correlated GBM processes as a $(D \times 1)$ vector $\mathbf{S}_j = \{S_j^{(d)} : d = 1, \ldots, D\}$ with a $(D \times 1)$ vector of continuously compounded dividend yields $\{q^{(d)} : 1, \ldots, D\}$, and a $(D \times D)$ positive definite covariance matrix $\Sigma = ACA'$, where A is a lower-triangular matrix



(a) $N = 10^3$; LSM (b) $N = 10^3$; IS LSM (c) $N = 10^3$; IS LSM-s



(d) $N = 10^4$; LSM (e) $N = 10^4$; IS LSM (f) $N = 10^4$; IS LSM-s

Figure 2.3.7: Bias without and with GHS drift.

Bias results are calculated with respect to a Cox et al. (1979) binomial tree solution for the LSM and the LSM-s estimators. The left column illustrates results for the LSM without importance sampling, the center column LSM with importance sampling, and the right column LSM-s with importance sampling. The top row presents results for estimators using 1000 paths and the bottom row for 10,000 paths.

and C is a diagonal matrix.[5] Discretized paths are simulated under the nominal measure with

$$\mathbf{S}_j = \mathbf{S}_{j-1} \exp\left(\left(\mu - \frac{\text{diag}(\Sigma)}{2}\right)\Delta t + A\sqrt{C}\sqrt{\Delta t} z_j\right), \tag{2.21}$$

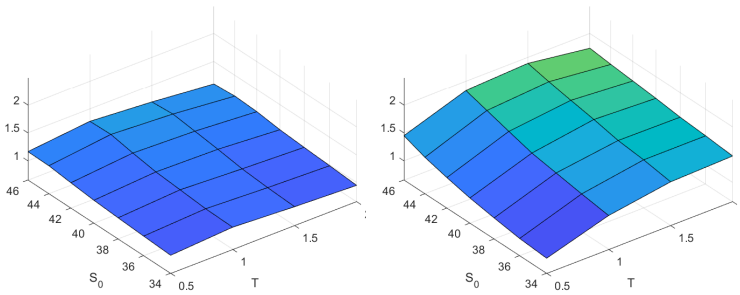
where $z_j \overset{IID}{\sim} \mathcal{N}(\mathbf{0}, I_D)$ is a $(D \times 1)$ vector of independent normal increments, and $\mu = [r - q^{(1)}, ..., r - q^{(D)}]'$ is a $(D \times 1)$ vector of mean log-returns defined as the difference between the continuously compounded risk-free rate and asset-specific dividend yield. Paths are simulated under the importance measure with the $(D\times1)$ drift adjustment Λ by using the same multivariate normal increments to generate $\tilde{z}_j \equiv z_j + \sqrt{\Delta t}\Lambda$. Under both measures the initial values of each path are the same, $\tilde{\mathbf{S}}_0 = \mathbf{S}_0$, and path are simulated under the importance measure with

$$\tilde{\mathbf{S}}_j = \tilde{\mathbf{S}}_{j-1} \exp\left(\left(\mu - \frac{\text{diag}(\Sigma)}{2} + A\sqrt{C}\Lambda\right)\Delta t + A\sqrt{C}\sqrt{\Delta t} z_j\right). \tag{2.22}$$

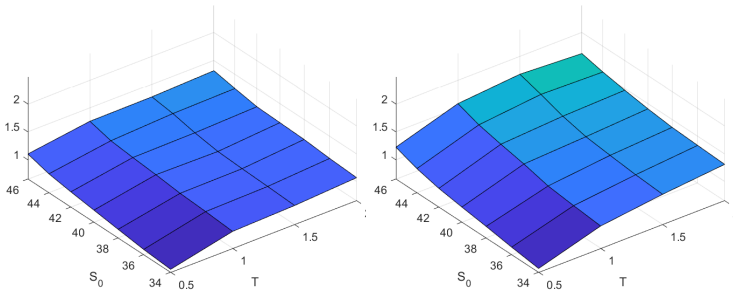
The k-step likelihood ratio for the discretized multivariate diffusion process takes the form

$$\frac{d\mathbb{P}}{d\tilde{\mathbb{P}}}(\tilde{\mathbf{S}}_j, k\Delta t) = \exp\left[-\Lambda k\sqrt{\Delta t}\left(\sum_{i=j-k}^{j} z_i\right) - \frac{1}{2}\|\Lambda\|^2 k\Delta t\right], \tag{2.23}$$

[5]Such a decomposition is easily computed with, for example, a Cholesky factorization.



(a) $N = 10^3$; IS LSM (b) $N = 10^3$; IS LSM-s



(c) $N = 10^4$; IS LSM (d) $N = 10^4$; IS LSM-s

Figure 2.3.8: RMSE efficiency with GHS drift.

RMSE efficiency results are calculated from a sample of $R = 10^7/N$ replications of the LSM and the LSM-s estimators. The left column illustrates results for the LSM without importance sampling, the center column LSM with importance sampling, and the right column LSM-s with importance sampling. The top row presents results for estimators using 1000 paths and the bottom row for 10,000 paths.

for any $j = 0, \ldots, J$ and $k = 0, \ldots, j$. One can then optimize the importance drift Λ with a gradient descent routine or use the saddle point approximation of Glasserman et al. (1999) when a closed-form solution to the analogous European-style option price is available.

In addition to the multivariate generalizations, it may also be desirable to consider importance sampling in models with fat-tailed distributions and variance dynamics for the underlying assets as exhibited in financial time series. When it comes to accommodating non-Gaussian fat-tailed returns, it is important to note that importance sampling techniques are not at all restricted to GBMs, as they only require that the importance distribution be absolutely continuous with respect to the nominal distribution. In such cases, the Radon–Nikodym derivative is well defined, and if it can be computed in a cost-effective way, importance sampling yields appreciable variance reduction for security pricing (see Boyle et al. (1997)). When it come to allowing for more flexible variance dynamics, Glasserman et al. (1999) also note that their approach can be used with the Hull and White (1987) stochastic volatility model or the Cox et al. (1977) mean-reverting square-root volatility model for interest rate derivatives.[6] For all these cases, the optimization of the drift term can be carried out with a gradient descent for any model specification or option payoff (Su and Fu 2000).

[6]They also present a more general framework for importance sampling in the Heath et al. (1990) model.

The shifted regression method proposed herein for American options can in principle be used with (all combinations of) the extensions outlined above simply by adjusting the likelihood ratio accordingly. Extensions to the multivariate setting are particular straightforward and simple to implement. However, to our knowledge applications of importance sampling for American option pricing in the presence of stochastic volatility and conditional heteroskedasticity (i.e., GARCH-type models) has yet to be thoroughly examined in the literature. The extension to the class of GARCH models, in particular with fat-tailed conditional distributions, is particularly interesting and challenging and it is currently the subject of ongoing research.[7]

2.5 Conclusion

This paper proposes a new method for pricing American options when using importance sampling as a variance reduction technique in simulation-and-regression based methods. Our suggested method uses regressions under the importance measure directly to determine the optimal early exercise strategy. The standard method for using importance sampling is instead to use regressions under the nominal measure.

We show that our proposed method offers several important benefits. First, our proposed method requires no prior knowledge of regression coefficients since the regressions are carried out directly with the importance paths and our approach, therefore, removes the overhead related to estimating an exercise strategy under the nominal measure. Second, our proposed method successfully reduces the bias that plague the standard LSM method with and without importance sampling by improving the accuracy of an exercise strategy. Third, the shifted LSM approach preserves the variance reduction in the standard implementation of importance sampling.

As a result, when a low number of paths is used, our method always improves on the standard method and reduces average root means squared error of estimated option prices by 22.5%. Hence, the Monte Carlo valuation of deep out of the money options, knock-in options or barrier options with American-style features may greatly benefit from the approach introduced in this paper. Moreover, our methods lends itself to simple parallel implementation where many independent jobs (with small sample size) are serially farmed out to processors in parallel because of the reduced bias compared to standard (and unshifted) estimators.

[7] We thank one of the reviewers for suggesting this extension.

Appendix

2.A Tables with detailed results

This appendix provides detailed results for the performance of the various methods used in this paper under the two different choices of change of measures used. Tables 2.A.1 and 2.A.2 present bias results, calculated with respect to a Cox et al. (1979) binomial tree solution, for the LSM, IS LSM, and IS LSM-s estimators across the 28 options considered for different number of paths and repetitions with the optimal drift and GHS approximate drift, respectively. Tables 2.A.3–2.A.6 present corresponding results for the standard deviation and root mean squared error, RMSE, respectively.

Table 2.A.1: Detailed bias results ($\times 10^3$) with optimized drift.

		$N = 10^3; R = 10^4$			$N = 10^4; R = 10^3$			$N = 10^5; R = 10^2$		
T	K	$\hat{g}$	$\hat{\bar{g}}$	$\hat{\bar{g}}^{(s)}$	$\hat{g}$	$\hat{\bar{g}}$	$\hat{\bar{g}}^{(s)}$	$\hat{g}$	$\hat{\bar{g}}$	$\hat{\bar{g}}^{(s)}$
0.5	34	117.1	-73.3	28.5	17.6	-18.9	1.7	-2.3	-4.6	-2.0
0.5	36	110.1	-62.7	24.1	15.8	-17.8	1.5	-3.1	-3.8	-1.4
0.5	38	104.2	-52.1	21.5	14.4	-15.5	3.0	-1.2	-5.8	-0.9
0.5	40	95.0	-42.1	17.6	15.9	-12.8	2.1	-0.0	-3.9	-1.5
0.5	42	90.9	-33.8	15.3	14.1	-11.0	2.1	-1.0	-3.3	-0.7
0.5	44	81.4	-30.0	12.2	15.4	-7.8	1.1	0.6	-3.9	-0.4
0.5	46	75.4	-30.2	9.1	12.0	-6.6	0.5	-1.2	-2.3	-1.1
1	34	138.5	-91.5	30.1	21.3	-19.6	0.1	4.6	-3.9	-2.3
1	36	138.6	-81.6	26.6	17.8	-18.8	2.4	-2.4	-6.4	-0.8
1	38	132.0	-74.2	25.4	15.1	-16.6	2.4	-3.3	-4.1	-1.1
1	40	128.1	-66.4	22.0	15.4	-15.5	1.5	1.6	-4.7	-1.0
1	42	124.4	-60.4	19.7	20.5	-17.0	2.4	2.7	-4.3	-0.9
1	44	118.7	-54.2	17.1	17.9	-12.6	1.5	-2.4	-3.8	-0.8
1	46	114.6	-54.9	15.8	18.4	-11.9	2.0	3.5	-4.3	0.3
1.5	34	141.5	-99.5	31.5	20.0	-20.7	3.4	-1.3	-4.4	-0.1
1.5	36	141.2	-93.2	29.2	22.5	-19.4	0.9	-0.1	-5.2	-1.8
1.5	38	141.9	-89.7	27.1	18.7	-18.6	3.6	-0.2	-5.3	-1.9
1.5	40	138.3	-82.4	24.8	13.7	-18.2	2.1	-0.1	-4.0	-0.3
1.5	42	134.2	-76.1	23.4	15.5	-15.8	2.4	0.9	-4.8	0.3
1.5	44	131.7	-72.6	20.0	15.2	-16.4	2.4	-0.5	-3.4	-1.7
1.5	46	129.2	-74.3	18.9	15.6	-14.3	1.7	-1.0	-4.0	-1.3
2	34	141.8	-104.3	30.6	16.7	-20.9	1.7	0.7	-4.5	-2.2
2	36	144.0	-98.8	29.0	13.1	-20.4	1.9	-1.0	-4.2	-2.1
2	38	142.0	-95.7	26.0	17.2	-17.5	2.4	-0.1	-4.1	-0.8
2	40	141.0	-91.8	24.8	14.5	-18.3	1.8	-1.1	-5.6	-1.7
2	42	139.4	-87.1	27.0	18.7	-17.9	2.3	5.3	-4.4	-1.4
2	44	141.6	-84.5	23.2	17.0	-18.5	3.1	3.2	-4.2	-1.2
2	46	137.6	-87.9	21.8	19.2	-16.8	1.1	1.3	-2.9	-0.7

Table 2.A.2: Detailed bias results ($\times 10^3$) with GHS drift.

		$N = 10^3; R = 10^4$			$N = 10^4; R = 10^3$			$N = 10^5; R = 10^2$		
T	K	$\hat{g}$	$\hat{\hat{g}}$	$\hat{\hat{g}}^{(s)}$	$\hat{g}$	$\hat{\hat{g}}$	$\hat{\hat{g}}^{(s)}$	$\hat{g}$	$\hat{\hat{g}}$	$\hat{\hat{g}}^{(s)}$
0.5	34	117.1	-3.0	26.7	17.6	-5.8	1.3	-2.3	-3.1	-2.2
0.5	36	110.1	-11.8	21.8	15.8	-7.2	2.1	-3.1	-3.5	-2.4
0.5	38	104.2	-20.3	18.7	14.4	-8.1	0.6	-1.2	-1.5	-2.0
0.5	40	95.0	-25.2	15.4	15.9	-8.9	-0.2	-0.0	-2.0	-1.5
0.5	42	90.9	-26.4	13.2	14.1	-8.5	-0.2	-1.0	-3.1	-2.2
0.5	44	81.4	-28.6	10.2	15.4	-8.0	0.6	0.6	-3.0	-1.4
0.5	46	75.4	-32.2	8.9	12.0	-7.4	0.4	-1.2	-2.8	-0.9
1	34	138.5	-56.7	23.8	21.3	-14.5	-0.2	4.6	-2.3	-0.6
1	36	138.6	-59.7	22.2	17.8	-13.2	0.5	-2.4	-1.9	-0.7
1	38	132.0	-59.7	21.2	15.1	-13.5	1.6	-3.3	-5.0	-1.0
1	40	128.1	-61.3	18.9	15.4	-14.7	0.2	1.6	-3.7	0.1
1	42	124.4	-58.5	17.1	20.5	-12.5	0.7	2.7	-2.3	-1.4
1	44	118.7	-54.8	14.7	17.9	-14.0	0.1	-2.4	-2.5	-1.8
1	46	114.6	-57.2	14.8	18.4	-13.4	1.5	3.5	-3.8	-0.4
1.5	34	141.5	-81.4	23.1	20.0	-15.6	0.5	-1.3	-4.6	-0.6
1.5	36	141.2	-85.0	23.5	22.5	-16.4	1.4	-0.1	-3.6	-1.5
1.5	38	141.9	-83.5	22.0	18.7	-16.7	3.7	-0.2	-4.0	-2.6
1.5	40	138.3	-80.4	21.6	13.7	-19.0	1.0	-0.1	-4.1	-2.1
1.5	42	134.2	-76.1	19.1	15.5	-17.0	1.7	0.9	-6.1	-1.8
1.5	44	131.7	-73.0	18.3	15.2	-15.7	3.1	-0.5	-4.3	-1.7
1.5	46	129.2	-75.4	17.8	15.6	-15.2	2.0	-1.0	-2.7	-1.6
2	34	141.8	-95.2	27.0	16.7	-21.4	1.8	0.7	-0.4	-2.9
2	36	144.0	-96.1	27.3	13.1	-18.8	3.6	-1.0	-6.6	-1.5
2	38	142.0	-94.0	27.0	17.2	-17.7	1.7	-0.1	-4.4	-4.8
2	40	141.0	-91.1	24.7	14.5	-17.2	3.5	-1.1	-5.3	-1.4
2	42	139.4	-87.7	25.5	18.7	-18.3	2.1	5.3	-3.9	-1.1
2	44	141.6	-84.6	23.8	17.0	-16.1	3.7	3.2	-3.3	-0.7
2	46	137.6	-89.2	21.7	19.2	-15.7	2.5	1.3	-3.6	0.4

Table 2.A.3: Detailed standard deviation results ($\times 10^2$) with optimized drift.

		$N = 10^3; R = 10^4$			$N = 10^4; R = 10^3$			$N = 10^5; R = 10^2$		
T	K	$\hat{g}$	$\hat{\bar{g}}$	$\hat{\bar{g}}^{(s)}$	$\hat{g}$	$\hat{\bar{g}}$	$\hat{\bar{g}}^{(s)}$	$\hat{g}$	$\hat{\bar{g}}$	$\hat{\bar{g}}^{(s)}$
0.5	34	18.2	8.1	8.1	5.7	2.5	2.5	1.7	0.8	0.9
0.5	36	17.6	7.5	7.7	5.6	2.2	2.4	1.5	0.7	0.8
0.5	38	16.8	7.0	7.2	5.5	2.2	2.1	1.7	0.6	0.7
0.5	40	15.9	6.3	6.5	4.9	1.9	2.0	1.4	0.7	0.7
0.5	42	14.7	5.7	5.9	4.8	1.8	1.9	1.3	0.5	0.6
0.5	44	13.4	4.9	5.1	4.4	1.6	1.5	1.4	0.5	0.4
0.5	46	12.2	4.3	4.5	3.9	1.3	1.3	1.4	0.4	0.4
1	34	21.6	10.0	9.9	6.6	3.0	3.0	2.3	0.8	1.0
1	36	21.2	9.4	9.4	6.6	2.9	2.9	2.0	0.9	0.9
1	38	21.1	8.9	9.1	6.6	2.9	2.9	2.2	0.9	0.9
1	40	20.4	8.4	8.7	6.4	2.6	2.5	1.8	0.9	0.8
1	42	19.4	7.9	8.1	6.3	2.5	2.5	1.8	0.8	0.8
1	44	18.3	7.3	7.5	6.0	2.2	2.3	1.9	0.6	0.7
1	46	17.4	6.7	6.9	5.6	2.1	2.1	1.8	0.6	0.7
1.5	34	23.6	10.8	10.9	7.8	3.2	3.3	2.2	1.0	1.1
1.5	36	23.6	10.6	10.5	7.8	3.2	3.3	2.2	1.1	1.1
1.5	38	23.1	10.1	10.1	7.6	3.1	3.1	2.1	1.0	1.0
1.5	40	22.5	9.7	9.8	7.1	2.9	3.1	2.3	0.9	0.9
1.5	42	22.0	9.3	9.4	6.9	2.9	2.9	2.2	0.8	0.9
1.5	44	21.3	8.8	9.0	6.8	2.6	2.7	2.1	0.7	0.9
1.5	46	20.7	8.3	8.5	6.5	2.5	2.5	2.1	0.8	0.7
2	34	25.2	11.4	11.4	8.0	3.5	3.5	2.6	1.0	1.2
2	36	25.1	11.3	11.3	8.1	3.5	3.5	2.3	1.0	1.1
2	38	24.7	10.9	11.1	7.4	3.4	3.4	2.4	1.0	1.1
2	40	24.5	10.5	10.6	7.6	3.3	3.2	2.3	1.1	1.0
2	42	24.2	10.2	10.2	7.6	3.0	3.1	2.2	0.9	1.1
2	44	23.2	9.7	9.9	7.2	2.9	3.0	2.3	0.9	0.9
2	46	22.6	9.3	9.4	7.1	2.8	2.8	2.1	0.9	0.9

Table 2.A.4: Detailed standard deviation results ($\times 10^2$) with GHS drift.

		$N = 10^3; R = 10^4$			$N = 10^4; R = 10^3$			$N = 10^5; R = 10^2$		
T	K	$\hat{g}$	$\hat{\hat{g}}$	$\hat{\hat{g}}^{(s)}$	$\hat{g}$	$\hat{\hat{g}}$	$\hat{\hat{g}}^{(s)}$	$\hat{g}$	$\hat{\hat{g}}$	$\hat{\hat{g}}^{(s)}$
0.5	34	18.2	10.9	10.9	5.7	3.6	3.6	1.7	1.2	1.1
0.5	36	17.6	10.4	10.3	5.6	3.2	3.2	1.5	1.0	1.0
0.5	38	16.8	9.4	9.5	5.5	2.9	3.0	1.7	0.8	0.9
0.5	40	15.9	8.7	8.6	4.9	2.7	2.7	1.4	0.8	0.9
0.5	42	14.7	7.7	7.7	4.8	2.4	2.3	1.3	0.8	0.8
0.5	44	13.4	6.7	6.7	4.4	2.1	2.1	1.4	0.6	0.7
0.5	46	12.2	5.9	5.8	3.9	1.8	1.9	1.4	0.6	0.6
1	34	21.6	11.5	11.2	6.6	3.4	3.7	2.3	1.1	1.1
1	36	21.2	10.9	10.8	6.6	3.3	3.5	2.0	0.9	0.9
1	38	21.1	10.2	10.1	6.6	3.2	3.1	2.2	0.9	0.8
1	40	20.4	9.5	9.3	6.4	3.0	2.9	1.8	1.0	1.0
1	42	19.4	8.7	8.6	6.3	2.7	2.8	1.8	0.7	0.8
1	44	18.3	8.0	8.0	6.0	2.5	2.6	1.9	0.8	0.8
1	46	17.4	7.3	7.3	5.6	2.3	2.3	1.8	0.7	0.7
1.5	34	23.6	11.6	11.2	7.8	3.7	3.7	2.2	1.1	1.0
1.5	36	23.6	11.2	10.8	7.8	3.3	3.3	2.2	1.1	1.0
1.5	38	23.1	10.5	10.4	7.6	3.2	3.3	2.1	1.0	0.9
1.5	40	22.5	9.9	9.7	7.1	3.1	3.1	2.3	1.0	0.9
1.5	42	22.0	9.5	9.4	6.9	2.9	2.9	2.2	1.0	0.9
1.5	44	21.3	8.9	8.9	6.8	2.6	2.7	2.1	1.0	0.8
1.5	46	20.7	8.5	8.3	6.5	2.6	2.6	2.1	0.9	0.8
2	34	25.2	11.6	11.3	8.0	3.6	3.7	2.6	1.0	1.1
2	36	25.1	11.2	11.0	8.1	3.4	3.4	2.3	1.2	1.1
2	38	24.7	11.0	10.7	7.4	3.5	3.4	2.4	1.2	1.1
2	40	24.5	10.5	10.5	7.6	3.3	3.2	2.3	1.0	0.9
2	42	24.2	10.1	10.2	7.6	3.1	3.2	2.2	0.9	1.0
2	44	23.2	9.8	10.1	7.2	3.1	2.9	2.3	0.8	1.0
2	46	22.6	9.5	9.5	7.1	3.0	2.7	2.1	0.9	0.9

Table 2.A.5: Detailed RMSE efficiency results with optimized drift.

		$N = 10^3; R = 10^4$		$N = 10^4; R = 10^3$		$N = 10^5; R = 10^2$	
T	K	$\hat{\hat{g}}$	$\hat{\hat{g}}^{(s)}$	$\hat{\hat{g}}$	$\hat{\hat{g}}^{(s)}$	$\hat{\hat{g}}$	$\hat{\hat{g}}^{(s)}$
0.5	34	0.98	1.52	0.90	1.43	0.81	0.95
0.5	36	1.12	1.56	1.05	1.42	0.99	1.05
0.5	38	1.28	1.64	1.13	1.63	0.94	1.39
0.5	40	1.43	1.73	1.22	1.55	0.86	1.10
0.5	42	1.61	1.85	1.39	1.64	1.29	1.29
0.5	44	1.71	1.98	1.63	2.13	1.18	2.10
0.5	46	1.72	2.13	1.71	2.13	2.03	2.34
1	34	0.89	1.48	0.94	1.30	1.50	1.33
1	36	1.03	1.59	0.97	1.34	0.80	1.21
1	38	1.15	1.63	1.05	1.36	1.18	1.41
1	40	1.25	1.68	1.15	1.57	0.77	1.13
1	42	1.32	1.76	1.17	1.66	0.94	1.33
1	44	1.40	1.83	1.49	1.69	1.55	1.83
1	46	1.40	1.94	1.46	1.83	1.41	1.72
1.5	34	0.87	1.43	1.09	1.42	0.97	1.09
1.5	36	0.95	1.53	1.16	1.46	0.89	0.94
1.5	38	1.00	1.58	1.17	1.49	0.86	0.99
1.5	40	1.07	1.61	1.12	1.31	1.23	1.44
1.5	42	1.14	1.67	1.13	1.48	1.24	1.33
1.5	44	1.19	1.72	1.25	1.56	1.58	1.29
1.5	46	1.19	1.79	1.37	1.69	1.39	1.98
2	34	0.88	1.46	1.01	1.34	1.36	1.06
2	36	0.93	1.48	1.04	1.35	1.13	0.94
2	38	0.96	1.50	0.99	1.27	1.17	1.29
2	40	1.02	1.59	1.06	1.44	0.88	1.31
2	42	1.08	1.64	1.26	1.55	1.27	1.08
2	44	1.12	1.68	1.16	1.44	1.39	1.46
2	46	1.07	1.75	1.27	1.63	1.31	1.40

Table 2.A.6: Detailed RMSE efficiency results with GHS drift.

		$N = 10^3; R = 10^4$		$N = 10^4; R = 10^3$		$N = 10^5; R = 10^2$	
T	K	$\hat{\hat{g}}$	$\hat{\hat{g}}^{(s)}$	$\hat{\hat{g}}$	$\hat{\hat{g}}^{(s)}$	$\hat{\hat{g}}$	$\hat{\hat{g}}^{(s)}$
0.5	34	0.98	0.93	0.65	0.66	0.35	0.53
0.5	36	0.99	0.97	0.78	0.81	0.54	0.57
0.5	38	1.05	1.04	0.88	0.91	0.96	0.92
0.5	40	1.05	1.11	0.77	0.86	0.65	0.48
0.5	42	1.12	1.22	0.96	1.13	0.66	0.69
0.5	44	1.16	1.33	1.05	1.20	0.99	0.94
0.5	46	1.14	1.45	1.08	1.16	1.22	1.18
1	34	1.00	1.25	0.84	0.88	1.01	1.02
1	36	1.04	1.30	0.91	0.95	1.22	1.17
1	38	1.11	1.41	0.96	1.18	1.05	1.62
1	40	1.13	1.52	0.97	1.25	0.73	0.80
1	42	1.19	1.63	1.22	1.37	1.33	1.10
1	44	1.25	1.68	1.17	1.44	1.24	1.24
1	46	1.25	1.81	1.20	1.60	1.23	1.61
1.5	34	0.95	1.42	0.99	1.16	0.84	1.21
1.5	36	0.96	1.49	1.18	1.46	1.00	1.12
1.5	38	1.01	1.55	1.19	1.33	0.96	1.25
1.5	40	1.07	1.65	1.01	1.36	1.06	1.47
1.5	42	1.12	1.68	1.11	1.48	0.90	1.43
1.5	44	1.17	1.75	1.28	1.60	0.92	1.57
1.5	46	1.15	1.88	1.23	1.60	1.22	1.42
2	34	0.93	1.48	0.93	1.20	1.62	1.27
2	36	0.96	1.54	1.09	1.38	0.64	0.96
2	38	0.97	1.58	0.96	1.25	0.93	1.06
2	40	1.03	1.61	1.09	1.41	1.02	1.48
2	42	1.08	1.65	1.17	1.47	1.33	1.29
2	44	1.10	1.63	1.14	1.54	1.67	1.38
2	46	1.04	1.70	1.19	1.67	1.29	1.29

Chapter 3

Bias Correction in the Least-Squares Monte Carlo Algorithm

3.1 Introduction

The pricing of derivatives with early-exercise features rests on a dynamic programming representation of the optimal stopping-time problem (Tilley 1993). The valuation task then boils down to estimating a series of continuation values, which can be approximated with regressions in a Monte Carlo simulation framework (Carriere 1996). Along with the seminal work of Boyle (1977) on simulation-based option pricing, these breakthroughs fostered the advent of a rich strand of literature that aims to develop efficient regression-based Monte Carlo simulation techniques for American-style option pricing. One important advantage of regression-based estimators is their flexibility, which tremendously eases the implementation in multivariate settings in comparison to, for instance, the stochastic mesh of Broadie and Glasserman (2004). Notable examples include the estimators proposed by Tsitsiklis and Van Roy (2001) and Longstaff and Schwartz (2001). The latter method, also called the Least-Squares Monte Carlo (LSM) method, enjoys exceptional popularity among practitioners due to its robustness, and reliability (Stentoft 2004*a*, 2014).

The LSM estimator, though almost-surely convergent and hence consistent (Clément et al. 2002, Stentoft 2004*b*), is affected by both positive and negative sources of bias in any actual application when a finite parameterization of the early exercise strategy and number of simulated paths are used. The bias is notoriously difficult to deal with because it has several sources. Previous related work in Kan and Reesor (2012) focused on the out-of-sample LSM estimator to approximate the bias caused by a sub-optimal estimator strategy. This paper focusses on the bias that occurs when the early-exercise strategy depends on future option values in a regression-based simulation method for American option pricing, thereby explicitly accounting for both sub-optimality and foresight bias. We first derive a theoretical expression for the foresight bias that holds true for general diffusion processes and payoff structures. We then show that a normal approximation of continuation values can lead to a tractable expression of local foresight bias. Local foresight bias corrections, complemented by the local sub-optimality bias corrections presented in Kan and Reesor (2012) then allow to efficiently correct the bias of an in-sample LSM estimator. Numerical results show that our proposed method reduces overall

estimator bias across a wide range of option characteristics.

To set the scene for the discussion to follow, consider a partition of time indexed by $j = 0, 1, 2, \ldots, J$ and denote as $\mathcal{V}_j(\mathcal{X}_j)$ the time-j perfect (yet unfeasible) price estimator given a cross-section of simulated asset price paths $\mathcal{X}_j$. That is, the operator $\mathcal{V}_j(\cdot)$ uses the true optimal early-exercise strategy at time j. Regression-based estimators in general and the LSM estimator in particular uses a finite parameterization of the early exercise strategy, and we write this as $v(\hat{\gamma}(\mathcal{X}_j, \mathcal{X}_{j+1}, \ldots, \mathcal{X}_J), \mathcal{X}_j)$, because the estimated early-exercise strategy $\hat{\gamma}(\mathcal{X}_j, \mathcal{X}_{j+1}, \ldots, \mathcal{X}_J)$ depends on all future asset prices and exercise decisions are made with the cross-section of asset paths $\mathcal{X}_j$ from the same sample that is used to estimate the early-exercise strategy. The (local to time j) LSM estimator bias can then be expressed as $\mathbb{E}\left[v(\hat{\gamma}(\mathcal{X}_j, \mathcal{X}_{j+1}, \ldots, \mathcal{X}_J), \mathcal{X}_j) - \mathcal{V}_j(\mathcal{X}_j)\right]$.

We can now explain how this bias can be decomposed into three specific components: approximation bias, sub-optimality bias, and foresight bias. To do so, we define two alternative LSM estimators. The first of these, which we denote $v(\gamma_j, \mathcal{X}_j)$, uses the optimal parameterization of the finite order approximation of the true exercise strategy with no model error. The second, which we denote $v(\hat{\gamma}(\tilde{\mathcal{X}}_j), \mathcal{X}_j)$, uses an independent but identically distributed cross-section of simulated paths $\tilde{\mathcal{X}}_j$ to estimate the early-exercise strategy. We then decompose the LSM estimator bias as

$$\mathbb{E}\left[\underbrace{v(\hat{\gamma}(\mathcal{X}_j, \mathcal{X}_{j+1}, \ldots, \mathcal{X}_J), \mathcal{X}_j) - v(\hat{\gamma}_j(\tilde{\mathcal{X}}_j), \mathcal{X}_j)}_{\text{Foresight bias}} + \underbrace{v(\hat{\gamma}_j(\tilde{\mathcal{X}}_j), \mathcal{X}_j) - v(\gamma_j, \mathcal{X}_j)}_{\text{Sub-optimality bias}} + \underbrace{v(\gamma_j, \mathcal{X}_j) - \mathcal{V}_j(\mathcal{X}_j)}_{\text{Approximation bias}}\right], \tag{3.1}$$

where the first term is the foresight bias studied in this paper. Numerical approximations of the relative size of each bias components are presented in Tables 3.A.1 and 3.A.2 of Appendix 3.A for various LSM configurations. These results demonstrate that foresight bias is an important contribution, often the largest one in relative size, to the LSM estimator bias.

We now briefly detail the "flaws" in the LSM algorithm that give rise to these different biases. First, the approximation bias relates to expected difference between an estimator that uses the true optimal exercise strategy and an estimator that uses the true approximation of the exercise strategy and is therefore due to the misspecification of the finite-order polynomial approximations of continuation values. As a square-integrable random variable, the continuation value belongs to the set of Hilbert spaces, and hence takes the form of a countable (infinite) linear combination of filtration-adapted basis functions (Royden 1988). Owing to the computational constraint of using a finite set of basis functions, an optimal parameterization of the LSM strategy is sub-optimal. Approximation bias is therefore a source of negative bias that is inherent to the user-chosen set of basis functions. As shown in Table 3.A.1 of Appendix 3.A, expanding the set of basis functions used in the regressions reduces approximation bias, such that we report values between -1.79% and -0.08% of the option price when using a cubic approximation of the continuation value, compared to values between -10.87% and -1.10% and when using a linear specification. We also note in Table 3.A.2 of Appendix 3.A that the approximation bias increases as the frequency of exercise opportunities grow. Indeed, more exercise opportunities also mean more opportunities to make exercise errors, which is reflected by the relative size increase of approximation bias.

Second, the sub-optimality bias relates to the expected difference between an estimator

that uses the true approximation of the exercise strategy and the estimator that estimates the approximation of the exercise strategy with an independent sample and therefore arises from the uncertainty of regression coefficients. This bias is also negatively signed because of the resulting sub-optimal exercise strategy. The first and most obvious reason for model estimation errors is the fact that regressions are performed with finite samples. Incorrect stopping times estimated from a poor or inadequate sample of option payoffs are particularly prevalent when valuing a deep out-of-the-money (OTM) option written on several underlying assets. This issue can be resolved directly, either by increasing the number of simulated paths or using importance sampling (Moreni 2003, Boire et al. 2021*a*). A second cause of model estimation error is the fact that classical assumptions of the ordinary least-squares (OLS) framework are generally violated in the LSM. For example, Fabozzi et al. (2017) suggest that error heteroskedasticity can be dealt with by using a weighted ordinary least-squares (wOLS).[1] Alternatively, the sub-optimality bias approximations can be used to correct the bias of cross-sections of LSM option values (Kan and Reesor 2012). It is clear from Tables 3.A.1 and 3.A.2 of Appendix 3.A that sub-optimality bias is an important source of bias in small samples. Indeed, it ranges between -11.23% and -2.12% of the option price when the sample size is 50, but is always above -0.31% when the sample size is 12800. We also observe that sub-optimality bias increases with the polynomial approximation order and the frequency of admissible exercise times.

Finally, there is foresight bias. Kan and Reesor (2012) gives continuation values using the same set of basis functions and regression coefficient estimated using an independent set of sample paths. We define foresight bias as the positive bias that arises due to the dependence between the regression-based exercise strategy and future option values. To be specific, when too many basis functions are used in proportion to the number of simulated paths, overfitting provides a look ahead for the estimated continuation values. The most straightforward way to reduce foresight bias is then to increase the number of simulated paths. Indeed, as the relative influence of each path is diluted in a large sample, overfitting becomes a negligible factor. With a fixed sample size, regularization and shrinkage methods can be used to reduce overfitting (Friedman et al. 2001), but the cost of cross-validating shrinkage parameters far exceeds the reward. Otherwise, foresight bias is difficult to alleviate without increasing other sources of bias. For instance, Table 3.A.1 shows that a lower number of regressors reduces overfitting and foresight bias, but also increases approximation bias. For the sample sizes considered in our experiments, we find that foresight bias is the most prominent source of variance in nearly all LSM configurations, reaching 53.73% of the option price when using fourth order polynomials with 50 simulated paths, but always below 1% when using 12800 paths. Furthermore, Table 3.A.2 of Appendix 3.A shows that that, like approximation bias, foresight bias is exacerbated as the frequency of exercise opportunities grows.

We are the first to propose a method to approximate and correct foresight bias, allowing practitioners to implement the in-sample LSM with a large number of regressors and a small number of simulated paths without being concerned with overfitting. We first generalize the bias approximation of Kan et al. (2009) and Kan and Reesor (2012) to cases where exercise decisions are allowed to depend on future cash flows along a given path. Next, we derive an expression of the foresight bias caused by exercise decisions at each period, termed *local* foresight bias, and show how sample quantities can be used to approximate it. Like the local

[1] Control variates can also be used to improve continuation value predictions (Rasmussen 2005).

sub-optimality bias approximation of Kan et al. (2009), Whitehead et al. (2012), and Kan and Reesor (2012), the approximation presented herein holds for general multivariate asset price processes and payoff structures. Finally, we propose a foresight bias-corrected LSM estimator. We show that the resulting estimator has reduced bias in finite samples compared with the standard LSM estimator along with the large-sample convergence properties.

Our numerical experiments clearly indicate the FS-LSM estimator reduces the bias of the LSM estimator across a wide range of option characteristics and underlying processes. Generally, we find that the performance of the bias-corrected estimator increases as the sample size decreases, is independent of the dimension of the underlying asset, increases with the number of basis functions, and increases with the frequency of exercise opportunities. These findings persist when a jump-diffusion process is used for the underlying. Finally, the bias-corrected estimator variance is the same as the standard estimator, implying the general results noted above for bias hold for estimator RMSE.

Other methods have been proposed to estimate and remove the American option bias. For instance, Broadie and Glasserman (1995) use bootstrapping techniques in the stochastic tree to reduce the bias. Alternatively, Chau and Oosterlee (2019) use a bundling approach in the context of a stochastic grid. In the LSM algorithm, Woo et al. (2018) employ leave-one-out cross-validation techniques to remove overfitting from the estimation of continuation values. These techniques, however, complicate the implementation of the algorithms and incur significant computational costs. Compared to these bias correction methods, our proposed approach is fast, only requires to add a few lines of code to the LSM program, and provides an explicit approximation that gives valuable insight on the relationship between continuation value estimates and local bias.

The paper is organized as follows. Section 3.2 outlines the LSM algorithm. Section 3.3 presents the derivation of an analytical approximation of foresight bias, and a modified LSM algorithm to correct for foresight bias. Section 3.4 presents numerical results for estimators with various types of bias corrections. Finally, Section 3.5 concludes. The appendixes contain additional numerical results.

3.2 Pricing American options

On a complete continuous-time probability space $(\Omega, \mathcal{F}, (\mathcal{F})_{t\geq 0}, \mathbb{P})$, the value of an American option is the solution to an optimal stochastic control problem where we want to maximize the expected discounted payoff from exercising the option. For the sake of clarity, we omit discount factors, and write the value of an American option as

$$V_t = \sup_{\tau \in \mathcal{T}(t,T)} \mathbb{E}_t \left[X_\tau \right], \tag{3.2}$$

where $(X_t)_{t\geq 0}$ is a non-negative payoff process of the underlying asset $(S_t)_{t\geq 0}$ with an adequately chosen numéraire satisfying the martingale property of discounted option prices under the measure $\mathbb{Q}$ equivalent to $\mathbb{P}$. Finally, $\mathcal{T}(t,T) \subseteq [t,T]$ is a class of $\mathcal{F}$-adapted stopping times. An $\mathcal{F}$-adapted stopping time can be seen as the first time τ^* such that the payoff process X_t is in the exercise region $\mathcal{E}_t$, written as

$$\tau^* = \inf_{\tau \in \mathcal{T}(t,T)} \{\tau : X_\tau \in \mathcal{E}_\tau\}, \tag{3.3}$$

where $\mathcal{E}_t \in \mathcal{F}_t$ for all $t \geq 0$.

The first step in implementing any type of numerical algorithm and a Monte Carlo simulation approach in particular to price American options is to assume that time can be discretized. Thus, we consider a discrete-time formulation of the above problem with an evenly spaced partition of J time steps of length $\Delta t = T/J$ and N simulated paths. We slightly adjust the above notation such that the path-n time-j underlying asset price process is $(S_{n,j} : n = 1, \ldots, N; j = 0, \ldots, J)$, with corresponding discounted payoff process $(X_{n,j} : n = 1, \ldots, N; j = 0, \ldots, J)$, and estimated continuation value $\hat{C}_{n,j}$. Options with a discrete early exercise feature of this sort are termed Bermudan options. The continuously exercisable American option price is approximated by letting N tend to infinity (Glasserman 2003).

In the discrete-time regression-based simulation LSM approach proposed by Longstaff and Schwartz (2001) continuation values approximate the expected discounted option value conditional on delaying the exercise one more period. They are computed by regressing discounted option values on a set of basis function related to the current underlying price process, written as the $1 \times (L+1)$ vector $\psi(S_{n,j}) = (\psi_0, \psi_1(S_{n,j}), \ldots, \psi_L(S_{n,j}))$. We refer to a polynomial regression model of the form

$$\hat{V}_{n,j+1} = \sum_{\ell=0}^{L} \psi_\ell(S_{n,j})\gamma_{j,\ell} + \varepsilon_{n,j+1}, \tag{3.4}$$

for $n = 1, \ldots, N$ and $j = 0, \ldots, J-1$. The $N \times 1$ vector of regression errors $\varepsilon_{j+1} = (\varepsilon_{1,j+1}, \ldots, \varepsilon_{N,j+1})$ is such that $\mathbb{E}_j\left[\varepsilon_{j+1}\right] = \mathbf{0}$ and $\mathbb{E}_j\left[\varepsilon_{j+1}\varepsilon'_{j+1}\right] = W_{j+1}$. The $(L+1) \times 1$ vector of coefficient estimates $\hat{\gamma}_j = (\hat{\gamma}_{j,0}, \ldots, \hat{\gamma}_{j,L})$ is obtained by regressing the time-$(j+1)$ cross-section of discounted option cash flows, denoted by the $N \times 1$ vector $\hat{\mathbf{V}}_{j+1}$, on the $N \times (L+1)$ matrix of basis functions evaluated at time-j, denoted by Ψ_j. One can then write

$$\hat{\gamma}_j = \left(\Psi'_j \Psi_j\right)^{-1} \Psi'_j \hat{\mathbf{V}}_{j+1}. \tag{3.5}$$

The path-n, time-j cross-section of fitted continuation value then takes the form

$$\hat{C}_{n,j} = \psi(S_{n,j})\hat{\gamma}_j. \tag{3.6}$$

This least-squares regression fit is at the heart of the LSM algorithm as it is recursively compared to the payoff to determine exercise times. Denoting the path-n, time-j discounted estimated option value process as $\hat{V}_{n,j}$, the LSM dynamic program is written as

$$\begin{cases} \hat{V}_{n,J} = X_{n,J} \\ \hat{V}_{n,j} = \begin{cases} X_{n,j} & \text{if } (X_{n,j} \geq \hat{C}_{n,j}) \text{ and } \left(X_{n,j} > 0\right) \\ \hat{V}_{n,j+1} & \text{if } (X_{n,j} < \hat{C}_{n,j}) \text{ or } \left(X_{n,j} = 0\right) \end{cases} \quad j = J-1, \ldots, 1. \end{cases} \tag{3.7}$$

It is apparent that $\hat{C}_{n,j}$ locally determines the exercise strategy for path n at time j. Finally, the sample size-N LSM estimator is defined as

$$\hat{V}_0^{(N)} = \frac{1}{N}\sum_{i=0}^{N} \hat{V}_{i,0} = \frac{1}{N}\sum_{i=0}^{N} X_{i,k_i} \tag{3.8}$$

where $\tau_i = k_i \Delta t$ is the path-i stopping time from the LSM exercise strategy and X_{i,k_i} is the corresponding discounted payoff.

3.3 LSM bias correction

In the following, we outline the steps to construct a bias-corrected in-sample LSM estimator. Our bias-correction method extends the method of Kan and Reesor (2012) to cases where continuation values depend on future option values. Following the approach of Kan and Reesor (2012), we find an analytic expression of local bias when exercise decisions are allowed to depend on future option values, leading to a new expression of local bias for in-sample LSM estimators that includes foresight bias. Second, we show that we can approximate local bias in the in-sample LSM method by assuming that continuation values are normally distributed (i.e., this normal distribution is an approximation given by a central limit theorem). We provide details on how sample quantities can be used to obtain these approximations, and propose a modified in-sample LSM estimator that corrects for local bias. Furthermore, we argue that removing local bias also removes global bias.

3.3.1 Local bias decomposition

Since the overall LSM bias is not easily tractable, we instead analyze local bias, or the bias caused by incorrect decisions along specific paths at specific times. Furthermore, in the following we do not address the approximation bias that results from using a finite set of basis functions. Focusing on local bias simplifies the analysis of the LSM bias because it focuses on current exercise errors only and is independent from all previous errors that were potentially propagated in the algorithm. For the in-sample LSM, local bias comprises two components: the negative local sub-optimality bias discussed in Whitehead et al. (2012) and Kan and Reesor (2012), and the positive local foresight bias presented in this article. The fact that local bias components have opposite signs illustrates how the in-sample LSM estimator mixes elements of positive and negative bias.[2] In practice, when the sample size N is low in proportion to the number of basis functions L, the positive foresight bias is the most prominent of the two, and the estimator is then typically biased high. As N/L increases, foresight bias is quickly dominated by the negative sub-optimality bias, leading to a low-biased estimator. For large sample sizes the estimator is asymptotically unbiased when compared to the true approximated option value.

Total local bias

To analyze local foresight and sub-optimality biases, one has to disentangle their individual effects on the propagated option values along given paths at specific times. In turn, the sum of local sub-optimality and local foresight biases yields an expression of total local bias (excluding approximation bias) that holds true for general payoff structures, diffusion processes, and number of dimensions in an in-sample LSM setting.

We now introduce a bit of notation, where we drop the path-indexing subscript n for ease of exposition. Let $\bar{C}_j = \mathbb{E}_j[\hat{V}_{j+1}] = \mathbb{E}_j[\hat{C}_j]$ be the time-j continuation value obtained from a model with no estimation error. We denote the event of holding the option in the LSM algorithm as $\mathcal{H}_j = (X_j < \hat{C}_j) \cup (X_j = 0)$, and conversely the event of exercising as $\mathcal{E}_j = (X_j \geq \hat{C}_j) \cap (X_j > 0)$.

[2] In contrast, the estimators of Carriere (1996) and Tsitsiklis and Van Roy (2001) propagates high biased option values through the dynamic program, and the out-of-sample LSM propagates low biased values.

Similarly, we denote the event that holding the option is optimal as $\bar{\mathcal{H}}_j = (X_j < \bar{C}_j) \cup (X_j = 0)$, and conversely the event that exercise is optimal as $\bar{\mathcal{E}}_j = (X_j \geq \bar{C}_j) \cap (X_j > 0)$.

A key insight to obtain tractable local bias expressions is to recognize that the bias of the time-j estimated option value is the result of all previous exercise errors in the backward recursion of the dynamic program. Therefore, we say that the time-j *global* bias is attributed to all LSM exercise errors along the path at times $k : j+1 < k \leq J-1$. The time-j *local* bias is the bias caused by the exercise decision at time-$(j+1)$. Note that the time-$(J-1)$ local bias is zero, since the exact continuation value is given by the European option price. Both sources of bias come from the sequence of continuation values as estimated in Equation (3.6).

At any exercise time $t_j : j = 0, \ldots, J-2$, we then decompose the bias along each path into *local* and *global* bias components as follows

$$\begin{aligned}
\bar{C}_j - \mathbb{E}_j\left[V_{j+1}\right] &= \mathbb{E}_j\left[\hat{V}_{j+1} - V_{j+1}\right] \\
&= \mathbb{E}_j\left[\hat{V}_{j+2}\mathbb{1}_{\{\mathcal{H}_{j+1}\}} + X_{j+1}\mathbb{1}_{\{\mathcal{E}_{j+1}\}} - \max\left(V_{j+2}, X_{j+1}\right)\right] \\
&= \mathbb{E}_j\left[\hat{V}_{j+2}\mathbb{1}_{\{\mathcal{H}_{j+1}\}} + X_{j+1}\mathbb{1}_{\{\mathcal{E}_{j+1}\}} - \max\left(\bar{C}_{j+1}, X_{j+1}\right)\right] \quad \text{(local)} \\
&\quad + \mathbb{E}_j\left[\max\left(\bar{C}_{j+1}, X_{j+1}\right) - \max\left(V_{j+2}, X_{j+1}\right)\right], \qquad \text{(global)}
\end{aligned} \tag{3.9}$$

where $\mathbb{1}_{\{\omega\}}$ is the indicator function that is equal to one for the event $\omega \in \Omega$, and equal to zero otherwise. At time t_{J-2}, there is no global error because the optimal exercise strategy is known at maturity (i.e., $\bar{C}_J = X_J$). Therefore, the global bias component is zero, and the bias is composed of local bias only. Moreover, it immediately follows from the identity $|\max(x,y) - \max(u,v)| \leq |x-u| + |y-v|$ and Jensen's inequality that the time-j absolute global bias is bounded from above by the time-$(j+1)$ absolute global bias.

$$\begin{aligned}
\left|\mathbb{E}_j\left[\max\left(\bar{C}_{j+1}, X_{j+1}\right) - \max\left(V_{j+2}, X_{j+1}\right)\right]\right| &\leq \mathbb{E}_j\left[\left|\bar{C}_{j+1} - V_{j+2}\right|\right] \\
&= \mathbb{E}_j\left[\left|\bar{C}_{j+1} - \mathbb{E}_{j+1}[V_{j+2}]\right|\right].
\end{aligned} \tag{3.10}$$

As we move backwards in the recursion, incorrect (local) exercise decisions "contaminate" the sample of future cash flows. In turn, fitting continuation values to a contaminated sample tends to incur even more incorrect exercise decisions. All local biases then accumulate into global bias in the backward recursion. Since global bias can be thought of as the aggregate effect of future local biases, global bias can be suppressed by recursively correcting local biases at every time step. Therefore, we need only focus on local bias correction.

Observe that the $\mathcal{F}_{j+1}$ conditional expectation of $(\hat{V}_{j+2} - \bar{C}_{j+1})\mathbb{1}_{\{\bar{\mathcal{H}}_{j+1}\}}$ is zero because $\mathbb{1}_{\{\bar{\mathcal{H}}_{j+1}\}}$ is $\mathcal{F}_{j+1}$-measurable, and $\bar{C}_{j+1} = \mathbb{E}_{j+1}[\hat{V}_{j+2}]$. A new expression for the time-j local bias can then be obtained by subtracting this term from the local bias and using nested expectations as

follows

$$
\begin{aligned}
&\mathbb{E}_j\left[\hat{V}_{j+2}\mathbb{1}_{\{\mathcal{H}_{j+1}\}} + X_{j+1}\mathbb{1}_{\{\mathcal{E}_{j+1}\}} - \max\left(\bar{C}_{j+1}, X_{j+1}\right)\right] \\
=&\mathbb{E}_j\left[\hat{V}_{j+2}\mathbb{1}_{\{\mathcal{H}_{j+1}\}} + X_{j+1}\mathbb{1}_{\{\mathcal{E}_{j+1}\}} - \bar{C}_{j+1}\mathbb{1}_{\{\bar{\mathcal{H}}_{j+1}\}} - X_{j+1}\mathbb{1}_{\{\bar{\mathcal{E}}_{j+1}\}}\right] \\
=&\mathbb{E}_j\left[\mathbb{E}_{j+1}\left[\hat{V}_{j+2}\mathbb{1}_{\{\mathcal{H}_{j+1}\}} + X_{j+1}\mathbb{1}_{\{\mathcal{E}_{j+1}\}} - \bar{C}_{j+1}\mathbb{1}_{\{\bar{\mathcal{H}}_{j+1}\}} - X_{j+1}\mathbb{1}_{\{\bar{\mathcal{E}}_{j+1}\}} - (\hat{V}_{j+2} - \bar{C}_{j+1})\mathbb{1}_{\{\bar{\mathcal{H}}_{j+1}\}}\right]\right] \\
=&\mathbb{E}_j\left[\hat{V}_{j+2}\mathbb{1}_{\{\mathcal{H}_{j+1}\}} + X_{j+1}\mathbb{1}_{\{\mathcal{E}_{j+1}\}} - X_{j+1}\mathbb{1}_{\{\bar{\mathcal{E}}_{j+1}\}} - \hat{V}_{j+2}\mathbb{1}_{\{\bar{\mathcal{H}}_{j+1}\}}\right] \\
=&\mathbb{E}_j\left[\underbrace{\mathbb{1}_{\{\mathcal{H}_{j+1}\cap\bar{\mathcal{E}}_{j+1}\}}}_{\text{Incorrect hold}}\left(\hat{V}_{j+2} - X_{j+1}\right) + \underbrace{\mathbb{1}_{\{\mathcal{E}_{j+1}\cap\bar{\mathcal{H}}_{j+1}\}}}_{\text{Incorrect exercise}}\left(X_{j+1} - \hat{V}_{j+2}\right)\right].
\end{aligned}
\tag{3.11}
$$

When expressed in this way, it is clear that no local bias can come from correctly exercising or holding the option. Even if the estimated continuation value is a severely flawed exercise criterion, only incorrect exercise decisions lead to local bias.

The in-sample LSM estimator written in Equation (3.7) uses the same sample of simulated paths to estimate continuation values and option values. It is then apparent from Equation (3.6) that continuation values and future option values are $\mathcal{F}_j$-dependent. To derive an expression for the expected local bias at time j, we need to account for this dependence with the idendity

$$
\mathbb{E}_{j+1}\left[\mathbb{1}_{\{\mathcal{H}_{j+1}\}}\hat{V}_{j+2}\right] = \mathrm{Cov}_{j+1}\left[\mathbb{1}_{\{\mathcal{H}_{j+1}\}}, \hat{V}_{j+2}\right] + \mathbb{E}_{j+1}\left[\mathbb{1}_{\{\mathcal{H}_{j+1}\}}\right]\bar{C}_{j+1}. \tag{3.12}
$$

Taking the $\mathcal{F}_{j+1}$-conditional expectation inside the $\mathcal{F}_j$-conditional expectation expression of local bias in Equation (3.11), and using $\mathbb{E}_{j+1}[\hat{V}_{j+2}] = \bar{C}_{j+1}$ then yields

$$
\mathbb{E}_j\left[\mathbb{1}_{\{\mathcal{H}_{j+1}\cap\bar{\mathcal{E}}_{j+1}\}}\left(\hat{V}_{j+2} - X_{j+1}\right) + \mathbb{1}_{\{\mathcal{E}_{j+1}\cap\bar{\mathcal{H}}_{j+1}\}}\left(X_{j+1} - \hat{V}_{j+2}\right)\right] = \zeta_j + \xi_j, \tag{3.13}
$$

where

$$
\zeta_j = \mathbb{E}_j\left[\mathbb{1}_{\{\mathcal{H}_{j+1}\cap\bar{\mathcal{E}}_{j+1}\}}\left(\bar{C}_{j+1} - X_{j+1}\right) + \mathbb{1}_{\{\mathcal{E}_{j+1}\cap\bar{\mathcal{H}}_{j+1}\}}\left(X_{j+1} - \bar{C}_{j+1}\right)\right], \tag{3.14}
$$

and

$$
\xi_j = \mathbb{E}_j\left[\mathrm{Cov}_{j+1}\left[\mathbb{1}_{\{\mathcal{H}_{j+1}\}}, \hat{V}_{j+2}\right]\right]. \tag{3.15}
$$

In Sections 3.3.1 and 3.3.1, we interpret ζ_j and ξ_j as local sub-optimality bias and local foresight bias, respectively. We provide some intuition as to why $\zeta_j \leq 0$ and $\xi_j \geq 0$.

Local sub-optimality bias

When the fitted continuation value is estimated independently from future cash flow along a given simulation path, the exercise criterion is independent from future option values, such that the covariance term in Equation (3.15) is null and $\xi_j = 0$. Absent foresight bias, the time-j local bias is then entirely composed of ζ_j, the negative bias from a sub-optimal exercise strategy, termed *sub-optimality* bias. The derivations of local bias discussed in Kan and Reesor (2012) and Whitehead et al. (2012) make such independence assumptions in the case of an out-of-sample LSM pricing approach, and in a stochastic tree whose transition weights are estimated independently from future node values, respectively. In both cases, the continuation value model ensures that $\mathcal{H}_j$ and $\hat{V}_{j+1}$ are $\mathcal{F}_j$-independent for all $j = 0, \ldots, J-1$.

Moreover is easy to see that the local sub-optimality bias is negative by simply observing that $\hat{C}_{j+1} \geq X_{j+1}$ on the set $\bar{\mathcal{H}}_{j+1}$ and conversely $\hat{C}_{j+1} \leq X_{j+1}$ on the set $\bar{\mathcal{E}}_{j+1}$. Therefore, when the independence is satisfied, the estimator contains only negative local bias elements, and the local bias of the estimator is thus also negative. Correcting this negative source of bias would therefore yield an estimator that has a larger expected value. It is therefore not recommended to implement sub-optimality bias corrections alone in an in-sample implementation of the LSM, as it would exacerbate the positive bias observed in smaller samples. As shown in Tables 3.A.1 and 3.A.2 of Appendix 3.A, foresight bias is dominated by sub-optimality bias as N increases, and the sub-optimality bias correction alone will successfully reduce the estimator bias. That being said, we aim to find a bias correction term that is adequate for all sample sizes, reflecting the fact that there are conflicting sources of bias.

Local foresight bias

As suggested in Equation (3.15), local foresight bias measures the covariance between current exercise decisions and future option values. Indeed, positive bias arises because continuation values are based on future information that should not be available to derive an optimal exercise strategy. If there is significant overfitting, it is as if the strategy can predict future price movements. Such a strategy is then more likely to exercise an option right before it loses value, and more likely to keep it alive right before it gains value. The positive correlation between the event of exercise and future cash flows highlights the fact that in-sample LSM exercise times do not fit the definition of a $\mathcal{F}$-adapted stopping time. Indeed, because the in-sample LSM strategy is not restricted to the set of $\mathcal{F}$-adapted strategies, it can potentially outperform the optimal $\mathcal{F}$-adapted strategy. Moreover, it is apparent that ξ_j is positive as $\mathbb{1}_{\{\mathcal{H}_{j+1}\}}$ is monotonically non-decreasing in $\hat{C}_{j+1}$. The inequality follows from the fact that $\hat{C}_{j+1}$ and $\hat{V}_{j+1}$ are positively correlated.

3.3.2 Bias approximation

To derive the bias approximations, we need to make an assumption about the distribution of the continuation value estimates. Under general conditions, the $\mathcal{F}_j$-conditional distribution of regression coefficients $\hat{\gamma}_j$ given a sufficiently large sample size, follows a multivariate normal distribution (White 2001)

$$\hat{\gamma}_j \mid \mathcal{F}_j \sim \mathcal{N}\left(\gamma_j, s_j^2\right), \tag{3.16}$$

where $\mathcal{N}(\mu, \Sigma)$ denotes a multivariate normal random vector with $M \times 1$ mean vector μ and $M \times M$ variance covariance matrix Σ. Seeing that the fitted values are linear combinations of regression coefficients, a simple application of the Cramér-Wold device yields that the $\mathcal{F}_j$-conditional distribution of estimated fitted values $\hat{C}_{n,j}$ is also asymptotically normal. Specifically,

$$\hat{C}_{n,j} \mid \mathcal{F}_j \sim \mathcal{N}\left(\bar{C}_{n,j}, \sigma_{n,j}\right) \tag{3.17}$$

for all $j = 0, \ldots, J-1$ and $n = 1, \ldots, N$. Moreover, denoting the time-j variance-covariance matrix of the regression coefficients as $s_j^2 = \mathbb{E}_j\left[(\hat{\gamma}_j - \gamma_j)(\hat{\gamma}_j - \gamma_j)'\right]$, and the path-$n$, time-$j$

variance of the fitted value as $\sigma^2_{n,j} = \mathbb{E}_j\left[(\hat{C}_{n,j} - \bar{C}_{n,j})^2\right]$, we can write

$$s^2_j = \left(\Psi'_j\Psi_j\right)^{-1}\Psi'_j W_j \Psi_j \left(\Psi'_j\Psi_j\right)^{-1} \tag{3.18a}$$

$$\sigma^2_{n,j} = \psi_j(S_{n,j}) s^2_j \psi'_j(S_{n,j}). \tag{3.18b}$$

We can then estimate these quantities from regression errors $\hat{\varepsilon}_{j+1}$ as follows

$$\hat{s}^2_j = \left(\Psi'_j\Psi_j\right)^{-1}\Psi'_j \hat{W}_j \Psi_j \left(\Psi'_j\Psi_j\right)^{-1} \tag{3.19a}$$

$$\hat{\sigma}^2_{n,j} = \psi_j(S_{n,j}) \hat{s}^2_j \psi'_j(S_{n,j}), \tag{3.19b}$$

where $\hat{W}_j = \text{diag}(\hat{\varepsilon}_{j+1}\hat{\varepsilon}'_{j+1})$ is a heteroskedasticity-consistent (HC) estimator of the variance-covariance matrix W_j. Assuming that the limiting distribution of continuation values in (3.17) holds for $N < \infty$, we write the density as

$$f_{\hat{C}_{n,j}|\bar{C}_{n,j},\sigma_{n,j}}(\hat{c}) = \frac{1}{\sigma_{n,j}}\phi\left(\frac{\hat{c} - \bar{C}_{n,j}}{\sigma_{n,j}}\right).$$

Kan and Reesor (2012) then derive the following estimator of the time-j LSM local sub-optimality bias under the assumption that exercise decisions and future option values are independent

$$\hat{\zeta}_j(\hat{C}_{j+1}, X_{j+1}, \hat{\sigma}_{j+1}) = \left|\hat{C}_{j+1} - X_{j+1}\right| \Phi\left(\frac{-\left|\hat{C}_{j+1} - X_{j+1}\right|}{\hat{\sigma}_{j+1}}\right) - \hat{\sigma}_{j+1}\phi\left(\frac{\hat{C}_{j+1} - X_{j+1}}{\hat{\sigma}_{j+1}}\right), \tag{3.20}$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ are the cumulative distribution function (CDF) and the probability density function (PDF) of a standard normal random variable, respectively.

Using similar arguments (see derivations in Appendix 3.B), we obtain the following approximation of the time-j foresight bias

$$\hat{\xi}_j(\hat{C}_{j+1}, X_{j+1}, \hat{\sigma}_{j+1}) = \frac{\hat{\sigma}_{j+1}}{\sqrt{2}}\phi\left(\frac{\hat{C}_{j+1} - X_{j+1}}{\hat{\sigma}_{j+1}\sqrt{2}}\right). \tag{3.21}$$

Equations (3.20) and (3.21) suggests that local sub-optimality and foresight biases, respectively, are determined by the distance between the continuation value and the cash flow $|\hat{C}_{j+1} - X_{j+1}|$ and the volatility σ_{j+1} of the fitted value. First, it is apparent that the local foresight bias approximation decreases as $|\hat{C}_{j+1} - X_{j+1}|$ increases. This is consistent with the idea that an exercise error is more likely to occur if the estimated continuation value is close to the payoff, because small variations in continuation values are more likely to lead to different exercise decisions. In other words, overfitting is more liable to cause exercise errors if the optimal exercise frontier is close to the current option payoff. Conversely, if overfitting has a strong influence on the fit far away from the optimal exercise frontier, no additional bias will result from the continuation value error. Second, foresight bias increases as the volatility of the fitted continuation value increases. If we think of an overfitted model as a model that has low bias and high variance, our proposed approximation agrees with the idea that overfitting is the root cause of foresight bias.[3]

[3]This also provides insight as to how variance reduction techniques applied to continuation values (see Rasmussen (2005)) may be effective for reducing foresight bias. This is the subject of ongoing research.

3.3.3 Bias-corrected LSM estimator

We can now construct an estimator of total local bias by adding the estimator of local sub-optimality bias to the estimator of local foresight bias. Recursively subtracting this term from the option values therefore correct for both sub-optimality and foresight bias locally. As later discussed in Section 3.4, our proposed approach effectively reduces foresight bias.

Local bias correction

Letting $\hat{B}(\hat{C}_{n,j+1}, X_{n,j+1}, \hat{\sigma}_{n,j+1})$ be the path-n, time-j local bias estimator, we modify the dynamic program in Equation (3.7) and denote the path-n, time-j corrected discounted option value process as $\hat{V}^*_{n,j}$. The bias-corrected LSM dynamic program is

$$\begin{cases} \hat{V}^*_{n,J} = X_{n,J} \\ \hat{V}^*_{n,j} = \begin{cases} X_{n,j} - \hat{B}(\hat{C}^*_{n,j}, X_{n,j}, \hat{\sigma}^*_{n,j}) & \text{if } (X_{n,j} \geq \hat{C}^*_{n,j}) \text{ and } (X_{n,j} > 0) \\ V^*_{n,j+1} - \hat{B}(\hat{C}^*_{n,j}, X_{n,j}, \hat{\sigma}^*_{n,j}) & \text{if } (X_{n,j} < \hat{C}^*_{n,j}) \text{ and } (X_{n,j} > 0) \\ V^*_{n,j+1} & \text{if } X_{n,j} = 0 \end{cases} & j = J-1, \ldots, 0, \end{cases} \tag{3.22}$$

where the path-n, time-j continuation value approximation denoted by $\hat{C}^*_{n,j}$ in Equation (3.22) is the $\mathcal{F}_j$-conditional expectation of the corresponding bias-corrected option value $\hat{V}^*_{n,j}$. Similar to Equation (3.5), regression coefficients are now obtained as

$$\hat{\gamma}^*_j = \left(\Psi'_j \Psi_j\right)^{-1} \Psi'_j \hat{\mathbf{V}}^*_{j+1}, \tag{3.23}$$

where $\hat{\mathbf{V}}^*_{j+1}$ denotes the $N \times 1$ vector of time-$(j+1)$ corrected discounted option cash flows. In the modified LSM algorithm, the time-j cross-section of fitted continuation values then takes the form

$$\hat{\mathbf{C}}^*_j = \Psi_j \hat{\gamma}^*_j = \Psi_j \left(\Psi'_j \Psi_j\right)^{-1} \Psi'_j \hat{\mathbf{V}}^*_{j+1}. \tag{3.24}$$

The variance of the fitted continuation value $\hat{C}^*_{n,j}$, denoted $\hat{\sigma}^*_{n,j}$, is computed as in Equation (3.19) using the regression residuals $\hat{\varepsilon}^*_{j+1} = (\hat{\varepsilon}^*_{1,j+1} \ldots, \hat{\varepsilon}^*_{N,j+1})$ instead, where $\hat{\varepsilon}^*_{n,j+1} = \hat{V}^*_{n,j+1} - \hat{C}^*_{n,j}$. The modified LSM estimator with sample size-N is then defined as

$$\hat{V}^{*(N)}_0 = \frac{1}{N} \sum_{i=0}^{N} \hat{V}^*_{i,0} = \frac{1}{N} \sum_{i=0}^{N} X_{i,k^*_i}, \tag{3.25}$$

where $\tau^*_i = k^*_i \Delta t$ is the path-i stopping time from the modified LSM exercise strategy and X_{i,k^*_i} is the corresponding discounted payoff.

In the next section, our numerical experiments study two bias-corrected LSM estimators, in addition to the standard LSM in Equation (3.8). The first bias-corrected estimator is termed the F-LSM, because it only corrects for local foresight bias such that the local bias corrections are written

$$\hat{B}(\hat{C}^*_{n,j+1}, X_{n,j+1}, \hat{\sigma}^*_{n,j+1}) = \hat{\xi}_j(\hat{C}^*_{n,j+1}, X_{n,j+1}, \hat{\sigma}^*_{n,j+1}). \tag{3.26}$$

Table 3.3.1: Local bias corrections for the LSM, F-LSM, and FS-LSM algorithms.

Bias correction	LSM	F-LSM	FS-LSM
Approximation (-)			
Sub-optimality (-)			✓
Foresight (+)		✓	✓

Bias corrections for the LSM, F-LSM, and FS-LSM algorithms are summarized in terms of three sources of bias. A checkmark (✓) indicates that the source of bias is corrected.

The second bias-corrected estimator is termed the FS-LSM, because it corrects for both foresight and sub-obtimality biases locally such that the local bias corrections are written

$$\hat{B}(\hat{C}^*_{n,j+1}, X_{n,j+1}, \hat{\sigma}^*_{n,j+1}) = \hat{\zeta}_j(\hat{C}^*_{n,j+1}, X_{n,j+1}, \hat{\sigma}^*_{n,j+1}) + \hat{\xi}_j(\hat{C}^*_{n,j+1}, X_{n,j+1}, \hat{\sigma}^*_{n,j+1}). \tag{3.27}$$

Note also that for $j = J-1, \ldots, 0$, the local bias correction is only applied to ITM paths. Indeed, no local bias can come from holding an out-of-the-money (OTM) option as it is always optimal to hold. Thus, a local bias correction would be unwarranted in such instances.

If the bias approximations are consistent estimates of the local bias, the net bias of the LSM, F-LSM, and FS-LSM estimators can be summarized in Table 3.3.1. In this table, the approximation bias is the negative bias that comes from using a mispecified model with $L < \infty$, the sub-optimality bias comes from the estimation error of regression coefficients and is also negatively signed, and the foresight bias comes from overfitting and is positively signed. Recalling the three sources of bias discussed in the introduction (i.e., approximation bias, sub-optimality bias, and foresight bias), we can see that the in-sample LSM mixes elements of positive and negative biases, whereas the F-LSM and FS-LSM leave only negative sources of bias uncorrected. The uncorrected biases of the F-LSM estimator are approximation and sub-optimality biases, which should yield negatively-biased estimator. Similarly, the uncorrected bias of the FS-LSM is simply the approximation bias and should yield a negatively-baised estimator too.

Convergence

Consider a continuation value model similar to that in Equation (3.4) for $j = 1, \ldots, J-1$, except that we are using bias-corrected discounted cash flows, namely

$$\hat{V}^*_{n,j+1} = \sum_{\ell=0}^{L} \psi_\ell(S_{n,j})\gamma^*_{j,l} + \varepsilon^*_{n,j+1}. \tag{3.28}$$

We now prove that our proposed local bias approximation, $\hat{\zeta}_j + \hat{\xi}_j$, converges to zero in probability as $N \to \infty$ under the following assumptions:

1. $\mathbb{E}_0[X_{n,j}] < \infty, \quad \mathbb{E}_0\left[\psi(S_{n,j})'\psi(S_{n,j})\right] < \infty$
2. $\mathbb{E}_0\left[\psi(S_{n,j})'\psi(S_{n,j})\right]$ is non-singular

3. $\mathbb{E}_j[\varepsilon^*_{n,j+1}] = 0, \quad \mathbb{E}_j\left[\varepsilon^{*2}_{n,j+1}\right] = w^2_{n,j+1} < \infty$

4. $\mathbb{E}_0\left[\varepsilon^{*4}_{n,j+1}\right] < \infty, \quad \mathbb{E}_0\left[\left(\psi_\ell(S_{n,j})'\psi_{\tilde{\ell}}(S_{n,j})\right)^2\right] < \infty, \quad$ for all $\ell, \tilde{\ell} = 1, \ldots, \text{rank}(\Psi_j)$

5. $\mathbb{P}(\hat{C}^*_{n,j} = X_{n,j}) \overset{a.s.}{=} 0.$

Kan and Reesor (2012) showed that the sub-optimality bias approximation $\hat{\zeta}_j$ converges to zero in probability under these assumptions. Using similar arguments, it is straightforward to show that the same holds for the foresight bias approximation $\hat{\xi}_j$.

First, observe that $\hat{\gamma}^*_j$ converges to γ_j almost surely (see Lemma 3.2 of Clément et al. (2002)), and that $\hat{W}^*_j = \text{diag}(\hat{\varepsilon}^*_{j+1}\hat{\varepsilon}'^*_{j+1})$ converges in probability to W_j (White 1980) under Assumptions 1 to 4. It follows that $\hat{\sigma}^*_{n,j}$ converges in probability to $\sigma_{n,j}$, which in turn converges in probability to zero as $N \to \infty$. Moreover, $\hat{C}^*_{n,j}$ converges in probability to $\bar{C}_{n,j}$. Next, observe that $\bar{C}_j - X_j$ is bounded under Assumption 1, and almost surely not equal to zero under Assumption 4. Then, by virtue of Slutzky's theorem and the continuous mapping theorem, $(\hat{C}^*_{n,j} - X_{n,j})/\hat{\sigma}^*_{n,j}$ converges in probability to $-\infty$ or ∞, depending on whether $\bar{C}_j - X_j$ is negative or positive, respectively, because $\hat{\sigma}^*_{n,j} \to 0$ as $N \to \infty$. Finally, observe that the normal PDF $\phi(\cdot)$ is a continuous function to conclude that $\hat{\xi}_j$ converges to zero in probability as $N \to \infty$ by applying the continuous mapping theorem once more. It immediately follows that total local bias $\hat{\zeta}_j + \hat{\xi}_j$ converges to zero in probability as $N \to \infty$. The consistency of the LSM estimator is therefore preserved for the bias-corrected F-LSM and FS-LSM estimators.

Numerical issues

In the following, we express a number of concerns regarding numerical issues that may be absent in a standard LSM approach, but arise when computing the bias-corrected estimator. These are mainly due to the fact that the variance of the continuation value is now a parameter of interest required to compute local bias estimates. Consequently, extra care might be needed when performing regressions, and computing the variance of the fitted values.

The LSM method of Longstaff and Schwartz (2001) performs regressions using only ITM paths. This approach can enhance the regression fit for the relevant sample of ITM paths where an exercise decision is considered. It also reduces the computational burden of OLS by reducing the number of observations used to estimate regression coefficients. It is perfectly adequate to use this approach in standard LSM because the regression fit is the only object of interest. This is not true in a bias-corrected LSM because the estimated variances of continuation values tend to be unreliable. Indeed, when pricing OTM options, small samples of ITM paths near the beginning of the option life prohibit the computation of fitted value variances. For instance, we cannot compute the variance of a regression fit that uses a single observation. This is especially problematic because in such cases, the continuation value enjoys perfect foresight by simply interpolating the option value at the next step. Our solution to this problem is therefore to use all simulated paths in the regressions.

Furthermore, because we are using fewer paths, high-order polynomial regressions have a tendency to overfit future cash flows such that the convexity of the continuation value function is not satisfied. Although this is not particularly concerning for an in-sample bias-corrected estimator, it is worth mentioning that in an out-of-sample implementation of LSM, continuation

values are not fitted values, but *predictions* from the model in Equation (3.4). Since continuation value estimates can be highly unreliable when they are extrapolated from high-order polynomials, sub-optimality bias correction terms can be severely over-estimated and lead to equally severe estimator bias.

Another issue that comes from the regression method is that of correlated regressors. Indeed, standard OLS assumptions require regressors to be uncorrelated, or the consistency of the estimated variance estimates will be compromised. In the LSM algorithm, it is commonplace to use a set of monomials as basis functions. This typically does not affect the regression fit in-sample, and the effect of multicollinearity on the estimator bias is trivial in a standard LSM. However, multicollinearity does have a more subtle impact on the bias of a bias-corrected estimator, because correlated regressors can potentially cause the well-known variance inflation problem. Fortunately, numerical results (not reported here) show that monomials do not incur significant bias compared to orthogonal bases like Chebyshev, Hermite, Laguerre, and Lagrange polynomials.[4]

Finally, the HC variance-covariance matrix estimator $\hat{W}_j$ is a key component of the approximation of local bias. A common choice is the biased, but consistent White (1980) estimator, simply written $\text{diag}(\hat{\varepsilon}_{j+1}\hat{\varepsilon}'_{j+1})$. As L/N increases, one may consider more sophisticated HC estimators that are better-suited to (small) finite samples, like the unbiased HC estimator of Hinkley (1977) that corrects for the number of degrees of freedom. Other methods to enhance the finite sample properties of HC estimators include residual studentization (Horn et al. 1975), and bootstrapping of estimator variance, as the jackknife estimator of Efron (1982), for instance. See MacKinnon and White (1985) for a comprehensive discussion on the topic, and a comparison of these estimators. Computational efforts for more sophisticated estimators of $\hat{W}_j$ may not be worthwhile, however, and is left for future research.

Another way to carry out robust inference of OLS parameters in the presence of heteroskedasticity is to use a weighted least-squares approach. The weighted LSM of Fabozzi et al. (2017) employs variance-based regression weights to preserve error homoskedasticity. Although not presented, all derivations in this paper hold true in the context of a weighted least-squares approach. This may improve the accuracy of local bias approximations, and is also left for future work.

3.4 Numerical results

In this section, we report and compare the performances of the three in-sample estimators: the standard LSM (LSM) in Equation (3.8), the local bias-corrected LSM (FS-LSM) in Equation (3.25), which corrects for both foresight and sub-optimality bias as in Equation (3.27), and the foresight bias-corrected LSM (F-LSM), which corrects for foresight bias only as in Equation (3.26). We test estimators across a wide range of option characteristics, including option maturities of $T \in \{0.5, 1\}$ years, initial stock prices of $S_0 \in \{90, 100, 110\}$, and underlying asset volatilities of $\sigma \in \{0.1, 0.2, 0.4\}$. Throughout this article, we fix the strike price at $K = 100$, the continuous risk-free rate at $r = 0.05$, and the continuous dividend yield at $q = 0.1$.

Two sets of experiments are presented in this section to convey how the in-sample FS-LSM

[4] See Abramowitz and Stegun (1948) for more details on the efficient computation of orthogonal bases.

improves on the standard in-sample LSM for general diffusion processes and payoff structures. The first set of experiments measures estimator relative biases for a call option written on a single Geometric Brownian Motion (GBM) underlying asset. The second set of experiments examines how the bias-corrected estimators perform when we extend the valuation problem to pricing options written on multiple GBM underlying assets, or options written on underlying assets governed by a jump-diffusion process. Each set of experiments compares relative bias results for various sample sizes $N \in \{50, 100, 200, 400, 800, 1600, 3200, 6400, 12800\}$, polynomial regression orders $L \in \{1, \ldots, 4\}$ and frequencies of early-exercise opportunities $J/T \in \{5, 10, 15, \ldots, 50\}$. Unless specified otherwise, all studied options permit weekly exercise opportunities (i.e., $J/T = 50$) and use a cubic approximation of the early-exercise strategy (i.e., $L = 3$).

Denoting the number of underlying assets as D, a budget of $1,280,000 \times D \times J$ random numbers is used to obtain each price estimate. In other words, for a given sample size N, we use a sample of $M = 1,280,000/N$ independently repeated option valuations to compute four statistics of performance: average relative bias, standard error, root-mean squared error (RMSE), and foresight bias. The bias is computed as the difference between the average of M estimator valuations and the benchmark price. For our purposes, an adequate benchmark price should be free of foresight bias, but still include the same approximation bias. Therefore, the benchmark price used in this article is the average of 50 independent repetitions of an *out-of-sample* estimator with 10^6 simulated paths and using the same set of basis functions. Though negatively biased, this benchmark is sufficiently precise for the present exposition purposes. The relative bias is then simply computed as the ratio of the bias and the benchmark price. Moreover, we compute estimator standard errors as the sample standard deviation of M values divided by $\sqrt{M}$. Finally, RMSE is used to compute the efficiency of the FS-LSM algorithms, relative to the standard LSM estimator with the same number of paths and polynomial regression order. The RMSE efficiency is defined as

$$\text{Eff}(\hat{V}_0^*) = \frac{\text{RMSE}(\hat{V}_0)}{\text{RMSE}(\hat{V}_0^*)} - 1, \tag{3.29}$$

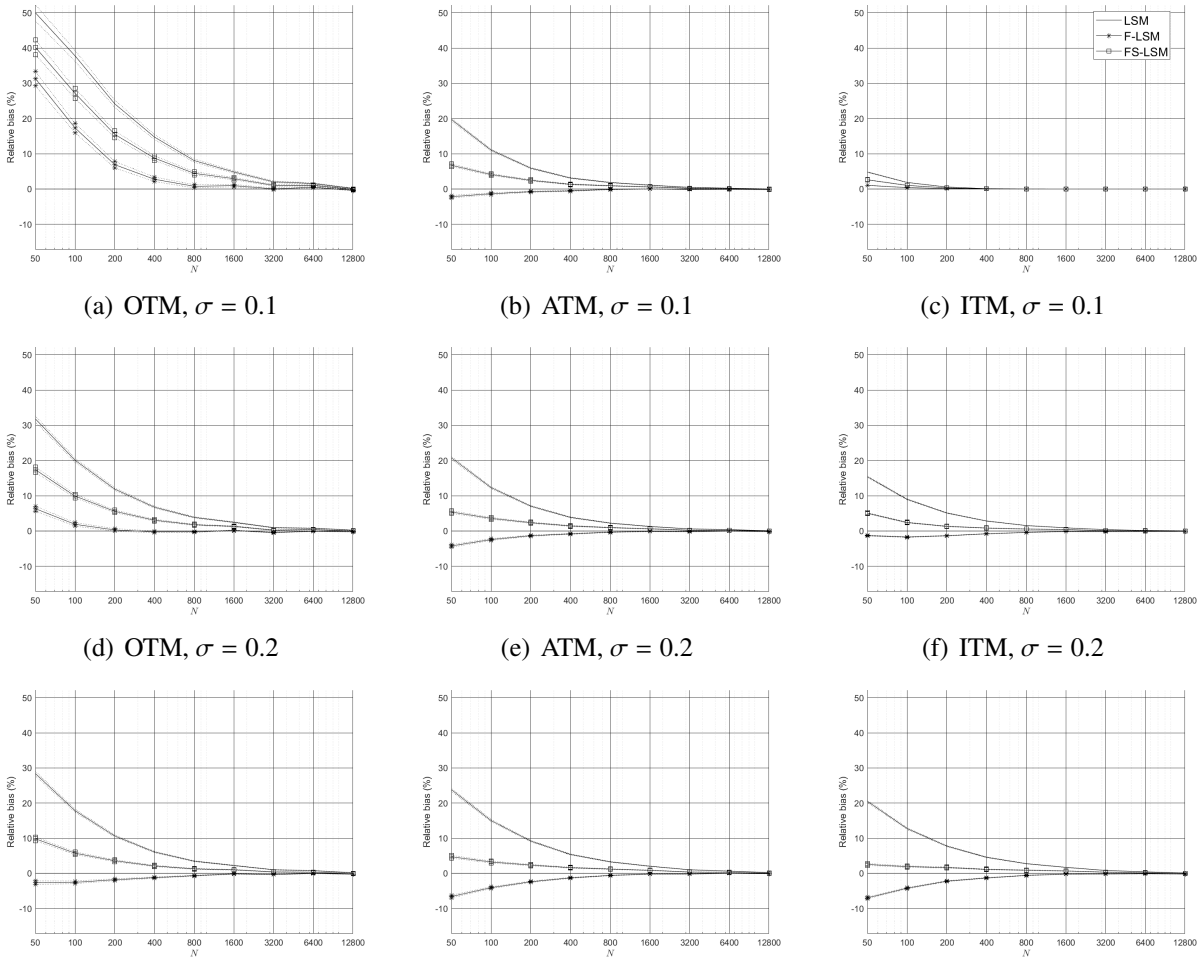
and a positive efficiency thus indicates that the bias-corrected estimator improves on the standard LSM estimator.

3.4.1 Bias correction results

The first set of experiments measures estimator relative biases for a call option written on a single GBM underlying asset across different levels of option moneyness $S_0/K \in \{0.9, 1, 1.1\}$, maturity $T \in \{0.5, 1\}$, and asset volatility $\sigma \in \{0.1, 0.2, 0.4\}$. Next, we fix the maturity to $T = 1$ and the volatility to $\sigma = 0.2$ to examine how relative bias changes as a function of N, J/T, and L. We find that the FS-LSM successfully corrects the relative bias across all levels of maturity, moneyness, and volatility. In addition, the FS-LSM relative bias is robust to the number of basis functions and, contrary to that of the standard LSM, decreases with the frequency of early-exercise opportunities.

In a first experiment, we show that the FS-LSM effectively reduces bias of option with weekly exercise opportunities (i.e., $J/T = 50$) across various asset volatilities and option maturities. Figures 3.4.1, 3.4.2, and 3.4.3 show relative biases of six-month, one-year, and two-year

options, respectively, as a function of the number of simulated paths for a single-asset American call option ($D = 1$). As expected, all three estimators converge to the benchmark value as N grows. Furthermore, we can see that F-LSM, and FS-LSM prices are both consistently below LSM prices. The F-LSM estimator is also consistently lower than the FS-LSM estimator because it does not correct for local sub-optimality bias. As expected, the magnitude of the bias decreases rapidly as the number of simulated paths increase. In addition observe that the relative bias of the LSM estimator depends on option moneyness S_0/K. When the number of sample paths is very low, the LSM has a larger positive bias for OTM options as seen in Panels 3.1(a), 3.1(d), 3.1(g) for six-month options, in particular. In such cases, the FS-LSM and F-LSM estimators are also positively biased, suggesting poor foresight bias approximations. In contrast, the bias of FS-LSM prices is quite small for ATM and ITM options, with a relative pricing error of less than 10% for sample sizes as small as $N = 50$, as presented in Table 3.C.1 in Appendix 3.D. To sum up, the FS-LSM and F-LSM estimators has smaller absolute bias than the LSM for moderately small sample sizes. For large sample sizes, say $N > 12800$, the bias of the corrected and uncorrected estimators tend to coincide, consistent with the convergence result presented in Section 3.3.3. Results for a one-year put option are reported in Figure 3.D.1 of Appendix 3.D to show that our conclusions hold irrespective of the option's payoff style.



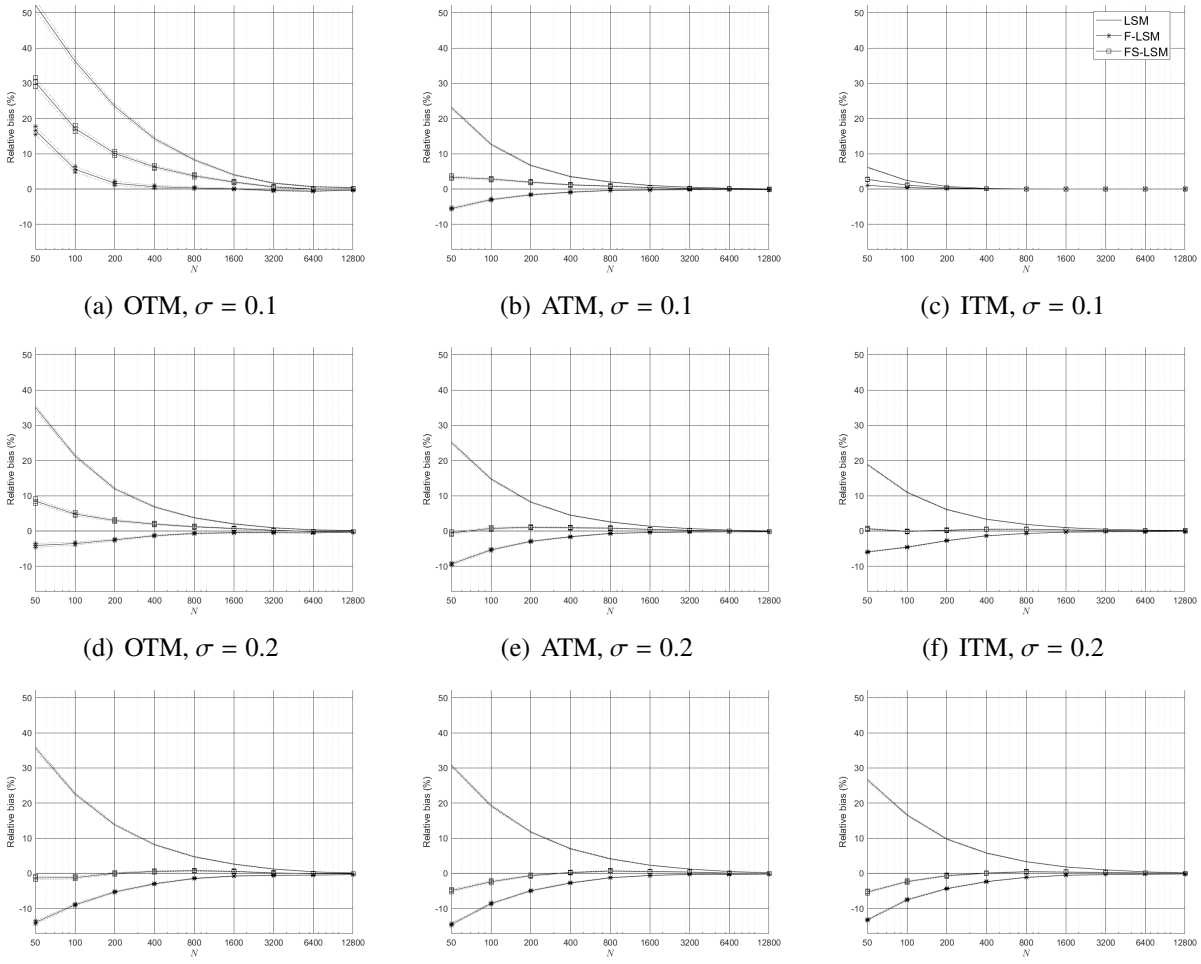
(a) OTM, $\sigma = 0.1$ (b) ATM, $\sigma = 0.1$ (c) ITM, $\sigma = 0.1$

(d) OTM, $\sigma = 0.2$ (e) ATM, $\sigma = 0.2$ (f) ITM, $\sigma = 0.2$

(g) OTM, $\sigma = 0.4$ (h) ATM, $\sigma = 0.4$ (i) ITM, $\sigma = 0.4$

Figure 3.4.1: Relative bias with GBM underlying asset, $T = 0.5$.

Bias as a percentage of the benchmark price for single-asset American call LSM, F-LSM, and FS-LSM price estimators. Results for 6-month OTM, ATM, and ITM options with weekly exercise opportunities (i.e. $J/T = 50$) are reported across asset volatilities $\sigma \in \{0.1, 0.2, 0.4\}$. The left column shows results for OTM options, the center column for ATM options, and the right column for ITM options. The top row shows results for $\sigma = 0.1$, the center row for $\sigma = 0.2$, and the bottom row for $\sigma = 0.4$. A cubic approximation of the exercise strategy (i.e. $L = 3$) is used to compute estimates and benchmark prices.



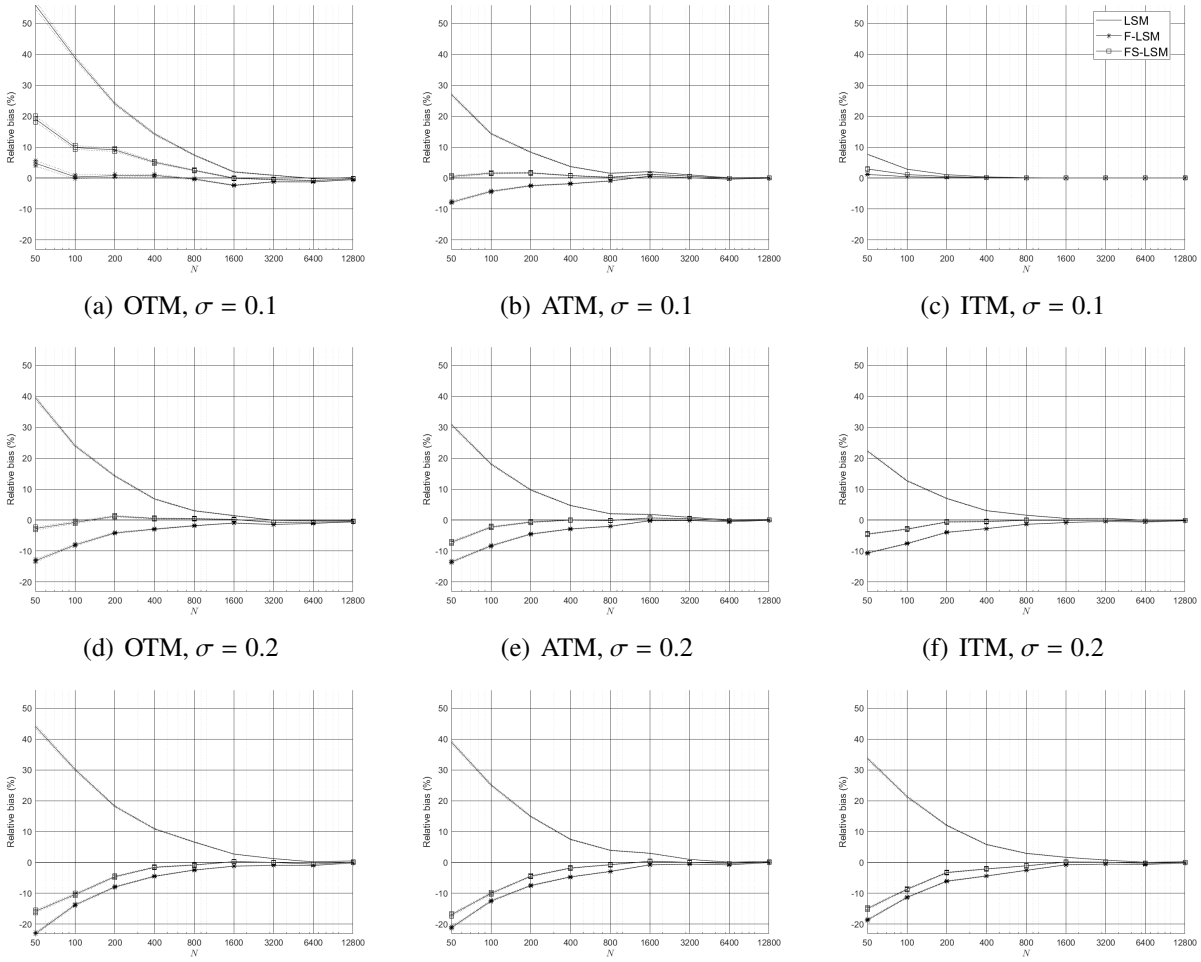
(a) OTM, $\sigma = 0.1$ (b) ATM, $\sigma = 0.1$ (c) ITM, $\sigma = 0.1$

(d) OTM, $\sigma = 0.2$ (e) ATM, $\sigma = 0.2$ (f) ITM, $\sigma = 0.2$

(g) OTM, $\sigma = 0.4$ (h) ATM, $\sigma = 0.4$ (i) ITM, $\sigma = 0.4$

Figure 3.4.2: Relative bias with GBM underlying asset, $T = 1$.

Bias as a percentage of the benchmark price for single-asset American call LSM, F-LSM, and FS-LSM price estimators. Results for 1-year OTM, ATM, and ITM options with weekly exercise opportunities (i.e. $J/T = 50$) are reported across asset volatilities $\sigma \in \{0.1, 0.2, 0.4\}$. The left column shows results for OTM options, the center column for ATM options, and the right column for ITM options. The top row shows results for $\sigma = 0.1$, the center row for $\sigma = 0.2$, and the bottom row for $\sigma = 0.4$. A cubic approximation of the exercise strategy (i.e. $L = 3$) is used to compute estimates and benchmark prices.



(a) OTM, $\sigma = 0.1$ (b) ATM, $\sigma = 0.1$ (c) ITM, $\sigma = 0.1$

(d) OTM, $\sigma = 0.2$ (e) ATM, $\sigma = 0.2$ (f) ITM, $\sigma = 0.2$

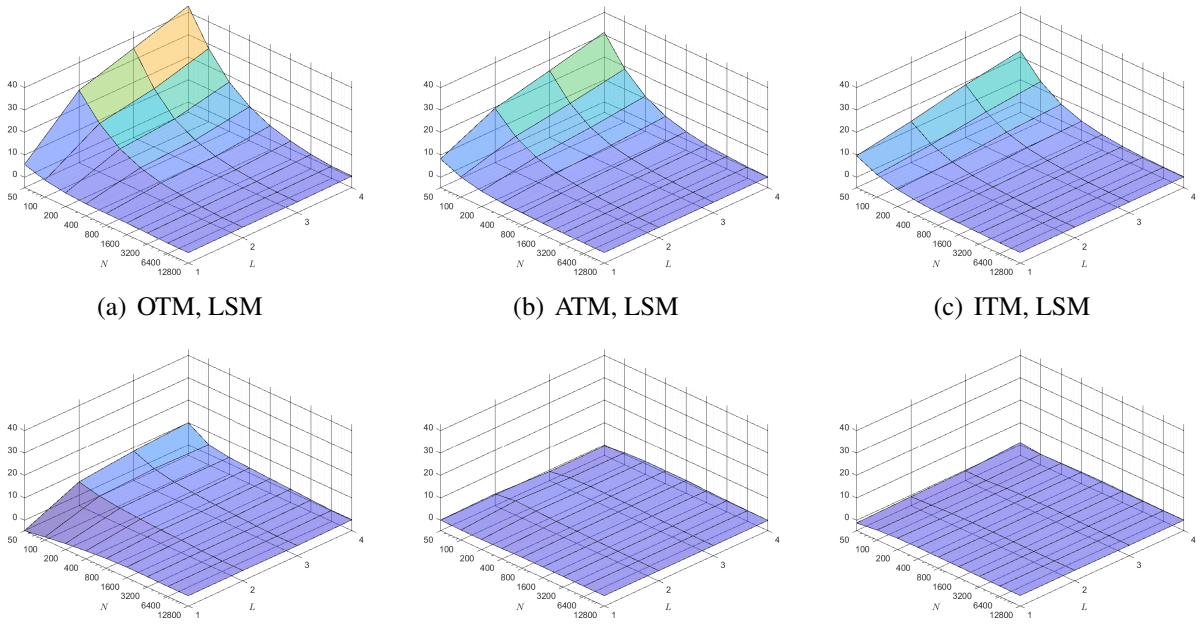
(g) OTM, $\sigma = 0.4$ (h) ATM, $\sigma = 0.4$ (i) ITM, $\sigma = 0.4$

Figure 3.4.3: Relative bias with GBM underlying asset, $T = 2$.

Bias as a percentage of the benchmark price for single-asset American call LSM, F-LSM, and FS-LSM price estimators. Results for 1-year OTM, ATM, and ITM options with weekly exercise opportunities (i.e. $J/T = 50$) are reported across asset volatilities $\sigma \in \{0.1, 0.2, 0.4\}$. The left column shows results for OTM options, the center column for ATM options, and the right column for ITM options. The top row shows results for $\sigma = 0.1$, the center row for $\sigma = 0.2$, and the bottom row for $\sigma = 0.4$. A cubic approximation of the exercise strategy (i.e. $L = 3$) is used to compute estimates and benchmark prices.

In the second experiment, we fix the asset volatility to $\sigma = 0.2$ and the option maturity to $T = 1$ and measure the effect of increasing the frequency of exercise opportunities J/T on LSM and FS-LSM estimators. Figure 3.4.5 illustrates the relative bias of LSM and FS-LSM estimators as a function of the sample size N and the frequency of exercise opportunities J/T. We observe that the LSM estimators are more positively biased as J/T increases, suggesting that continuation value models are more prone to overfitting when the frequency of exercise opportunities is high. Indeed, the more times foresight affects local exercise decisions, the more positive bias will arise. The FS-LSM exhibits an opposite tendency, where the positive bias tends toward zero as J/T increases. This suggests that local bias estimates are more accurate

when the time steps are short, and that the bias-corrections are more effective when they are frequent. Table 3.C.2 in Appendix 3.D presents results for LSM, F-LSM, and FS-LSM prices in this experiments. We observe that FS-LSM prices for OTM options with weekly exercise opportunities have less than 10% bias with as few as $N = 50$ paths, whereas the LSM relative pricing bias nears 35%. In comparison, using $N = 50$ paths to price an ATM option with weekly exercise opportunities yields relative biases of 26.6% for the LSM and 0.2% for the FS-LSM. For ITM options, the relative bias is 19.3% for the LSM, and -0.3% for the FS-LSM. Given that American options are approximated when a large number of exercise opportunities is considered, the FS-LSM permits a better approximation of the American option price because practitioners can increase J/T without being concerned about introducing bias. The bias coming from the "Bermudification" of American options is then more easily suppressed.



(a) OTM, LSM (b) ATM, LSM (c) ITM, LSM

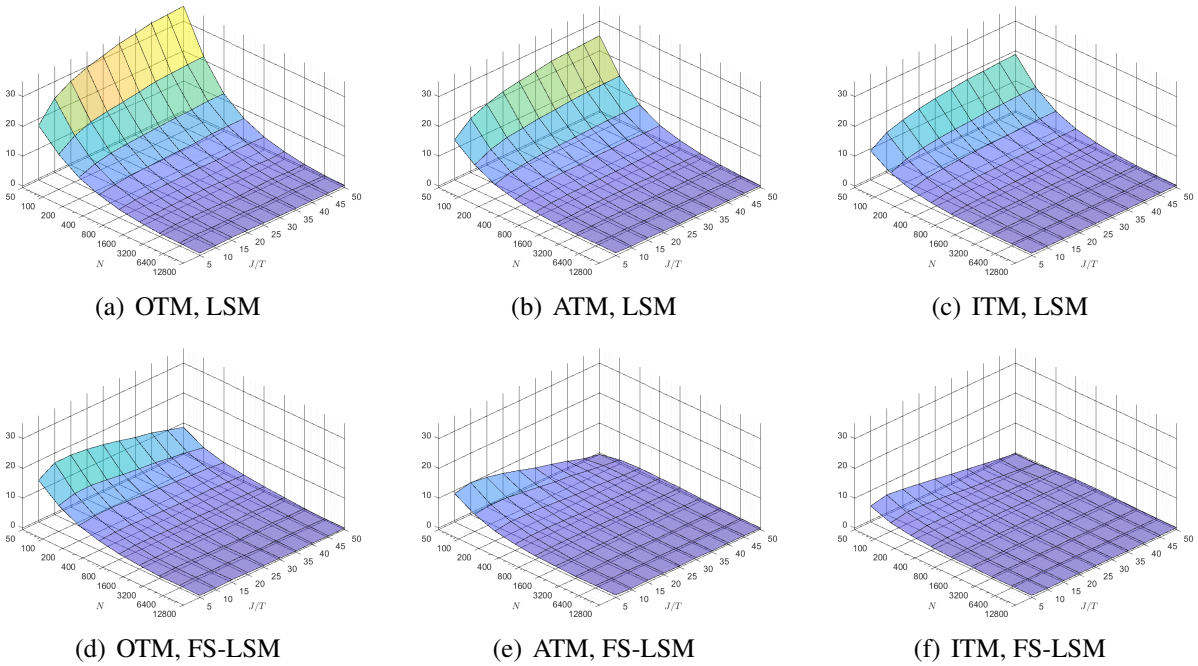
(d) OTM, FS-LSM (e) ATM, FS-LSM (f) ITM, FS-LSM

Figure 3.4.4: Relative bias of single-asset call for varying N and L.

Bias as a percentage of the benchmark price for single-asset American call LSM and FS-LSM price estimators. Results for one-year OTM, ATM, and ITM options with weekly exercise opportunities (i.e. $J/T = 50$), and asset price volatilities of $\sigma = 0.2$, are reported across the number of simulated paths $N \in \{50, 100, 200, 400, 800, 1600, 3200, 6400, 12800\}$ and polynomial order $L \in \{1, \ldots, 4\}$. The left column shows results for OTM options, the center column for ATM options, and the right column for ITM options. The top row shows results for the LSM estimator and the bottom row for the FS-LSM. For every option, the benchmark is computed with the corresponding polynomial order L.

Finally, we fix the asset volatility to $\sigma = 0.2$ and the option maturity to $T = 1$ and measure the effect of increasing the number of polynomials L. We can see in Figure 3.4.4 that the bias of the LSM estimator is very sensitive to the number of basis functions L, especially in small samples. Indeed, the relative bias of an OTM call LSM estimator goes above 40% when $N = 50$ and $L = 4$, whereas the relative bias of an FS-LSM estimator is about 10%. For ATM

and ITM options, the LSM exhibits large positive bias when the sample size is small, but the FS-LSM relative bias is always lower than 2% in absolute value. Detailed results are presented in Table 3.C.3 in Appendix 3.D along with corresponding F-LSM estimates.



(a) OTM, LSM (b) ATM, LSM (c) ITM, LSM

(d) OTM, FS-LSM (e) ATM, FS-LSM (f) ITM, FS-LSM

Figure 3.4.5: Relative bias of single-asset call for varying N and J/T.

Bias as a percentage of the benchmark price for single-asset American call LSM and FS-LSM price estimators. Results for one-year OTM, ATM, and ITM options with a cubic approximation of the early-exercise strategy ($L = 3$), and asset price volatilities of $\sigma = 0.2$, are reported across the number of simulated paths $N \in \{50, 100, 200, 400, 800, 1600, 3200, 6400, 12800\}$ and exercise opportunity frequency $J/T \in \{5, 10, 15, \ldots, 50\}$. The left column shows results for OTM options, the center column for ATM options, and the right column for ITM options. The top row shows results for the LSM estimator and the bottom row for the FS-LSM. For every option, the benchmark is computed with the corresponding exercise opportunity frequency J/T.

3.4.2 Robustness checks

The second set of experiments tests the robustness of the FS-LSM to the payoff structure and the underlying diffusion process. First, we compare the bias of LSM and FS-LSM estimators in the context of a American max-call option whose payoff is a function of $D = 5$ underlying assets $(S_t)_{t\geq 0} = (S_{1,t}, \ldots, S_{5,t})_{t\geq 0}$ (c.f. Broadie and Glasserman (2004), Whitehead et al. (2012), and Kan and Reesor (2012)). The general D-dimensional GBM is written as

$$\frac{dS_{d,t}}{S_{d,t}} = (r - q_d)dt + \sigma_d d\mathcal{W}_{d,t}, \tag{3.30}$$

where $d\mathcal{W}_{d,t} : d = 1, \ldots, D$ are (potentially correlated) Brownian increments. In our experiments, we assume that all asset prices are independent, i.e., $\mathbb{E}\left[d\mathcal{W}_{i,t}, d\mathcal{W}_{j,t}\right] = \mathbb{1}_{\{i=j\}}dt$, and

identically distributed, such that they share the same initial price S_0, volatility σ, and continuous dividend and risk-free rates, $q = 0.1$ and $r = 0.05$, respectively. Denoting the time-j sorted assets as $(S_{(1),j}, \ldots, S_{(D),j})$, where $S_{(d),j} \leq S_{(d+1),j}$ for $d = 1, \ldots, D-1$, we write the payoff of the max-call option $X_t = (S_{(D),t} - K)^+$, where $(x)^+$ is the positive part of x.

Second, to further illustrate the robustness of the FS-LSM, we compare estimator biases when the underlying asset is governed by the Jump Diffusion Model (JDM) of Merton (1976). The JDM generalizes GBM by allowing price discontinuities that take the form of multiplicative jumps. In the JDM, jumps are modelled as a compound counting process written as

$$J_t = \sum_{j=1}^{N_t} (Y_j - 1), \tag{3.31}$$

where jump sizes $Y_1, Y_2, \ldots$ are i.i.d. log-normal, and N_t is the Poisson counting process with constant intensity λ. The Poisson process, jump sizes and Brownian motion process are mutually independent. Note that the the GBM is a special case of the JDM model where $\lambda = 0$. This allows to measure the effect of jumps on foresight by comparing the foresight bias under the JDM with the GBM results in Section 3.4.1. Using the notation of Glasserman (2003), we can write the JDM diffusion under the risk-neutral measure as

$$\frac{dS_t}{S_t} = (r - q - \lambda m)dt + \sigma d\mathcal{W}_t + dJ_t, \tag{3.32}$$

where $m = \mathbb{E}[Y] - 1$.

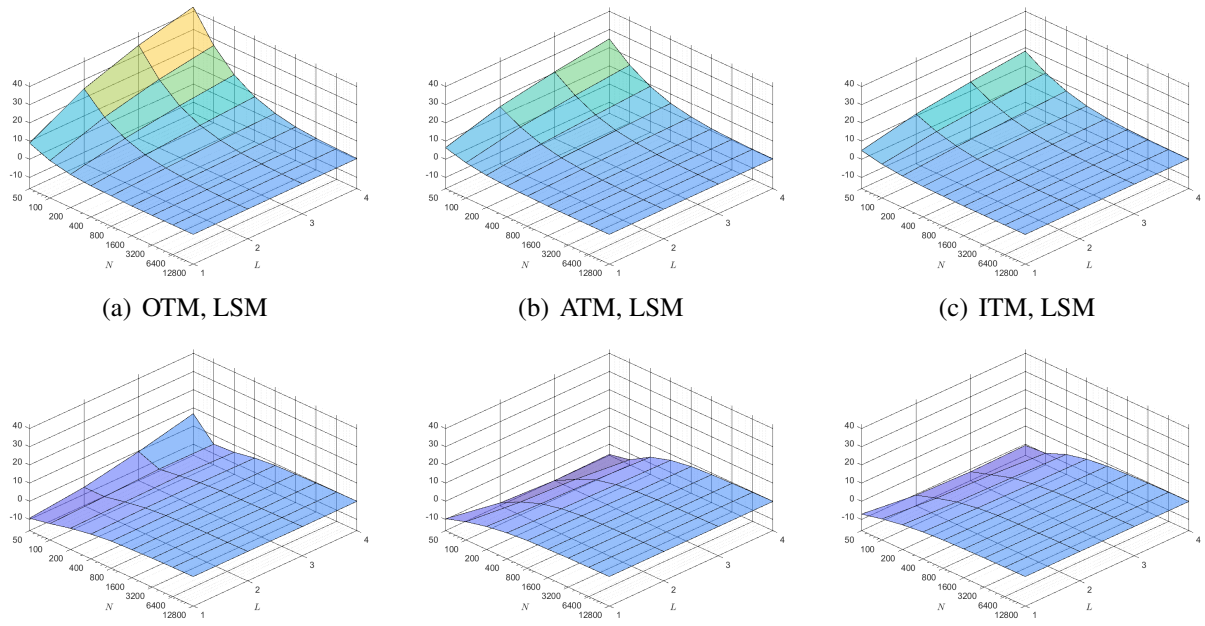
The JDM has several desirable properties for Monte Carlo applications. First, the JDM retains much of the tractability of the GBM, faciliating the discretization of the model in a Monte Carlo setting. Second, one can simulate JDM asset prices at fixed dates (i.e., $\tau_j : j = 0, \ldots, J$) without explicitly simulating exact jump times. Indeed, the simulation only requires one to simulate the number of jumps over a given time step. If the simulated Poisson process counts n jumps between two time points, a single normal variate can be used to simulate the compound process by observing that

$$\sum_{j=1}^{n} \log(Y_j) \sim \mathcal{N}\left(an, b^2 n\right), \tag{3.33}$$

because the JDM has log-normal jump sizes. Exponentiating simulated log-returns then yields JDM asset price paths (Glasserman 2003). In our experiment, all pure diffusion parameters r, q, σ remain the same as in the GBM, and we set jump parameters to $a = 0$, $b = 0.2$, meaning that each additive jump of log-returns is normally distributed about 0 with a standard deviation of 0.2. We report results for annual jump intensities $\lambda \in \{1, 10\}$, meaning that log-returns exhibit λ additive jump per year on average.

In Figure 3.4.6 are depicted average relative biases for a max-call option with $D = 5$ underlying assets, evolving according to Equation (3.30), as a function of L.[5] We can see that in such high-dimensional option pricing problems, the FS-LSM successfully reduces bias of

[5] In our implementation, a polynomial order of L across $D = 5$ assets implies that the set of basis functions is $\psi(S_j) = (\psi_0, \psi_\ell(S_{(d),j}) : \ell = 1, \ldots, L, d = 1, \ldots, D)$. In other words, regressors are made of up to four monomials in each of the D assets without cross-products, for a total of $(L \times D + 1)$ regression parameters.



(a) OTM, LSM (b) ATM, LSM (c) ITM, LSM

(d) OTM, FS-LSM (e) ATM, FS-LSM (f) ITM, FS-LSM

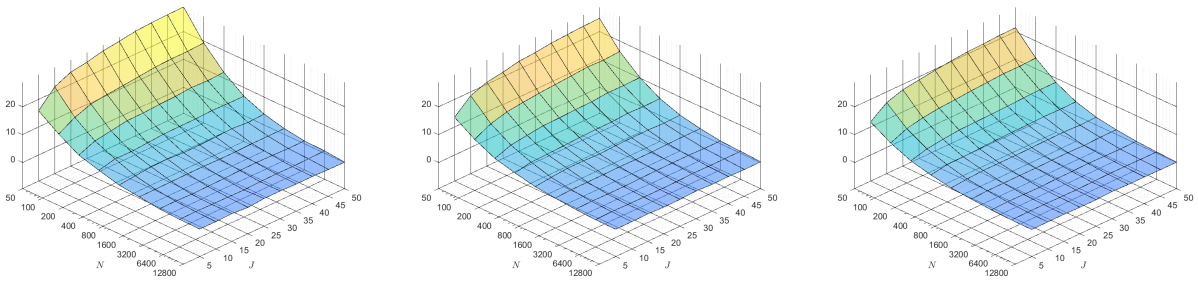
Figure 3.4.6: Relative bias of five-asset max-call for varying N and L.

Bias as a percentage of the benchmark price for five-asset American max-call LSM and FS-LSM price estimators. Results for one-year OTM, ATM, and ITM options with weekly exercise opportunities (i.e. $J/T = 50$), and asset price volatilities of $\sigma = 0.2$, are reported across the number of simulated paths $N \in \{50, 100, 200, 400, 800, 1600, 3200, 6400, 12800\}$ and polynomial order $L \in \{1, \ldots, 4\}$. The left column shows results for OTM options, the center column for ATM options, and the right column for ITM options. The top row shows results for the LSM estimator and the bottom row for the FS-LSM. For every option, the benchmark is computed with the corresponding polynomial order L.

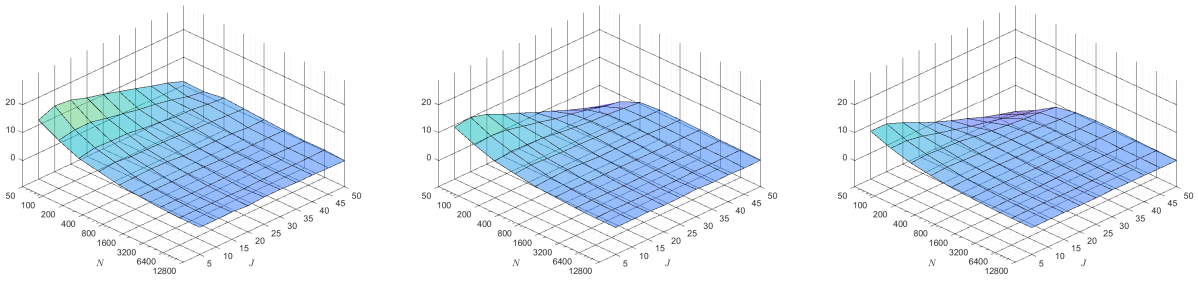
the standard approach across all sample sizes, polynomial orders, and levels of moneyness. As seen in Table 3.C.5 of Appendix 3.C, for sample sizes of about $N = 800$, the relative pricing error is less than 1% for all FS-LSM options. The relative importance of the improvements made through bias reduction is most noticeable when pricing OTM options with a large number of basis functions. Comparing Panels 3.6(a) and 3.6(d) shows that for very small sample sizes $N = 50$, the relative bias of the FS-LSM is about 5% in absolute value for all polynomial orders, while the LSM relative bias reaches 30% for $L = 4$. Indeed, for small samples sizes, the foresight bias of a standard LSM estimator increases quickly in the polynomial order L whereas the FS-LSM estimator bias is less sensitive to the number of basis functions. For larger samples, the LSM estimator bias is quite robust to L, and the advantages of FS-LSM are then subdued.[6]

Results under the five-dimensional GBM diffusion are interesting because they inform us on the effect of multiple stochastic elements on the reliability of local bias estimations. Indeed, the variance of continuation values are increasingly difficult to estimate as the number

[6]Figure 3.D.2 in Appendix 3.D shows that the negative bias of the bias-corrected estimators decreases when the number of exercise opportunities is decreased. Detailed results are reported in Table 3.C.4 of Appendix 3.C.



(a) OTM, LSM (b) ATM, LSM (c) ITM, LSM



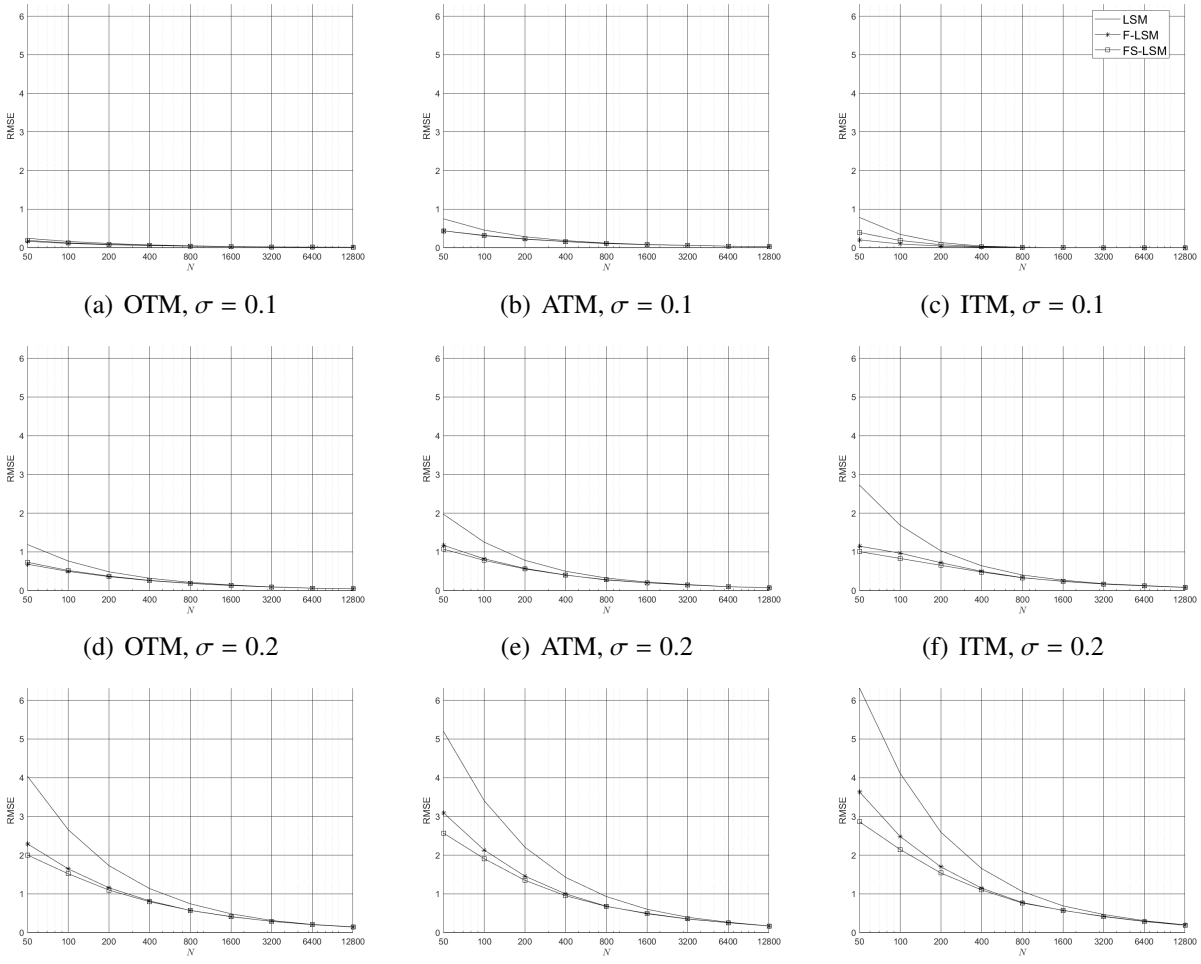
(d) OTM, FS-LSM (e) ATM, FS-LSM (f) ITM, FS-LSM

Figure 3.4.7: Relative bias of single-asset call with JDM underlying for varying N and J/T.
Bias as a percentage of the benchmark price for single-asset American call LSM and FS-LSM price estimators. Results for one-year OTM, ATM, and ITM options with a cubic approximation of the early-exercise strategy ($L = 3$), and asset price volatilities of $\sigma = 0.2$, are reported across the number of simulated paths $N \in \{50, 100, 200, 400, 800, 1600, 3200, 6400, 12800\}$ and exercise opportunity frequency $J/T \in \{5, 10, 15, \ldots, 50\}$. The compound Poisson process in parameterized by $\lambda = 1, a = 0, b = 0.2$. The left column shows results for OTM options, the center column for ATM options, and the right column for ITM options. The top row shows results for the LSM estimator and the bottom row for the FS-LSM. For every option, the benchmark is computed with the corresponding exercise opportunity frequency J/T.

of regressors increase, hampering the consistency of local bias estimates. The results in Figure 3.4.6 further suggest that local bias corrections effectively deal with foresight for general option characteristics. Contrary to the standard LSM approach, FS-LSM prices are more robust to the choice of polynomial order L. This means that an FS-LSM estimator with a large number of basis functions requires fewer sample paths than LSM to obtain the same bias. Reducing the impact of foresight bias facilitates the pricing of American options that require a large number of basis functions, like basket options, or other derivatives based on multiple correlated stochastic factors. Indeed, the number of basis functions grows prohibitively fast when approximating multivariate polynomials with cross-products.

Results under the JDM are interesting because they inform us on the effect of price discontinuities on the relative pricing bias. Figure 3.4.7 graphs relative bias results for one-year options as a function of the number simulated paths and of the frequency of exercise opportunities. Corresponding detailed results are presented in Table 3.C.6 in Appendix 3.C. These results show that the FS-LSM again has smaller bias than the LSM in all cases considered. FS-LSM option prices are roughly below 10% for sample sizes as small as $N = 50$, and have

less than 1% bias for $N \geq 800$.[7] Unreported results for a JDM underlying process with $\lambda = 10$ indicates that the performance of local bias corrections is robust to even a very high jump intensity. We leave the thorough examination of robustness to more general classes of underlying diffusion processes like stochastic volatility models to further research.



(a) OTM, $\sigma = 0.1$ (b) ATM, $\sigma = 0.1$ (c) ITM, $\sigma = 0.1$

(d) OTM, $\sigma = 0.2$ (e) ATM, $\sigma = 0.2$ (f) ITM, $\sigma = 0.2$

(g) OTM, $\sigma = 0.4$ (h) ATM, $\sigma = 0.4$ (i) ITM, $\sigma = 0.4$

Figure 3.4.8: RMSE with GBM underlying, $T = 1$.

RMSE for single-asset American call LSM, F-LSM, and FS-LSM price estimators. Results for 1-year OTM, ATM, and ITM options with weekly exercise opportunities (i.e. $J/T = 50$) are reported across asset volatilities $\sigma \in \{0.1, 0.2, 0.4\}$. The left column shows results for OTM options, the center column for ATM options, and the right column for ITM options. The top row shows results for $\sigma = 0.1$, the center row for $\sigma = 0.2$, and the bottom row for $\sigma = 0.4$. A cubic approximation of the exercise strategy (i.e. $L = 3$) is used to compute estimators, and benchmark prices.

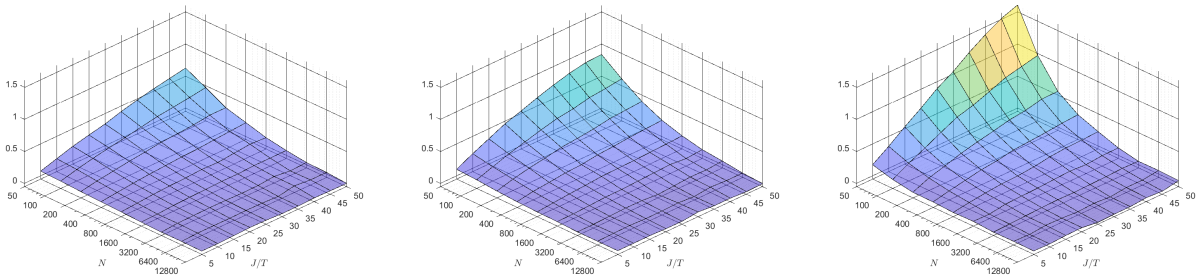
[7] Figure 3.D.3 in Appendix 3.D shows that the negative bias of the bias-corrected estimators decreases when the number of exercise opportunities is decreased.

3.4.3 Discussion

Our proposed FS-LSM estimator reduces the bias significantly even with a small number of paths and eliminates it completely when the number of paths is moderately large. Moreover, the implementation of local bias corrections also reduces the variance of the estimator. We explain this by emphasizing that the bias corrections not only reduce the bias locally, but also enhance all future regressions in the algorithm. Indeed, when regressions are performed on a bias-corrected cross-section of option values, the fitted values that determine exercise decisions are themselves less biased. Local bias corrections therefore reduce the variance stemming from random exercise errors.

The net effect of reducing both bias and standard error is, of course, a reduced RMSE. Figure 3.4.8 graphs estimator RMSE versus sample size for one-year single-asset options with GBM underlying, showing that corrected estimators are most efficient in terms of RMSE when the number of simulation paths is low, and when the level of moneyness is low. Indeed, it is in such cases that LSM estimators are most prone to foresight bias, so it stands to reason that it is in such cases that the corrected estimators are most relevant.

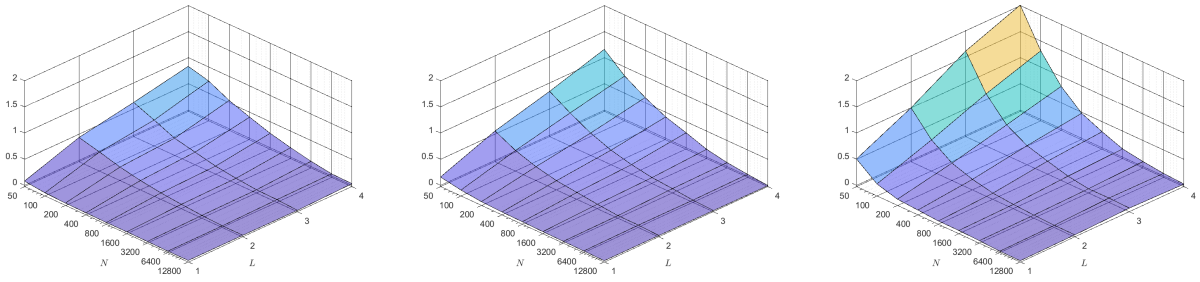
Finally, we see in Figures 3.4.9 and 3.4.10 that the frequency of exercise opportunities and the number of polynomials have an important impact on the RMSE efficiency of the FS-LSM. For OTM and ATM options, the bias corrections are most effective when J/T and L are high and when N is low. For ITM options, the RMSE efficiency decreases when J/T is greater than 25, due to increasing estimator volatility.



(a) OTM (b) ATM (c) ITM

Figure 3.4.9: RMSE efficiency of FS-LSM single-asset call prices for varying N and J/T.

RMSE efficiency for single-asset American call FS-LSM price estimators. Results for one-year OTM, ATM, and ITM options with a cubic approximation of the early-exercise strategy ($L = 3$), and asset price volatilities of $\sigma = 0.2$, are reported across the number of simulated paths $N \in \{50, 100, 200, 400, 800, 1600, 3200, 6400, 12800\}$ and exercise opportunity frequency $J/T \in \{5, 10, 15, \ldots, 50\}$. The left column shows results for OTM options, the center column for ATM options, and the right column for ITM options. The top row shows results for the F-LSM estimator and the bottom row for the FS-LSM. For every option, the benchmark is computed with the corresponding exercise opportunity frequency J/T.



(a) OTM (b) ATM (c) ITM

Figure 3.4.10: RMSE efficiency of FS-LSM single-asset call prices for varying N and L.

RMSE efficiency for single-asset American call FS-LSM price estimators. Results for one-year OTM, ATM, and ITM options with weekly exercise opportunities (i.e. $J/T = 50$), and asset price volatilities of $\sigma = 0.2$, are reported across the number of simulated paths $N \in \{50, 100, 200, 400, 800, 1600, 3200, 6400, 12800\}$ and polynomial order $L \in \{1, \ldots, 4\}$. The left column shows results for OTM options, the center column for ATM options, and the right column for ITM options. The top row shows results for the F-LSM estimator, and the bottom row for the FS-LSM. For every option, the benchmark is computed with the corresponding polynomial order L.

3.5 Conclusion

This article provides an in-depth discussion of the foresight bias that occurs when the early-exercise strategy depends on future option values in a regression based simulation method for American option pricing. We first derive a theoretical expression for the foresight bias that holds true for general diffusion processes and payoff structures. We then show that a normal approximation of continuation values can lead to a tractable expression of local foresight bias. Local foresight bias corrections, complemented by the local sub-optimality bias corrections presented in Kan et al. (2009), then allow to efficiently correct the bias of an in-sample LSM estimator.

For single asset options, we find that the FS-LSM estimator has a lower RMSE than the standard LSM. Not only do local corrections reduce bias, they also reduce the variance of the estimator. The overall performance of the FS-LSM is most noticeable for ATM and ITM options with small sample sizes. For OTM options, the estimation of the parameters used to compute local bias terms is hampered because the number of observations that contribute to the regression fit is too small to obtain a reliable regression variance-covariance matrix.

The FS-LSM also reduces the bias of the LSM estimator when considering options with complex payoffs that depend on multiple underlying assets. However, the performance of the FS-LSM is limited when there are too many regressors relative to the number of simulated paths. Indeed, although FS-LSM prices are more robust to the polynomial regression order, their variance increases, leading to a larger RMSE.

The foresight bias approximation introduced in this article reveals the root causes of foresight bias, namely the variance of continuation values, and the distance between the continuation value and the current payoff. Since a low-variance continuation value model reduces foresight bias, one may be interested in applying variance reduction techniques to the continuation values themselves, using the techniques of Rasmussen (2005) for instance. One may also

consider using importance sampling tools as in Boire et al. (2021*a*) to generate more balanced samples. Another interesting future research direction is to explore different assumptions about the continuation value distribution and the corresponding local bias estimators.

Appendix

3.A Bias decomposition

In the following, we provide rough estimates of the relative importance of the three bias components: approximation bias, sub-optimality bias, and foresight bias. To do so, we need estimators that are free from foresight bias. We therefore consider an alternative implementation of the LSM that uses an independent sample paths to estimate continuation values. This method, known as an *out-of-sample* implementation, removes foresight bias, but incurs a more hefty simulation task because the dynamic program needs to be carried out twice: once to estimate regression parameters, and once more to estimate stopping times and option values. The standard LSM is then called an *in-sample* LSM implementation of the LSM because the same set of paths is used both to estimate continuation values and to compute the option cash flows. In this algorithm, foresight bias is large when the number of paths used in the regressions is small compared to the number of regressors.

We numerically estimate approximation bias as the difference between the true option price approximated from a (Cox et al. 1979) binomial tree with a large number of time steps and an out-of-sample LSM with a large number of simulated paths. By using a large number of paths in an out-of-sample implementation, the latter estimator is designed to remove the foresight bias and most of the sub-optimality bias, therefore retaining only approximation bias. Sub-optimality bias is then estimated as the difference between the out-of-sample LSM with a large number of paths and a similar estimator with a modest number of paths. Using a modest number of paths in an out-of-sample representation prevents foresight bias, but allows sub-optimality bias. Finally, foresight bias is approximated as the difference between out-of-sample and in-sample LSM estimators with a modest number of simulated paths. The in-sample LSM estimator with a modest number of paths thus contains all three sources of bias.

Note that these are rough approximations of the three bias components. First, the out-of-sample LSM with a large number of simulated paths does not remove all sub-optimality bias, as this can only be done by using an infinite sample. Moreover, the out-of-sample estimator suffers from foresight in the estimation phase of the algorithm, and therefore exhibits more sub-optimality bias in the pricing phase than an estimator whose strategy is completely independent from future option values. Indeed, the out-of-sample estimator takes the form $v(\hat{\gamma}_j(\tilde{X}_j, \tilde{X}_{j+1}, \ldots, \tilde{X}_J), X_j)$, and it follows that

$$\mathbb{E}\left[v(\hat{\gamma}_j(\tilde{X}_j, \tilde{X}_{j+1}, \ldots, \tilde{X}_J), X_j)\right] \leq \mathbb{E}\left[v(\hat{\gamma}_j(\tilde{X}_j), X_j)\right]. \tag{3.34}$$

As a result, the sizes of sub-optimality and foresight bias estimates may be slightly over-estimated in the results reported in this section.

Table 3.A.1: Proportions of bias components as a function of L.

			N=50		N=1600		N=12800	
S_0	L	**Appr.**	**SubO.**	**For.**	**SubO.**	**For.**	**SubO.**	**For.**
90	1	-10.8704	-3.1323	8.2125	0.2309	0.2479	-0.0764	0.0069
90	2	-1.2491	-4.9723	32.2374	0.3301	1.4342	-0.0593	0.1541
90	3	-1.7877	-6.8896	41.3786	-0.0762	2.0515	-0.0895	0.2569
90	4	-0.6773	-11.2333	53.7376	-0.4742	4.0071	-0.0839	0.6013
100	1	-2.9413	-1.7071	9.6235	0.0941	0.3452	-0.0878	0.0584
100	2	-0.7659	-3.6314	23.0483	-0.1034	1.1624	-0.1780	0.1726
100	3	-0.4016	-6.3328	31.3901	-0.4179	1.7620	-0.2094	0.2358
100	4	-0.3267	-7.2512	38.2504	-0.7600	2.6105	-0.2793	0.3933
110	1	-1.0986	-2.1228	11.5901	0.0308	0.5409	-0.0966	0.1281
110	2	-0.2878	-4.7955	18.5024	-0.1785	0.8355	-0.0239	0.1639
110	3	-0.0785	-6.4435	25.3158	-0.4440	1.4022	-0.0428	0.1978
110	4	-0.0859	-7.3056	30.0972	-0.4688	1.6200	-0.0680	0.2445

This table shows approximation, sub-optimality, and foresight bias components as a function of the order of the polynomial approximation L for single-asset call options with strike price $K = 100$, maturity $T = 1$, and number of admissible exercise times $J = 50$. We set GBM diffusion parameters to $r = 0.05$, $q = 0.1$, and $\sigma = 0.2$. Results are expressed as a percentage of the price given by a Cox et al. (1979) binomial tree with 25,200 time steps and J exercise opportunities. Results are reported across number of paths $N \in \{50, 1600, 12800\}$, initial asset price $S_0 \in \{90, 100, 110\}$ and polynomial order $L \in \{1, 2, 3, 4\}$, and calculated from $M = 1280000/N$ independent replications.

Assuming that the equality holds in Equation (3.34), the relative size of each bias component for varying polynomial orders L and for varying frequencies of admissible early-exercise times are reported in Tables 3.A.1 and 3.A.2, respectively. From these tables it is apparent that as the polynomial regression order L increases, the size of approximation bias decreases, and that of sub-optimality and foresight biases increases, illustrating the difficulties of dealing with bias in the LSM. Indeed, the cost of reducing approximation bias is an unstable continuation value model that is prone to overfitting. Moreover, as the frequency of exercise times increases, the size of approximation and foresight bias components increases, while the sub-optimality bias appears to remain the same. Across all the configurations, though, the tables show that foresight bias is an important part of the total bias and often much more important than the other biases.

Table 3.A.2: Proportions of bias components as a function of J/T.

			N=50		N=1600		N=12800	
S_0	J/T	**Appr.**	**SubO.**	**For.**	**SubO.**	**For.**	**SubO.**	**For.**
90	5	-0.2714	-5.0073	22.6773	-0.3169	1.4776	-0.0670	0.1868
90	10	-0.5821	-5.9176	30.1095	-0.3735	1.7681	-0.1293	0.2874
90	25	-1.2009	-6.5213	37.1768	-0.0865	2.0307	-0.1138	0.3086
90	50	-1.7877	-6.8896	41.3786	-0.0762	2.0515	-0.0895	0.2569
100	5	-0.0763	-4.2942	17.1809	-0.1787	0.9490	0.1431	0.1149
100	10	-0.1421	-4.9356	22.4781	-0.4374	1.2434	-0.2061	0.1824
100	25	-0.2783	-5.5391	27.9103	-0.2533	1.5133	-0.1915	0.2022
100	50	-0.4016	-6.3328	31.3901	-0.4179	1.7620	-0.2094	0.2358
110	5	-0.0350	-4.2432	13.4277	-0.1003	0.5408	0.0610	0.0514
110	10	-0.0435	-4.6376	17.4864	-0.3249	0.9101	-0.1047	0.0987
110	25	-0.0613	-5.7294	22.3926	-0.3049	1.2090	-0.1256	0.1596
110	50	-0.0785	-6.4435	25.3158	-0.4440	1.4022	-0.0428	0.1978

This table shows approximation, sub-optimality, and foresight bias components as a function of the frequency of early-exercise opportunities J/T for single-asset call options with strike price $K = 100$, maturity $T = 1$, and using a polynomial regression of order $L = 3$. We set GBM diffusion parameters to $r = 0.05$, $q = 0.1$, and $\sigma = 0.2$. Results are expressed as a percentage of the price given by a Cox et al. (1979) binomial tree with 25,200 time steps and J exercise opportunities. Results are reported across number of paths $N \in \{50, 1600, 12800\}$, initial asset price $S_0 \in \{90, 100, 110\}$ and exercise frequency $J/T \in \{5, 10, 25, 50\}$, and calculated from $M = 1280000/N$ independent replications.

3.B Derivations

In the following are presented the derivations and relevant references to derive the local bias expressions used to correct the bias in the LSM algorithm. Under the assumption that continuation values are normally distributed, the local sub-optimality bias approximation is

$$\hat{\zeta}_j(\hat{C}_{j+1}, X_{j+1}, \sigma_{j+1}) \approx \left|\hat{C}_{j+1} - X_{j+1}\right| \Phi\left(\frac{-\left|\hat{C}_{j+1} - X_{j+1}\right|}{\sigma_{j+1}}\right) - \sigma_{j+1}\phi\left(\frac{\hat{C}_{j+1} - X_{j+1}}{\sigma_{j+1}}\right), \tag{3.35}$$

as presented in detail in Section 2 of Kan et al. (2009).

The expression of local foresight in Equation (3.15) can be approximated in the same heuristic fashion. To obtain the foresight bias approximation, it is useful to first observe that the time-j continuation value given that the option is held at time j is a random variable $\hat{C}_j^{\mathcal{H}} = (\hat{C}_j \mid \hat{C}_j > X_j)$ that belongs to the family of truncated normal random variables. We can therefore write

$$\hat{C}_j^{\mathcal{H}} \sim \mathcal{N}_{X_j}^{\infty}\left(\bar{C}_j, \sigma_j\right) \tag{3.36}$$

where $\mathcal{N}_a^b(\mu, \sigma)$ denotes a normal random variable with mean $\mu \in \mathbb{R}$ and standard deviation $\sigma > 0$ whose domain is truncated and constrained to be the interval $(a, b) : -\infty \leq a < b \leq \infty$. The density of a random variable $Y \sim \mathcal{N}_a^b(\mu, \sigma)$ takes the form

$$\begin{aligned} f_Y(y) &= \frac{1}{\sigma Z(a, b, \mu, \sigma)}\phi\left(\frac{y - \mu}{\sigma}\right) \\ Z(a, b, \mu, \sigma) &= \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right), \end{aligned} \tag{3.37}$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ are the cumulative distribution function (CDF) and the probability density function (PDF) of a standard normal random variable, respectively. Moreover, the mean of variable Y is written as

$$\mathbb{E}[Y] = \mu + \frac{\sigma}{Z(a, b, \mu, \sigma)}\left(\phi\left(\frac{a - \mu}{\sigma}\right) - \phi\left(\frac{b - \mu}{\sigma}\right)\right).$$

With $\mu = \bar{C}_j, \sigma = \sigma_j, a = X_j, b = \infty$, it immediately follows that

$$\mathbb{E}_j\left[\hat{C}_j^{\mathcal{H}}\right] = \bar{C}_j + \frac{\sigma_j}{\Phi\left(\frac{\bar{C}_j - X_j}{\sigma_j}\right)}\phi\left(\frac{X_j - \bar{C}_j}{\sigma_j}\right).$$

Moreover, since $\hat{C}_j$ converges to $\mathbb{E}_j\left[\hat{V}_{j+1}\right]$ almost everywhere, we have that $\mathbb{E}_{j+1}\left[\hat{C}_{j+1}^{\mathcal{H}}\right] = \mathbb{E}_{j+1}\left[\hat{V}_{j+2} \mid \mathcal{H}_{j+1}\right]$. Next, observe that $\mathbb{E}_{j+1}\left[\mathbb{1}_{\{\mathcal{H}_{j+1}\}}\hat{V}_{j+2}\right] = \mathbb{E}_{j+1}\left[\hat{C}_{j+1}^{\mathcal{H}}\right] \times \mathbb{P}_{j+1}\left[\mathcal{H}_{j+1}\right] = \mathbb{E}_{j+1}\left[\hat{C}_{j+1}^{\mathcal{H}}\right] \times \Phi\left(\frac{\bar{C}_j - X_j}{\sigma_j}\right)$ to write the local foresight bias term as

$$\begin{aligned} \xi_j &= \mathbb{E}_j\left[\Phi\left(\frac{\bar{C}_j - X_j}{\sigma_j}\right) \times \left(\mathbb{E}_{j+1}\left[\hat{C}_{j+1}^{\mathcal{H}}\right] - \bar{C}_{j+1}\right)\right] \\ &= \mathbb{E}_j\left[\sigma_{j+1}\phi\left(\frac{X_{j+1} - \bar{C}_{j+1}}{\sigma_{j+1}}\right)\right]. \end{aligned} \tag{3.38}$$

The expression in Equation (3.38) can be written in integral form to derive an approximation, and construct an estimator. In the derivations below, we start by approximating the joint density of $(\bar{C}_{j+1}, \sigma_{j+1} \mid \mathcal{F}_j)$ with the joint density of $(\hat{C}_{j+1}, \sigma_{j+1} \mid \mathcal{F}_j)$. Indeed, as the sample size increases $f_{\hat{C}_{j+1},\sigma_{j+1}|\mathcal{F}_j}(\hat{c}, \sigma)$ converges to $f_{\bar{C}_{j+1},\sigma_{j+1}|\mathcal{F}_j}(\bar{c}, \sigma)$ almost surely. By virtue of Fubini's theorem, we can then change the order of integration between $\hat{c}$ and $\bar{c}$. It is then possible to integrate out the unobserved quantity $\bar{C}_{j+1}$ to obtain the approximation.

$$
\begin{aligned}
&\int_0^\infty \int_{\mathbb{R}} \int_{\mathbb{R}} \sigma\phi\left(\frac{X_{j+1}-\bar{c}}{\sigma}\right) \times f_{\hat{C}_{j+1}|\bar{c},\sigma}(\hat{c}) \times f_{\bar{C}_{j+1},\sigma_{j+1}|\mathcal{F}_j}(\bar{c},\sigma) d\hat{c}d\bar{c}d\sigma \\
\approx &\int_0^\infty \int_{\mathbb{R}} \int_{\mathbb{R}} \sigma\phi\left(\frac{X_{j+1}-\bar{c}}{\sigma}\right) \times \frac{1}{\sigma}\phi\left(\frac{\hat{c}-\bar{c}}{\sigma}\right) \times f_{\hat{C}_{j+1},\sigma_{j+1}|\mathcal{F}_j}(\bar{c},\sigma) d\hat{c}d\bar{c}d\sigma \\
= &\int_0^\infty \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \phi\left(\frac{X_{j+1}-\bar{c}}{\sigma}\right)\phi\left(\frac{\hat{c}-\bar{c}}{\sigma}\right) d\bar{c}\right) \times f_{\hat{C}_{j+1},\sigma_{j+1}|\mathcal{F}_j}(\hat{c},\sigma) d\hat{c}d\sigma \\
= &\int_0^\infty \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \phi\left(\frac{\bar{c}-\frac{X+\hat{c}}{2}}{\sigma/\sqrt{2}}\right) d\bar{c}\right)\phi\left(\frac{\hat{c}-X_{j+1}}{\sigma\sqrt{2}}\right) \times f_{\hat{C}_{j+1},\sigma_{j+1}|\mathcal{F}_j}(\hat{c},\sigma) d\hat{c}d\sigma \\
= &\int_0^\infty \int_{\mathbb{R}} \frac{\sigma}{\sqrt{2}}\phi\left(\frac{\hat{c}-X_{j+1}}{\sigma\sqrt{2}}\right) \times f_{\hat{C}_{j+1},\sigma_{j+1}|\mathcal{F}_j}(\hat{c},\sigma) d\hat{c}d\sigma.
\end{aligned}
\tag{3.39}
$$

In expectation form, we simply write

$$
\xi_j \approx \mathbb{E}_j\left[\frac{\sigma_{j+1}}{\sqrt{2}}\phi\left(\frac{\hat{C}_{j+1}-X_{j+1}}{\sigma_{j+1}\sqrt{2}}\right)\right]. \tag{3.40}
$$

The sum of sub-optimality and foresight biases yields the local bias approximation in Equation (3.27) when continuation values are normal, which can be written as

$$
\begin{aligned}
&B_j(\hat{C}_{j+1}, X_{j+1}, \sigma_{j+1}) = \zeta_j(C_{j+1}, X_{j+1}, \sigma_{j+1}) + \xi_j(\hat{C}_{j+1}, X_{j+1}, \hat{\sigma}_{j+1}) \\
&= \mathbb{E}_j\Bigg[\underbrace{\left|\hat{C}_{j+1}-X_{j+1}\right|\Phi\left(\frac{-\left|\hat{C}_{j+1}-X_{j+1}\right|}{\sigma_{j+1}}\right) - \sigma_{j+1}\phi\left(\frac{\hat{C}_{j+1}-X_{j+1}}{\sigma_{j+1}}\right)}_{\text{Sub-optimality bias}} \\
&\quad + \underbrace{\frac{\sigma_{j+1}}{\sqrt{2}}\phi\left(\frac{\hat{C}_{j+1}-X_{j+1}}{\sigma_{j+1}\sqrt{2}}\right)}_{\text{Foresight bias}}\Bigg].
\end{aligned}
\tag{3.41}
$$

We can then construct the local bias estimators in Equations (3.20) and (3.21) by substituting σ_{j+1} by $\hat{\sigma}_{j+1}$ in Equations (3.35) and (3.40), respectively.

3.C Additional tables

Table 3.C.1: Detailed results for LSM, F-LSM, and FS-LSM prices with GBM underlying asset.

			LSM			F-LSM			FS-LSM		
T	S_0	σ	**Bias**	**S.E.**	**RMSE**	**Bias**	**S.E.**	**RMSE**	**Bias**	**S.E.**	**RMSE**
$N = 50, M = 25600$											
0.5	90	0.1	50.0	0.10	0.11	31.3	0.09	0.09	40.2	0.09	0.10
0.5	90	0.2	31.8	0.53	0.67	6.3	0.46	0.46	17.4	0.48	0.53
0.5	90	0.4	28.4	1.72	2.32	-2.7	1.42	1.43	9.8	1.47	1.57
0.5	100	0.1	19.7	0.40	0.56	-2.2	0.35	0.35	6.7	0.36	0.39
0.5	100	0.2	20.8	0.97	1.35	-4.2	0.83	0.86	5.4	0.85	0.89
0.5	100	0.4	23.8	2.26	3.27	-6.6	1.90	2.01	4.7	1.94	2.00
0.5	110	0.1	4.8	0.41	0.63	1.0	0.18	0.20	2.6	0.28	0.38
0.5	110	0.2	15.4	1.20	2.06	-1.3	0.70	0.71	5.1	0.85	1.01
0.5	110	0.4	20.5	2.72	4.23	-7.0	2.09	2.37	2.5	2.16	2.20
1	90	0.1	52.2	0.20	0.24	16.6	0.16	0.16	30.3	0.17	0.19
1	90	0.2	35.1	0.86	1.19	-4.1	0.67	0.68	8.5	0.70	0.73
1	90	0.4	35.8	2.65	4.04	-14.0	1.95	2.29	-1.3	2.00	2.00
1	100	0.1	23.1	0.51	0.75	-5.5	0.41	0.43	3.3	0.43	0.44
1	100	0.2	25.2	1.30	1.97	-9.4	1.03	1.17	-0.6	1.07	1.07
1	100	0.4	30.7	3.26	5.19	-14.5	2.43	3.09	-4.9	2.48	2.57
1	110	0.1	6.2	0.48	0.78	1.0	0.17	0.20	2.7	0.28	0.39
1	110	0.2	18.9	1.58	2.72	-6.0	0.91	1.15	0.6	1.00	1.01
1	110	0.4	26.7	3.82	6.32	-13.3	2.63	3.63	-5.3	2.68	2.86
2	90	0.1	55.9	0.32	0.43	4.8	0.23	0.23	19.1	0.25	0.26
2	90	0.2	39.5	1.24	1.90	-13.1	0.86	0.99	-2.7	0.90	0.90
2	90	0.4	44.1	4.01	6.67	-23.1	2.62	3.83	-15.9	2.66	3.28
2	100	0.1	27.0	0.63	0.98	-7.8	0.48	0.52	0.5	0.52	0.52
2	100	0.2	30.9	1.75	2.86	-13.6	1.27	1.62	-7.2	1.36	1.46
2	100	0.4	39.1	4.70	8.10	-21.2	2.98	4.65	-17.0	3.36	4.41
2	110	0.1	7.7	0.54	0.94	1.2	0.19	0.22	2.9	0.29	0.41
2	110	0.2	22.3	2.13	3.57	-10.7	1.10	1.75	-4.5	1.17	1.30
2	110	0.4	33.8	5.33	9.28	-18.7	3.19	5.27	-15.0	3.37	4.77
$N = 1600, M = 800$											
0.5	90	0.1	4.8	0.01	0.02	1.0	0.01	0.01	2.9	0.01	0.01
0.5	90	0.2	2.5	0.08	0.09	0.2	0.08	0.08	1.3	0.08	0.08
0.5	90	0.4	2.2	0.28	0.30	-0.2	0.26	0.26	1.0	0.27	0.27
0.5	100	0.1	1.1	0.06	0.07	0.0	0.06	0.06	0.6	0.06	0.06
0.5	100	0.2	1.3	0.16	0.17	-0.1	0.15	0.15	0.6	0.15	0.16
0.5	100	0.4	2.0	0.37	0.42	-0.2	0.35	0.35	0.8	0.36	0.37
0.5	110	0.1	0.0	0.00	0.00	0.0	0.00	0.00	0.0	0.00	0.00
0.5	110	0.2	0.9	0.18	0.21	-0.2	0.18	0.18	0.4	0.18	0.18
0.5	110	0.4	1.7	0.44	0.51	-0.2	0.42	0.42	0.7	0.42	0.44
1	90	0.1	4.0	0.03	0.03	0.0	0.03	0.03	2.0	0.03	0.03
1	90	0.2	2.0	0.13	0.14	-0.5	0.13	0.13	0.7	0.13	0.13
1	90	0.4	2.6	0.43	0.48	-0.8	0.40	0.41	0.6	0.41	0.41
1	100	0.1	1.0	0.08	0.08	-0.3	0.08	0.08	0.4	0.08	0.08
1	100	0.2	1.3	0.20	0.22	-0.4	0.20	0.20	0.4	0.20	0.20
1	100	0.4	2.3	0.52	0.60	-0.6	0.48	0.49	0.5	0.49	0.50
1	110	0.1	0.0	0.00	0.00	0.0	0.00	0.00	0.0	0.00	0.00
1	110	0.2	1.0	0.24	0.27	-0.3	0.23	0.24	0.3	0.24	0.24
1	110	0.4	1.8	0.60	0.69	-0.6	0.56	0.57	0.4	0.57	0.57
2	90	0.1	2.0	0.04	0.04	-2.4	0.04	0.04	-0.0	0.04	0.04
2	90	0.2	1.4	0.17	0.18	-1.0	0.15	0.16	0.2	0.17	0.17
2	90	0.4	2.7	0.47	0.57	-1.2	0.51	0.53	0.3	0.44	0.44
2	100	0.1	2.1	0.11	0.13	0.6	0.11	0.11	1.2	0.11	0.11
2	100	0.2	1.8	0.32	0.35	-0.2	0.30	0.30	0.7	0.31	0.31
2	100	0.4	3.0	0.68	0.84	-0.8	0.71	0.72	0.4	0.63	0.64
2	110	0.1	0.0	0.00	0.00	0.0	0.00	0.00	0.0	0.00	0.00
2	110	0.2	0.5	0.39	0.40	-0.8	0.41	0.42	-0.1	0.39	0.39
2	110	0.4	1.7	0.85	0.93	-0.8	0.82	0.83	0.1	0.85	0.85
$N = 12800, M = 100$											
0.5	90	0.1	0.2	0.01	0.01	-0.4	0.01	0.01	-0.1	0.01	0.01
0.5	90	0.2	0.2	0.03	0.03	-0.2	0.03	0.03	0.0	0.03	0.03

Continued on following page

Table 3.C.1, continued

			LSM			F-LSM			FS-LSM		
T	S_0	σ	**Bias**	**S.E.**	**RMSE**	**Bias**	**S.E.**	**RMSE**	**Bias**	**S.E.**	**RMSE**
0.5	90	0.4	0.2	0.10	0.10	-0.2	0.09	0.09	-0.0	0.09	0.09
0.5	100	0.1	0.0	0.02	0.02	-0.2	0.02	0.02	-0.1	0.02	0.02
0.5	100	0.2	0.1	0.05	0.06	-0.2	0.05	0.05	-0.0	0.05	0.05
0.5	100	0.4	0.2	0.12	0.13	-0.1	0.13	0.13	0.0	0.13	0.13
0.5	110	0.1	0.0	0.00	0.00	0.0	0.00	0.00	0.0	0.00	0.00
0.5	110	0.2	0.1	0.06	0.06	-0.1	0.06	0.06	-0.0	0.06	0.06
0.5	110	0.4	0.1	0.15	0.16	-0.2	0.15	0.15	-0.0	0.15	0.15
1	90	0.1	0.4	0.01	0.01	-0.3	0.01	0.01	0.0	0.01	0.01
1	90	0.2	0.2	0.05	0.05	-0.4	0.04	0.05	-0.1	0.05	0.05
1	90	0.4	0.1	0.15	0.15	-0.3	0.14	0.14	-0.1	0.14	0.14
1	100	0.1	0.0	0.03	0.03	-0.2	0.03	0.03	-0.1	0.03	0.03
1	100	0.2	0.0	0.07	0.07	-0.3	0.07	0.07	-0.1	0.07	0.07
1	100	0.4	0.2	0.17	0.17	-0.2	0.17	0.17	-0.0	0.17	0.17
1	110	0.1	0.0	0.00	0.00	0.0	0.00	0.00	0.0	0.00	0.00
1	110	0.2	0.2	0.08	0.08	-0.1	0.08	0.08	0.1	0.08	0.08
1	110	0.4	0.1	0.20	0.20	-0.2	0.19	0.19	-0.0	0.19	0.19
2	90	0.1	0.1	0.01	0.01	-0.5	0.01	0.01	-0.2	0.01	0.01
2	90	0.2	-0.2	0.04	0.05	-0.6	0.05	0.05	-0.4	0.05	0.05
2	90	0.4	0.5	0.12	0.14	-0.2	0.13	0.13	0.1	0.13	0.13
2	100	0.1	0.2	0.03	0.03	-0.1	0.03	0.03	0.1	0.03	0.03
2	100	0.2	0.2	0.07	0.07	-0.1	0.07	0.07	0.1	0.07	0.07
2	100	0.4	0.3	0.16	0.17	-0.0	0.17	0.17	0.1	0.16	0.16
2	110	0.1	0.0	0.00	0.00	0.0	0.00	0.00	0.0	0.00	0.00
2	110	0.2	0.1	0.08	0.08	-0.2	0.08	0.08	-0.1	0.08	0.08
2	110	0.4	0.2	0.18	0.18	-0.2	0.19	0.20	0.0	0.19	0.19

This table shows detailed pricing results for LSM, F-LSM, and FS-LSM prices for single-asset call options with strike price $K = 100$ and weekly exercise opportunities $J/T = 50$. Price processes follow a GBM with $r = 0.05$, $q = 0.1$, and $\sigma = 0.2$. Continuation values are estimated with a polynomial regression of order $L = 3$. The table reports relative bias ("Bias") as a percentage of the benchmark price, standard error ("S.E."), and RMSE. Detailed results are reported for $T \in \{0.5, 1, 2\}$, $S_0 \in \{90, 100, 110\}$, $\sigma \in \{0.1, 0.2, 0.3\}$ and $N \in \{50, 1600, 12800\}$, and calculated from $M = 1280000/N$ independent replications.

Table 3.C.2: Detailed results for LSM, F-LSM, and FS-LSM prices as a function of J/T with $D = 1$.

			LSM			F-LSM			FS-LSM		
S_0	J/T	N	**Bias**	**S.E.**	**RMSE**	**Bias**	**S.E.**	**RMSE**	**Bias**	**S.E.**	**RMSE**
90	5	50	17.7	0.88	0.97	8.1	0.82	0.85	13.5	0.85	0.90
90	5	1600	1.2	0.14	0.14	0.1	0.14	0.14	0.7	0.14	0.14
90	5	12800	0.1	0.05	0.05	-0.1	0.05	0.05	0.0	0.05	0.05
90	10	50	24.3	0.86	1.03	8.6	0.77	0.80	16.8	0.80	0.89
90	10	1600	1.4	0.14	0.14	-0.0	0.14	0.14	0.8	0.14	0.14
90	10	12800	0.2	0.05	0.05	-0.1	0.05	0.05	0.0	0.05	0.05
90	15	50	27.5	0.85	1.07	6.9	0.74	0.76	16.9	0.78	0.87
90	15	1600	1.7	0.14	0.15	0.1	0.14	0.14	1.0	0.14	0.14
90	15	12800	0.4	0.05	0.05	0.1	0.05	0.05	0.3	0.05	0.05
90	20	50	29.6	0.87	1.11	5.3	0.73	0.74	16.3	0.78	0.86
90	20	1600	1.8	0.13	0.14	-0.0	0.13	0.13	0.9	0.13	0.13
90	20	12800	0.3	0.04	0.04	-0.1	0.04	0.04	0.1	0.04	0.04
90	25	50	31.0	0.86	1.13	3.5	0.73	0.73	15.3	0.76	0.84
90	25	1600	2.0	0.13	0.14	-0.0	0.13	0.13	1.0	0.13	0.13
90	25	12800	0.2	0.05	0.05	-0.2	0.05	0.05	0.0	0.05	0.05
90	30	50	32.6	0.86	1.15	1.7	0.70	0.70	14.1	0.74	0.81

Continued on following page

Table 3.C.2, continued

			LSM			F-LSM			FS-LSM		
S_0	J/T	N	**Bias**	**S.E.**	**RMSE**	**Bias**	**S.E.**	**RMSE**	**Bias**	**S.E.**	**RMSE**
90	30	1600	1.8	0.13	0.14	-0.3	0.13	0.13	0.8	0.13	0.13
90	30	12800	0.1	0.05	0.05	-0.3	0.05	0.05	-0.0	0.05	0.05
90	35	50	33.4	0.86	1.16	-0.0	0.69	0.69	12.7	0.73	0.79
90	35	1600	1.8	0.13	0.14	-0.4	0.13	0.13	0.7	0.13	0.13
90	35	12800	0.1	0.05	0.05	-0.3	0.05	0.05	-0.1	0.05	0.05
90	40	50	34.4	0.86	1.18	-1.1	0.69	0.69	11.7	0.72	0.77
90	40	1600	2.1	0.13	0.14	-0.2	0.13	0.13	0.9	0.13	0.13
90	40	12800	0.1	0.05	0.05	-0.3	0.05	0.05	-0.1	0.05	0.05
90	45	50	34.7	0.86	1.18	-2.5	0.67	0.68	10.1	0.71	0.75
90	45	1600	1.8	0.14	0.14	-0.6	0.13	0.13	0.5	0.13	0.13
90	45	12800	0.1	0.05	0.05	-0.4	0.05	0.05	-0.2	0.05	0.05
90	50	50	35.1	0.86	1.19	-4.1	0.67	0.68	8.5	0.70	0.73
90	50	1600	2.0	0.13	0.14	-0.5	0.13	0.13	0.7	0.13	0.13
90	50	12800	0.2	0.05	0.05	-0.4	0.04	0.05	-0.1	0.05	0.05
100	5	50	12.9	1.31	1.51	3.5	1.25	1.26	9.0	1.27	1.37
100	5	1600	0.8	0.22	0.22	0.0	0.22	0.22	0.4	0.22	0.22
100	5	12800	0.3	0.08	0.08	0.2	0.08	0.08	0.2	0.08	0.08
100	10	50	17.6	1.28	1.64	2.2	1.17	1.18	10.1	1.20	1.34
100	10	1600	0.8	0.21	0.21	-0.2	0.20	0.20	0.4	0.21	0.21
100	10	12800	-0.0	0.08	0.08	-0.2	0.08	0.08	-0.1	0.08	0.08
100	15	50	19.8	1.28	1.73	0.2	1.13	1.13	9.4	1.17	1.29
100	15	1600	1.1	0.21	0.22	-0.0	0.21	0.21	0.6	0.21	0.21
100	15	12800	0.2	0.07	0.07	0.0	0.07	0.07	0.1	0.07	0.07
100	20	50	21.5	1.30	1.81	-1.6	1.11	1.12	8.2	1.15	1.24
100	20	1600	1.2	0.20	0.22	-0.1	0.20	0.20	0.6	0.20	0.21
100	20	12800	0.1	0.07	0.07	-0.1	0.07	0.07	-0.0	0.07	0.07
100	25	50	22.4	1.31	1.86	-3.6	1.11	1.13	6.6	1.14	1.20
100	25	1600	1.3	0.20	0.22	-0.1	0.21	0.21	0.6	0.20	0.21
100	25	12800	0.0	0.07	0.07	-0.2	0.07	0.07	-0.1	0.07	0.07
100	30	50	23.2	1.29	1.88	-5.2	1.08	1.12	4.9	1.12	1.15
100	30	1600	1.4	0.20	0.22	-0.1	0.20	0.20	0.7	0.20	0.21
100	30	12800	0.2	0.08	0.08	-0.0	0.08	0.08	0.1	0.07	0.07
100	35	50	24.1	1.30	1.92	-6.2	1.08	1.14	3.6	1.11	1.13
100	35	1600	1.2	0.20	0.21	-0.4	0.19	0.19	0.4	0.19	0.20
100	35	12800	-0.0	0.08	0.08	-0.3	0.08	0.08	-0.1	0.08	0.08
100	40	50	24.5	1.31	1.95	-7.5	1.06	1.15	2.2	1.09	1.09
100	40	1600	1.3	0.21	0.22	-0.4	0.20	0.20	0.5	0.20	0.20
100	40	12800	0.1	0.08	0.08	-0.2	0.08	0.08	-0.0	0.08	0.08
100	45	50	24.9	1.31	1.97	-8.6	1.06	1.17	0.5	1.08	1.08
100	45	1600	1.3	0.21	0.22	-0.4	0.20	0.20	0.5	0.21	0.21
100	45	12800	0.1	0.07	0.07	-0.2	0.07	0.08	-0.0	0.07	0.07
100	50	50	25.2	1.30	1.97	-9.4	1.03	1.17	-0.6	1.07	1.07
100	50	1600	1.3	0.20	0.22	-0.4	0.20	0.20	0.4	0.20	0.20
100	50	12800	0.0	0.07	0.07	-0.3	0.07	0.07	-0.1	0.07	0.07
110	5	50	9.2	1.64	1.95	-0.4	1.58	1.58	4.8	1.57	1.66
110	5	1600	0.4	0.25	0.26	-0.0	0.25	0.25	0.2	0.25	0.25
110	5	12800	0.1	0.10	0.10	0.1	0.09	0.10	0.1	0.10	0.10
110	10	50	12.9	1.59	2.19	-0.8	1.44	1.45	6.0	1.47	1.63
110	10	1600	0.6	0.26	0.27	-0.1	0.26	0.26	0.3	0.26	0.26
110	10	12800	-0.0	0.09	0.09	-0.1	0.09	0.09	-0.1	0.09	0.09
110	15	50	14.5	1.58	2.32	-2.1	1.35	1.37	5.4	1.38	1.51
110	15	1600	0.7	0.24	0.25	-0.1	0.23	0.23	0.4	0.24	0.24
110	15	12800	0.0	0.08	0.08	-0.1	0.08	0.08	-0.0	0.08	0.08
110	20	50	15.8	1.58	2.43	-3.1	1.27	1.32	4.6	1.32	1.43
110	20	1600	0.8	0.24	0.26	-0.1	0.24	0.24	0.4	0.24	0.24
110	20	12800	0.0	0.08	0.08	-0.1	0.08	0.09	-0.0	0.08	0.08
110	25	50	16.7	1.59	2.52	-4.1	1.20	1.29	3.8	1.26	1.33
110	25	1600	0.9	0.24	0.26	-0.1	0.24	0.24	0.5	0.24	0.24
110	25	12800	0.0	0.08	0.08	-0.1	0.08	0.08	-0.0	0.08	0.08
110	30	50	17.3	1.58	2.58	-4.9	1.17	1.30	2.8	1.22	1.26
110	30	1600	1.0	0.23	0.26	-0.1	0.23	0.23	0.5	0.23	0.24
110	30	12800	0.1	0.08	0.08	-0.1	0.08	0.08	-0.0	0.08	0.08
110	35	50	18.0	1.59	2.64	-5.5	1.13	1.30	2.0	1.17	1.19
110	35	1600	0.9	0.23	0.25	-0.2	0.22	0.23	0.4	0.23	0.23

Continued on following page

Table 3.C.2, continued

			LSM			F-LSM			FS-LSM		
S_0	J/T	N	**Bias**	**S.E.**	**RMSE**	**Bias**	**S.E.**	**RMSE**	**Bias**	**S.E.**	**RMSE**
110	35	12800	0.0	0.09	0.09	-0.2	0.09	0.09	-0.1	0.09	0.09
110	40	50	18.3	1.58	2.66	-6.2	1.08	1.30	1.2	1.11	1.12
110	40	1600	0.9	0.24	0.26	-0.4	0.24	0.24	0.3	0.24	0.24
110	40	12800	0.1	0.10	0.10	-0.1	0.09	0.09	-0.0	0.09	0.09
110	45	50	18.6	1.59	2.71	-6.7	1.05	1.31	0.4	1.07	1.07
110	45	1600	0.9	0.24	0.27	-0.4	0.24	0.24	0.3	0.24	0.24
110	45	12800	0.1	0.09	0.09	-0.1	0.09	0.09	0.0	0.09	0.09
110	50	50	18.9	1.58	2.72	-7.2	1.02	1.32	-0.3	1.04	1.04
110	50	1600	1.0	0.24	0.27	-0.3	0.23	0.24	0.3	0.24	0.24
110	50	12800	0.2	0.08	0.08	-0.1	0.08	0.08	0.1	0.08	0.08

This table shows detailed pricing results for LSM, F-LSM, and FS-LSM prices as a function of J/T for single-asset call options with strike price $K = 100$, and maturity $T = 1$. We set GBM diffusion parameters to $r = 0.05$, $q = 0.1$, and $\sigma = 0.2$. Continuation values are estimated with a polynomial regression of order $L = 3$. The table reports relative bias ("Bias") as a percentage of the benchmark price, standard error ("S.E."), and RMSE. Detailed results are reported for $N \in \{50, 1600, 12800\}$, $S_0 \in \{90, 100, 110\}$ and $J/T \in \{5, 10, \ldots, 50\}$, and calculated from $M = 1280000/N$ independent replications.

Table 3.C.3: Detailed results for LSM, F-LSM, and FS-LSM prices as a function of L with $D = 1$.

			LSM			F-LSM			FS-LSM		
S_0	L	N	**Bias**	**S.E.**	**RMSE**	**Bias**	**S.E.**	**RMSE**	**Bias**	**S.E.**	**RMSE**
90	1	50	5.7	0.74	0.75	-14.5	0.65	0.72	-4.7	0.69	0.69
90	1	1600	0.5	0.12	0.12	-0.4	0.12	0.12	0.1	0.12	0.12
90	1	12800	-0.1	0.04	0.04	-0.2	0.04	0.04	-0.1	0.04	0.04
90	2	50	27.6	0.86	1.08	-3.3	0.70	0.71	5.9	0.74	0.75
90	2	1600	1.8	0.14	0.15	0.0	0.13	0.13	0.9	0.14	0.14
90	2	12800	0.1	0.05	0.05	-0.2	0.05	0.05	-0.0	0.05	0.05
90	3	50	35.1	0.86	1.19	-4.1	0.67	0.68	8.5	0.70	0.73
90	3	1600	2.0	0.13	0.14	-0.5	0.13	0.13	0.7	0.13	0.13
90	3	12800	0.2	0.05	0.05	-0.4	0.04	0.05	-0.1	0.05	0.05
90	4	50	42.8	0.90	1.35	-6.0	0.66	0.67	10.0	0.70	0.74
90	4	1600	3.6	0.14	0.16	-0.3	0.13	0.13	1.3	0.14	0.14
90	4	12800	0.5	0.05	0.05	-0.1	0.05	0.05	0.2	0.05	0.05
100	1	50	8.2	1.20	1.29	-7.0	1.07	1.14	0.1	1.11	1.11
100	1	1600	0.5	0.20	0.20	-0.2	0.20	0.20	0.1	0.20	0.20
100	1	12800	-0.0	0.07	0.07	-0.1	0.07	0.07	-0.1	0.07	0.07
100	2	50	19.6	1.25	1.69	-6.3	1.05	1.11	0.6	1.09	1.09
100	2	1600	1.1	0.20	0.21	-0.3	0.19	0.19	0.4	0.19	0.19
100	2	12800	-0.0	0.07	0.07	-0.2	0.07	0.07	-0.1	0.07	0.07
100	3	50	25.2	1.30	1.97	-9.4	1.03	1.17	-0.6	1.07	1.07
100	3	1600	1.3	0.20	0.22	-0.4	0.20	0.20	0.4	0.20	0.20
100	3	12800	0.0	0.07	0.07	-0.3	0.07	0.07	-0.1	0.07	0.07
100	4	50	31.1	1.31	2.25	-11.1	1.02	1.21	-0.4	1.04	1.04
100	4	1600	1.9	0.21	0.23	-0.4	0.20	0.20	0.6	0.20	0.20
100	4	12800	0.1	0.07	0.07	-0.3	0.07	0.08	-0.1	0.07	0.07
110	1	50	9.6	1.44	1.82	-7.1	1.17	1.43	-1.2	1.20	1.21
110	1	1600	0.6	0.24	0.25	-0.0	0.24	0.24	0.3	0.24	0.24
110	1	12800	0.0	0.09	0.09	-0.1	0.09	0.09	-0.0	0.09	0.09
110	2	50	13.7	1.50	2.20	-6.5	1.07	1.31	-0.7	1.10	1.10
110	2	1600	0.7	0.23	0.24	-0.3	0.23	0.23	0.2	0.23	0.23
110	2	12800	0.1	0.08	0.08	-0.0	0.08	0.08	0.1	0.08	0.08
110	3	50	18.9	1.58	2.72	-7.2	1.02	1.32	-0.3	1.04	1.04
110	3	1600	1.0	0.24	0.27	-0.3	0.23	0.24	0.3	0.24	0.24

Continued on following page

Table 3.C.3, continued

			LSM			F-LSM			FS-LSM		
S_0	L	N	**Bias**	**S.E.**	**RMSE**	**Bias**	**S.E.**	**RMSE**	**Bias**	**S.E.**	**RMSE**
110	3	12800	0.2	0.08	0.08	-0.1	0.08	0.08	0.1	0.08	0.08
110	4	50	22.8	1.59	3.11	-7.1	0.96	1.28	1.1	1.03	1.04
110	4	1600	1.2	0.24	0.28	-0.4	0.24	0.24	0.4	0.24	0.24
110	4	12800	0.2	0.08	0.08	-0.1	0.08	0.08	0.1	0.08	0.08

This table shows detailed pricing results for LSM, F-LSM, and FS-LSM prices as a function of L for single-asset call options with strike price $K = 100$, maturity $T = 1$, and weekly exercise opportunities $J/T = 50$. We set GBM diffusion parameters to $r = 0.05$, $q = 0.1$, and $\sigma = 0.2$. The table reports relative bias ("Bias") as a percentage of the benchmark price, standard error ("S.E."), and RMSE. Detailed results are reported for $N \in \{50, 1600, 12800\}$, $S_0 \in \{90, 100, 110\}$ and $L \in \{1, 2, 3, 4\}$, and calculated from $M = 1280000/N$ independent replications.

Table 3.C.4: Detailed results for LSM, F-LSM, and FS-LSM prices as a function of J/T with $D = 5$.

			LSM			F-LSM			FS-LSM		
S_0	J/T	N	**Bias**	**S.E.**	**RMSE**	**Bias**	**S.E.**	**RMSE**	**Bias**	**S.E.**	**RMSE**
90	5	50	19.5	1.49	2.33	7.6	1.39	1.56	14.1	1.43	1.93
90	5	1600	1.3	0.24	0.27	-0.0	0.24	0.24	0.7	0.24	0.25
90	5	12800	0.2	0.08	0.08	-0.1	0.08	0.08	0.1	0.08	0.08
90	10	50	26.5	1.47	2.84	6.4	1.31	1.44	16.7	1.36	2.05
90	10	1600	1.5	0.25	0.28	-0.6	0.24	0.25	0.5	0.24	0.25
90	10	12800	0.1	0.09	0.10	-0.2	0.09	0.10	-0.0	0.09	0.09
90	15	50	29.6	1.44	3.06	2.9	1.26	1.29	15.9	1.31	1.95
90	15	1600	1.5	0.25	0.28	-1.0	0.24	0.26	0.3	0.24	0.24
90	15	12800	0.3	0.08	0.09	-0.2	0.08	0.08	0.1	0.08	0.08
90	20	50	31.6	1.44	3.22	-1.0	1.24	1.25	14.2	1.27	1.82
90	20	1600	1.7	0.25	0.29	-1.2	0.24	0.27	0.2	0.24	0.24
90	20	12800	0.3	0.09	0.09	-0.2	0.08	0.09	0.1	0.09	0.09
90	25	50	32.6	1.41	3.27	-5.0	1.20	1.28	11.9	1.22	1.63
90	25	1600	1.7	0.24	0.28	-1.6	0.23	0.27	0.0	0.23	0.23
90	25	12800	0.2	0.09	0.09	-0.4	0.09	0.09	-0.1	0.09	0.09
90	30	50	33.5	1.40	3.33	-8.7	1.19	1.43	9.6	1.21	1.49
90	30	1600	1.7	0.24	0.28	-1.8	0.23	0.29	-0.1	0.24	0.24
90	30	12800	0.2	0.08	0.08	-0.5	0.08	0.09	-0.2	0.08	0.08
90	35	50	34.0	1.43	3.38	-12.3	1.21	1.64	7.1	1.21	1.37
90	35	1600	1.6	0.24	0.28	-2.2	0.24	0.31	-0.4	0.24	0.24
90	35	12800	0.1	0.08	0.08	-0.6	0.08	0.09	-0.2	0.08	0.08
90	40	50	34.5	1.41	3.40	-15.7	1.21	1.86	4.7	1.19	1.26
90	40	1600	1.7	0.24	0.28	-2.4	0.23	0.32	-0.5	0.23	0.24
90	40	12800	0.3	0.08	0.08	-0.4	0.08	0.09	-0.1	0.08	0.08
90	45	50	35.1	1.40	3.44	-18.8	1.19	2.06	2.4	1.17	1.19
90	45	1600	1.8	0.24	0.29	-2.6	0.23	0.33	-0.6	0.23	0.24
90	45	12800	0.1	0.08	0.08	-0.7	0.08	0.10	-0.3	0.08	0.08
90	50	50	35.1	1.41	3.44	-22.0	1.18	2.29	-0.3	1.16	1.16
90	50	1600	1.7	0.24	0.28	-3.0	0.24	0.36	-0.9	0.23	0.25
90	50	12800	0.3	0.09	0.09	-0.5	0.09	0.10	-0.1	0.09	0.09
100	5	50	13.9	1.84	3.19	2.6	1.80	1.86	8.8	1.79	2.42
100	5	1600	0.8	0.32	0.36	-0.2	0.32	0.32	0.4	0.32	0.33
100	5	12800	0.1	0.10	0.11	-0.0	0.10	0.10	0.0	0.10	0.10
100	10	50	18.0	1.83	3.81	-1.8	1.80	1.84	8.2	1.75	2.32
100	10	1600	0.9	0.33	0.37	-0.6	0.33	0.35	0.1	0.33	0.33
100	10	12800	0.1	0.12	0.12	-0.1	0.12	0.12	-0.0	0.12	0.12
100	15	50	19.4	1.81	4.03	-7.2	1.85	2.28	5.6	1.74	2.02
100	15	1600	0.9	0.33	0.37	-0.9	0.32	0.36	0.0	0.32	0.32

Continued on following page

Table 3.C.4, continued

			LSM			F-LSM			FS-LSM		
S_0	J/T	N	**Bias**	**S.E.**	**RMSE**	**Bias**	**S.E.**	**RMSE**	**Bias**	**S.E.**	**RMSE**
100	15	12800	0.2	0.10	0.10	-0.1	0.10	0.10	0.0	0.10	0.10
100	20	50	20.3	1.81	4.16	-12.8	1.92	3.04	2.5	1.75	1.81
100	20	1600	1.0	0.32	0.37	-1.1	0.32	0.38	-0.1	0.32	0.32
100	20	12800	0.1	0.11	0.12	-0.2	0.11	0.12	-0.0	0.11	0.11
100	25	50	20.6	1.80	4.20	-17.8	2.02	3.84	-0.8	1.78	1.79
100	25	1600	0.9	0.32	0.36	-1.4	0.31	0.41	-0.3	0.32	0.32
100	25	12800	0.1	0.12	0.12	-0.2	0.12	0.13	-0.1	0.12	0.12
100	30	50	20.8	1.82	4.23	-22.5	2.06	4.61	-4.0	1.79	1.93
100	30	1600	1.0	0.32	0.36	-1.6	0.31	0.43	-0.4	0.32	0.32
100	30	12800	0.1	0.10	0.11	-0.3	0.10	0.11	-0.1	0.10	0.10
100	35	50	20.8	1.84	4.23	-26.9	2.15	5.37	-7.2	1.87	2.29
100	35	1600	0.9	0.33	0.36	-1.9	0.32	0.48	-0.6	0.32	0.34
100	35	12800	0.1	0.11	0.11	-0.4	0.11	0.13	-0.2	0.11	0.11
100	40	50	20.8	1.85	4.22	-30.5	2.20	5.99	-10.3	1.91	2.68
100	40	1600	1.0	0.32	0.37	-2.0	0.31	0.48	-0.6	0.31	0.33
100	40	12800	0.2	0.10	0.10	-0.3	0.10	0.11	-0.1	0.10	0.10
100	45	50	20.8	1.85	4.22	-33.9	2.21	6.57	-13.2	1.93	3.09
100	45	1600	1.0	0.32	0.36	-2.2	0.31	0.51	-0.7	0.31	0.33
100	45	12800	0.1	0.10	0.10	-0.4	0.10	0.13	-0.2	0.10	0.11
100	50	50	20.5	1.89	4.18	-37.1	2.23	7.11	-16.2	1.97	3.54
100	50	1600	0.9	0.32	0.36	-2.5	0.31	0.55	-0.9	0.31	0.36
100	50	12800	0.2	0.11	0.11	-0.4	0.11	0.13	-0.1	0.11	0.11
110	5	50	10.5	2.07	3.73	1.8	2.06	2.13	6.6	2.02	2.81
110	5	1600	0.6	0.37	0.41	-0.2	0.37	0.37	0.2	0.37	0.37
110	5	12800	0.1	0.12	0.12	-0.0	0.12	0.12	0.0	0.12	0.12
110	10	50	13.4	2.06	4.45	-1.5	2.09	2.14	6.1	2.00	2.69
110	10	1600	0.6	0.37	0.41	-0.5	0.37	0.40	0.1	0.37	0.37
110	10	12800	0.1	0.13	0.14	-0.1	0.13	0.14	-0.0	0.13	0.13
110	15	50	14.3	2.04	4.69	-5.5	2.13	2.67	4.1	1.99	2.33
110	15	1600	0.6	0.38	0.42	-0.7	0.37	0.43	-0.0	0.37	0.38
110	15	12800	0.1	0.11	0.11	-0.1	0.11	0.12	-0.0	0.11	0.11
110	20	50	15.0	2.04	4.84	-9.4	2.20	3.52	1.9	2.00	2.07
110	20	1600	0.7	0.37	0.42	-0.9	0.37	0.45	-0.1	0.37	0.37
110	20	12800	0.1	0.13	0.13	-0.1	0.12	0.13	-0.0	0.13	0.13
110	25	50	15.2	2.04	4.89	-12.9	2.29	4.41	-0.4	2.03	2.04
110	25	1600	0.7	0.37	0.42	-1.1	0.37	0.49	-0.3	0.37	0.38
110	25	12800	0.1	0.13	0.13	-0.2	0.13	0.14	-0.1	0.13	0.13
110	30	50	15.2	2.04	4.90	-16.1	2.30	5.24	-2.8	2.04	2.20
110	30	1600	0.7	0.37	0.42	-1.3	0.37	0.53	-0.4	0.37	0.38
110	30	12800	0.1	0.12	0.12	-0.2	0.12	0.14	-0.1	0.12	0.12
110	35	50	15.2	2.08	4.90	-19.0	2.35	6.01	-5.1	2.10	2.58
110	35	1600	0.6	0.37	0.41	-1.5	0.36	0.57	-0.5	0.37	0.40
110	35	12800	0.0	0.13	0.13	-0.3	0.12	0.15	-0.1	0.13	0.13
110	40	50	15.2	2.09	4.90	-21.5	2.39	6.70	-7.2	2.15	3.01
110	40	1600	0.7	0.37	0.42	-1.6	0.37	0.59	-0.5	0.37	0.40
110	40	12800	0.1	0.11	0.12	-0.2	0.11	0.13	-0.1	0.11	0.11
110	45	50	15.2	2.10	4.89	-23.8	2.38	7.32	-9.3	2.17	3.47
110	45	1600	0.7	0.36	0.42	-1.7	0.36	0.62	-0.6	0.36	0.40
110	45	12800	0.1	0.11	0.11	-0.3	0.11	0.14	-0.1	0.11	0.12
110	50	50	14.9	2.14	4.84	-25.7	2.38	7.84	-11.3	2.21	3.97
110	50	1600	0.6	0.37	0.41	-1.9	0.36	0.67	-0.7	0.36	0.42
110	50	12800	0.1	0.13	0.13	-0.3	0.13	0.15	-0.1	0.13	0.13

This table shows detailed pricing results for LSM, F-LSM, and FS-LSM prices as a function of J/T for five-asset max-call options with strike price $K = 100$, and maturity $T = 1$. We set GBM diffusion parameters to $r = 0.05$, $q = 0.1$, and $\sigma = 0.2$. Continuation values are estimated with a polynomial regression of order $L = 3$. The table reports relative bias ("Bias") as a percentage of the benchmark price, standard error ("S.E."), and RMSE. Detailed results are reported for $N \in \{50, 1600, 12800\}$, $S_0 \in \{90, 100, 110\}$ and $J/T \in \{5, 10, \ldots, 50\}$, and calculated from $M = 1280000/N$ independent replications.

Table 3.C.5: Detailed results for LSM, F-LSM, and FS-LSM prices as a function of L with $D = 5$.

			LSM			F-LSM			FS-LSM		
S_0	L	N	**Bias**	**S.E.**	**RMSE**	**Bias**	**S.E.**	**RMSE**	**Bias**	**S.E.**	**RMSE**
90	1	50	8.9	1.32	1.50	-25.0	1.21	2.37	-9.3	1.22	1.44
90	1	1600	0.1	0.23	0.23	-1.7	0.23	0.26	-0.8	0.23	0.24
90	1	12800	0.1	0.08	0.08	-0.2	0.08	0.08	-0.0	0.08	0.08
90	2	50	25.0	1.42	2.64	-23.5	1.22	2.42	-6.1	1.21	1.33
90	2	1600	1.0	0.25	0.26	-2.0	0.24	0.30	-0.5	0.24	0.25
90	2	12800	0.2	0.09	0.09	-0.2	0.09	0.09	0.0	0.09	0.09
90	3	50	35.1	1.41	3.44	-22.0	1.18	2.29	-0.3	1.16	1.16
90	3	1600	1.7	0.24	0.28	-3.0	0.24	0.36	-0.9	0.23	0.25
90	3	12800	0.3	0.09	0.09	-0.5	0.09	0.10	-0.1	0.09	0.09
90	4	50	42.3	1.44	4.06	-16.5	1.12	1.86	6.8	1.14	1.29
90	4	1600	2.3	0.25	0.32	-3.1	0.24	0.36	-0.6	0.24	0.25
90	4	12800	0.4	0.09	0.10	-0.4	0.09	0.10	-0.0	0.09	0.09
100	1	50	6.4	1.74	2.06	-22.2	1.77	4.24	-9.6	1.72	2.40
100	1	1600	0.1	0.31	0.32	-1.2	0.32	0.38	-0.5	0.31	0.33
100	1	12800	0.1	0.11	0.11	-0.1	0.11	0.11	-0.0	0.11	0.11
100	2	50	15.1	1.86	3.32	-30.1	2.21	5.91	-14.7	1.97	3.32
100	2	1600	0.6	0.32	0.34	-1.7	0.32	0.44	-0.6	0.32	0.34
100	2	12800	0.1	0.11	0.11	-0.2	0.11	0.11	-0.0	0.11	0.11
100	3	50	20.5	1.89	4.18	-37.1	2.23	7.11	-16.2	1.97	3.54
100	3	1600	0.9	0.32	0.36	-2.5	0.31	0.55	-0.9	0.31	0.36
100	3	12800	0.2	0.11	0.11	-0.4	0.11	0.13	-0.1	0.11	0.11
100	4	50	24.8	1.92	4.91	-40.3	1.94	7.58	-15.6	3.11	4.21
100	4	1600	1.3	0.32	0.40	-2.8	0.32	0.59	-0.9	0.32	0.35
100	4	12800	0.2	0.11	0.12	-0.3	0.11	0.13	-0.1	0.12	0.12
110	1	50	4.8	1.98	2.41	-16.2	2.07	5.04	-7.1	2.01	2.83
110	1	1600	0.1	0.36	0.36	-0.8	0.36	0.43	-0.4	0.36	0.38
110	1	12800	0.0	0.13	0.13	-0.1	0.13	0.13	-0.0	0.13	0.13
110	2	50	11.1	2.13	3.86	-21.4	2.48	6.71	-10.4	2.25	3.78
110	2	1600	0.4	0.36	0.38	-1.4	0.36	0.54	-0.5	0.36	0.39
110	2	12800	0.1	0.13	0.13	-0.1	0.13	0.14	-0.0	0.13	0.13
110	3	50	14.9	2.14	4.84	-25.7	2.38	7.84	-11.3	2.21	3.97
110	3	1600	0.6	0.37	0.41	-1.9	0.36	0.67	-0.7	0.36	0.42
110	3	12800	0.1	0.13	0.13	-0.3	0.13	0.15	-0.1	0.13	0.13
110	4	50	18.1	2.19	5.70	-27.3	2.03	8.19	-10.7	2.52	4.00
110	4	1600	0.9	0.37	0.45	-2.2	0.36	0.73	-0.7	0.36	0.42
110	4	12800	0.1	0.13	0.14	-0.3	0.13	0.16	-0.1	0.13	0.13

This table shows detailed pricing results for LSM, F-LSM, and FS-LSM prices as a function of L for five-asset max-call options with strike price $K = 100$, maturity $T = 1$, and weekly exercise opportunities $J/T = 50$. We set GBM diffusion parameters to $r = 0.05$, $q = 0.1$, and $\sigma = 0.2$. The table reports relative bias ("Bias") as a percentage of the benchmark price, standard error ("S.E."), and RMSE. Detailed results are reported for $N \in \{50, 1600, 12800\}$, $S_0 \in \{90, 100, 110\}$ and $L \in \{1, 2, 3, 4\}$, and calculated from $M = 1280000/N$ independent replications.

Table 3.C.6: Detailed results for LSM, F-LSM, and FS-LSM prices as a function of J/T with JDM underlying asset ($\lambda = 1$).

			LSM			F-LSM			FS-LSM		
S_0	J/T	N	**Bias**	**S.E.**	**RMSE**	**Bias**	**S.E.**	**RMSE**	**Bias**	**S.E.**	**RMSE**
90	5	50	15.9	1.90	2.07	7.1	1.77	1.81	11.8	1.82	1.92
90	5	1600	1.8	0.32	0.33	0.7	0.31	0.32	1.3	0.32	0.32
90	5	12800	0.4	0.11	0.11	0.2	0.11	0.11	0.3	0.11	0.11
90	10	50	21.8	1.92	2.23	7.1	1.71	1.74	14.2	1.78	1.93
90	10	1600	1.8	0.31	0.33	0.4	0.30	0.30	1.2	0.31	0.31
90	10	12800	0.4	0.11	0.11	0.1	0.11	0.11	0.2	0.11	0.11
90	15	50	24.4	1.94	2.31	4.7	1.65	1.66	13.6	1.73	1.87
90	15	1600	1.9	0.32	0.34	0.3	0.31	0.31	1.1	0.32	0.32
90	15	12800	0.5	0.11	0.11	0.3	0.11	0.11	0.4	0.11	0.11
90	20	50	25.2	1.93	2.33	1.6	1.60	1.61	11.7	1.69	1.80
90	20	1600	2.0	0.30	0.32	0.2	0.30	0.30	1.1	0.30	0.31
90	20	12800	0.1	0.10	0.10	-0.2	0.10	0.10	-0.1	0.10	0.10
90	25	50	26.6	1.93	2.37	-0.7	1.56	1.56	10.3	1.65	1.73
90	25	1600	2.0	0.30	0.32	-0.0	0.29	0.29	1.0	0.29	0.30
90	25	12800	0.6	0.10	0.10	0.3	0.10	0.10	0.4	0.10	0.10
90	30	50	27.3	1.92	2.39	-2.7	1.53	1.54	8.5	1.61	1.67
90	30	1600	1.8	0.30	0.31	-0.3	0.28	0.28	0.7	0.29	0.29
90	30	12800	0.6	0.09	0.10	0.3	0.09	0.09	0.4	0.09	0.10
90	35	50	28.2	1.97	2.45	-4.9	1.53	1.55	6.9	1.61	1.65
90	35	1600	1.9	0.30	0.32	-0.3	0.29	0.29	0.7	0.29	0.29
90	35	12800	0.3	0.10	0.10	0.0	0.10	0.10	0.2	0.10	0.10
90	40	50	28.5	1.94	2.44	-6.8	1.49	1.53	5.1	1.57	1.59
90	40	1600	2.0	0.32	0.33	-0.2	0.30	0.30	0.8	0.31	0.31
90	40	12800	0.4	0.10	0.10	0.0	0.10	0.10	0.2	0.10	0.10
90	45	50	28.8	1.95	2.45	-8.4	1.48	1.54	3.5	1.57	1.58
90	45	1600	2.0	0.31	0.32	-0.3	0.29	0.29	0.7	0.30	0.30
90	45	12800	0.2	0.09	0.09	-0.2	0.09	0.09	-0.0	0.09	0.09
90	50	50	28.8	1.97	2.47	-10.6	1.48	1.57	1.3	1.55	1.56
90	50	1600	2.0	0.31	0.33	-0.4	0.29	0.29	0.6	0.29	0.29
90	50	12800	0.1	0.09	0.09	-0.2	0.09	0.09	-0.1	0.09	0.09
100	5	50	13.5	2.41	2.72	4.2	2.27	2.30	9.2	2.31	2.46
100	5	1600	1.4	0.40	0.42	0.4	0.39	0.39	1.0	0.39	0.40
100	5	12800	0.3	0.13	0.13	0.2	0.13	0.13	0.3	0.13	0.13
100	10	50	18.6	2.44	2.99	2.3	2.18	2.19	10.1	2.23	2.42
100	10	1600	1.7	0.40	0.43	0.4	0.39	0.39	1.1	0.39	0.41
100	10	12800	0.3	0.14	0.14	0.2	0.14	0.14	0.2	0.14	0.14
100	15	50	20.7	2.43	3.10	-0.9	2.11	2.11	8.4	2.16	2.30
100	15	1600	1.7	0.42	0.44	0.3	0.41	0.41	1.0	0.41	0.42
100	15	12800	0.4	0.13	0.13	0.2	0.13	0.13	0.3	0.13	0.13
100	20	50	21.6	2.44	3.16	-4.3	2.07	2.11	5.7	2.12	2.18
100	20	1600	1.7	0.39	0.42	0.1	0.39	0.39	0.9	0.39	0.40
100	20	12800	0.2	0.12	0.12	-0.0	0.12	0.12	0.1	0.12	0.12
100	25	50	22.8	2.43	3.23	-7.3	2.03	2.14	3.5	2.07	2.09
100	25	1600	1.8	0.39	0.42	0.1	0.38	0.38	1.0	0.38	0.39
100	25	12800	0.3	0.13	0.13	0.1	0.13	0.13	0.1	0.12	0.13
100	30	50	23.2	2.44	3.26	-9.8	2.02	2.22	1.0	2.05	2.06
100	30	1600	1.7	0.38	0.42	0.0	0.37	0.37	0.9	0.37	0.38
100	30	12800	0.6	0.13	0.14	0.4	0.13	0.13	0.5	0.13	0.13
100	35	50	24.0	2.49	3.34	-12.1	1.99	2.29	-1.2	2.03	2.04
100	35	1600	1.8	0.39	0.43	-0.1	0.37	0.37	0.8	0.38	0.39
100	35	12800	0.3	0.13	0.13	0.0	0.13	0.13	0.1	0.13	0.13
100	40	50	24.3	2.45	3.33	-14.3	1.95	2.36	-3.6	1.97	2.00
100	40	1600	1.9	0.41	0.45	-0.0	0.39	0.39	0.8	0.40	0.41
100	40	12800	0.4	0.13	0.14	0.1	0.13	0.13	0.2	0.13	0.14
100	45	50	24.6	2.46	3.36	-16.1	1.92	2.44	-5.7	1.98	2.04
100	45	1600	2.0	0.40	0.44	0.0	0.38	0.38	0.9	0.39	0.39
100	45	12800	0.2	0.13	0.13	-0.1	0.12	0.13	-0.0	0.13	0.13
100	50	50	24.8	2.49	3.39	-18.0	1.91	2.54	-8.0	1.97	2.10
100	50	1600	2.0	0.39	0.43	-0.0	0.37	0.37	0.9	0.37	0.38
100	50	12800	0.2	0.12	0.12	-0.0	0.12	0.12	0.1	0.12	0.12
110	5	50	11.9	2.87	3.37	2.7	2.68	2.70	7.8	2.71	2.95
110	5	1600	1.0	0.48	0.50	0.2	0.47	0.47	0.7	0.47	0.49

Continued on following page

Table 3.C.6, continued

			LSM			F-LSM			FS-LSM		
S_0	J/T	N	**Bias**	**S.E.**	**RMSE**	**Bias**	**S.E.**	**RMSE**	**Bias**	**S.E.**	**RMSE**
110	5	12800	0.2	0.16	0.16	0.1	0.16	0.16	0.2	0.16	0.16
110	10	50	16.1	2.86	3.74	-0.0	2.48	2.48	7.8	2.54	2.79
110	10	1600	1.3	0.47	0.51	0.2	0.47	0.47	0.8	0.47	0.48
110	10	12800	0.3	0.16	0.17	0.1	0.16	0.16	0.2	0.16	0.17
110	15	50	18.0	2.83	3.90	-3.5	2.35	2.41	5.5	2.40	2.53
110	15	1600	1.3	0.46	0.51	0.2	0.46	0.46	0.8	0.46	0.47
110	15	12800	0.3	0.15	0.16	0.1	0.15	0.15	0.2	0.15	0.15
110	20	50	18.8	2.83	3.99	-6.9	2.19	2.42	2.7	2.27	2.31
110	20	1600	1.4	0.46	0.50	0.0	0.46	0.46	0.7	0.46	0.47
110	20	12800	0.1	0.14	0.14	-0.0	0.14	0.14	0.1	0.14	0.14
110	25	50	19.6	2.84	4.08	-9.6	2.13	2.57	0.3	2.20	2.20
110	25	1600	1.4	0.46	0.50	-0.1	0.44	0.44	0.6	0.45	0.46
110	25	12800	0.1	0.16	0.16	-0.1	0.16	0.16	0.0	0.16	0.16
110	30	50	20.3	2.83	4.15	-12.0	2.02	2.71	-2.0	2.09	2.11
110	30	1600	1.4	0.45	0.50	-0.1	0.43	0.43	0.6	0.44	0.45
110	30	12800	0.4	0.15	0.16	0.1	0.15	0.15	0.3	0.15	0.16
110	35	50	20.8	2.87	4.23	-13.7	1.96	2.83	-3.9	2.02	2.11
110	35	1600	1.5	0.46	0.51	-0.1	0.45	0.45	0.7	0.45	0.46
110	35	12800	0.2	0.15	0.15	-0.0	0.15	0.15	0.1	0.15	0.15
110	40	50	21.0	2.82	4.22	-15.8	1.83	2.98	-6.1	1.92	2.12
110	40	1600	1.3	0.47	0.51	-0.3	0.44	0.45	0.5	0.45	0.46
110	40	12800	0.2	0.15	0.15	-0.0	0.15	0.15	0.1	0.15	0.15
110	45	50	21.3	2.83	4.26	-17.4	1.76	3.13	-7.8	1.84	2.18
110	45	1600	1.6	0.46	0.52	-0.2	0.45	0.45	0.6	0.45	0.46
110	45	12800	0.1	0.15	0.15	-0.1	0.15	0.15	-0.0	0.14	0.14
110	50	50	21.4	2.89	4.31	-18.8	1.69	3.28	-9.5	1.79	2.28
110	50	1600	1.6	0.46	0.52	-0.1	0.43	0.43	0.6	0.44	0.45
110	50	12800	0.1	0.16	0.16	-0.1	0.15	0.16	-0.0	0.15	0.15

This table shows detailed pricing results for LSM, F-LSM, and FS-LSM prices as a function of J/T for single-asset call options with strike price $K = 100$, and maturity $T = 1$. Price processes follow a JDM with $r = 0.05$, $q = 0.1$, and $\sigma = 0.2$. The intensity parameter of the Poisson counting process is set to $\lambda = 1$ (annually), and log-jumps are normally distributed with mean zero and standard deviation 0.2. Continuation values are estimated with a polynomial regression of order $L = 3$. The table reports relative bias ("Bias") as a percentage of the benchmark price, standard error ("S.E."), and RMSE. Detailed results are reported for $N \in \{50, 1600, 12800\}$, $S_0 \in \{90, 100, 110\}$ and $J/T \in \{5, 10, \ldots, 50\}$, and calculated from $M = 1280000/N$ independent replications.

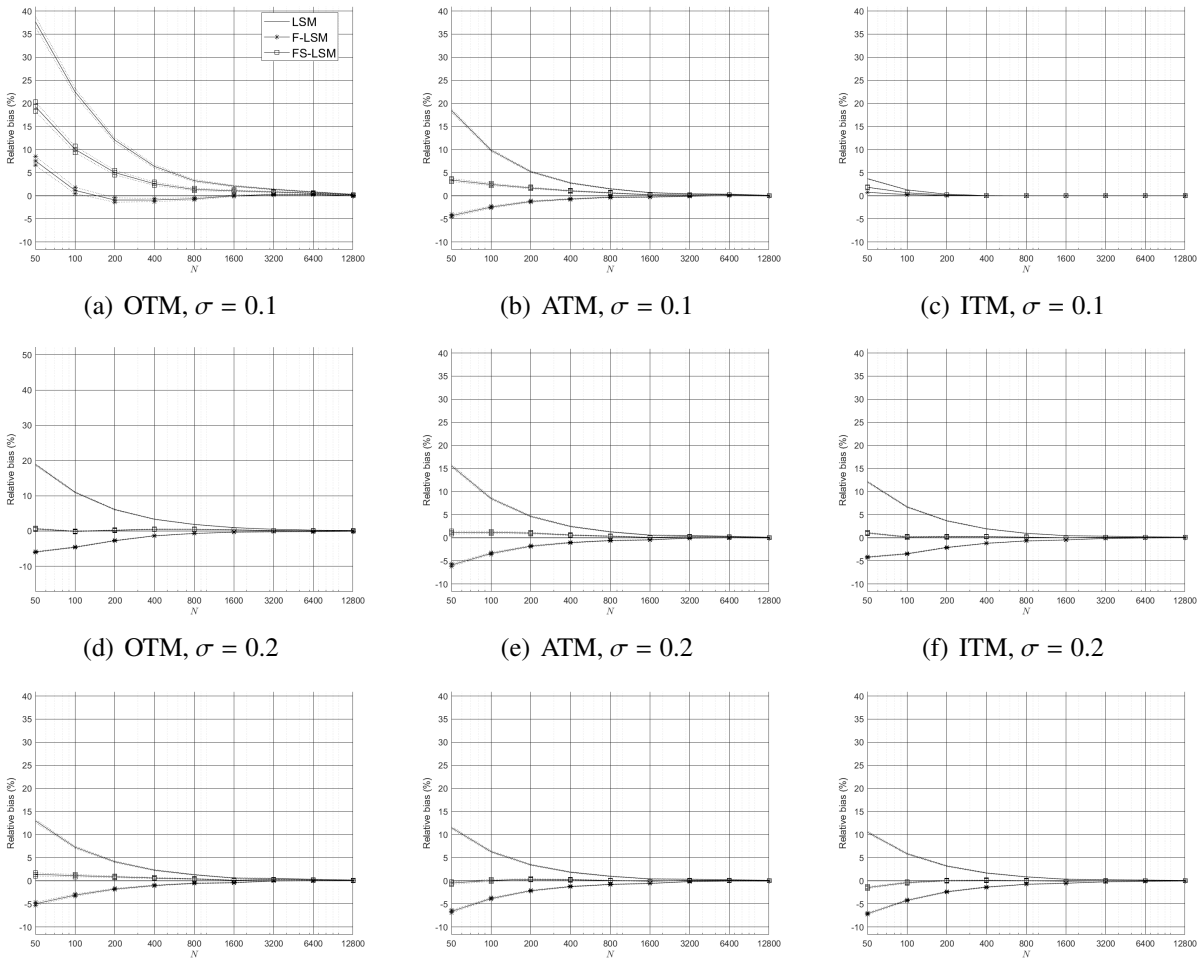
Table 3.C.7: Detailed results for LSM, F-LSM, and FS-LSM prices as a function of L with JDM underlying asset ($\lambda = 1$).

			LSM			F-LSM			FS-LSM		
S_0	L	N	**Bias**	**S.E.**	**RMSE**	**Bias**	**S.E.**	**RMSE**	**Bias**	**S.E.**	**RMSE**
90	1	50	7.0	1.87	1.90	-17.4	1.55	1.79	-7.1	1.63	1.67
90	1	1600	0.2	0.29	0.29	-1.2	0.29	0.29	-0.5	0.29	0.29
90	1	12800	-0.2	0.10	0.10	-0.4	0.09	0.10	-0.3	0.09	0.10
90	2	50	24.1	1.88	2.25	-6.5	1.50	1.54	2.5	1.58	1.59
90	2	1600	1.3	0.28	0.29	-0.5	0.28	0.28	0.3	0.28	0.28
90	2	12800	-0.0	0.09	0.09	-0.5	0.09	0.09	-0.3	0.09	0.09
90	3	50	28.8	1.97	2.47	-10.6	1.48	1.57	1.3	1.55	1.56
90	3	1600	2.0	0.31	0.33	-0.4	0.29	0.29	0.6	0.29	0.29
90	3	12800	0.1	0.09	0.09	-0.2	0.09	0.09	-0.1	0.09	0.09
90	4	50	34.1	1.98	2.65	-10.2	1.49	1.58	3.5	1.56	1.57
90	4	1600	2.7	0.30	0.33	-0.8	0.28	0.28	0.6	0.29	0.29
90	4	12800	-0.1	0.09	0.09	-0.6	0.09	0.09	-0.4	0.09	0.09
100	1	50	8.1	2.40	2.52	-14.6	2.03	2.44	-6.3	2.10	2.18
100	1	1600	0.4	0.37	0.38	-0.5	0.37	0.38	-0.1	0.37	0.37
100	1	12800	-0.2	0.13	0.13	-0.3	0.13	0.13	-0.2	0.13	0.13
100	2	50	19.0	2.37	2.95	-14.4	1.87	2.29	-5.7	1.93	2.00
100	2	1600	0.9	0.36	0.37	-0.5	0.36	0.36	0.1	0.36	0.36
100	2	12800	-0.2	0.12	0.12	-0.5	0.12	0.13	-0.4	0.12	0.12
100	3	50	24.8	2.49	3.39	-18.0	1.91	2.54	-8.0	1.97	2.10
100	3	1600	2.0	0.39	0.43	-0.0	0.37	0.37	0.9	0.37	0.38
100	3	12800	0.2	0.12	0.12	-0.0	0.12	0.12	0.1	0.12	0.12
100	4	50	28.2	2.49	3.62	-21.5	1.84	2.72	-9.3	1.89	2.08
100	4	1600	2.0	0.39	0.43	-0.6	0.36	0.37	0.5	0.37	0.38
100	4	12800	-0.0	0.12	0.12	-0.4	0.12	0.13	-0.2	0.12	0.13
110	1	50	9.7	2.77	3.13	-16.8	2.00	3.21	-9.0	2.11	2.50
110	1	1600	0.5	0.46	0.46	-0.4	0.46	0.46	0.1	0.46	0.46
110	1	12800	-0.2	0.16	0.16	-0.3	0.16	0.17	-0.2	0.16	0.17
110	2	50	16.0	2.78	3.66	-16.3	1.88	3.07	-8.2	1.96	2.31
110	2	1600	0.7	0.44	0.45	-0.6	0.43	0.44	-0.0	0.44	0.44
110	2	12800	-0.1	0.14	0.15	-0.4	0.14	0.16	-0.3	0.14	0.15
110	3	50	21.4	2.89	4.31	-18.8	1.69	3.28	-9.5	1.79	2.28
110	3	1600	1.6	0.46	0.52	-0.1	0.43	0.43	0.6	0.44	0.45
110	3	12800	0.1	0.16	0.16	-0.1	0.15	0.16	-0.0	0.15	0.15
110	4	50	24.2	2.91	4.65	-19.8	1.58	3.36	-9.0	1.72	2.19
110	4	1600	1.8	0.47	0.54	-0.5	0.44	0.44	0.5	0.45	0.46
110	4	12800	0.1	0.15	0.15	-0.2	0.15	0.15	-0.1	0.15	0.15

This table shows detailed pricing results for LSM, F-LSM, and FS-LSM prices as a function of L for single-asset call options with strike price $K = 100$, maturity $T = 1$, and weekly exercise opportunities $J/T = 50$. Price processes follow a JDM with $r = 0.05$, $q = 0.1$, and $\sigma = 0.2$. The intensity parameter of the Poisson counting process is set to $\lambda = 1$ (annually), and log-jumps are normally distributed with mean zero and standard deviation 0.2. The table reports relative bias ("Bias") as a percentage of the benchmark price, standard error ("S.E."), and RMSE. Detailed results are reported for $N \in \{50, 1600, 12800\}$, $S_0 \in \{90, 100, 110\}$ and $L \in \{1, 2, 3, 4\}$, and calculated from $M = 1280000/N$ independent replications.

3.D Additional figures

This appendix contains additional figures complementing those presented in the body of the paper.



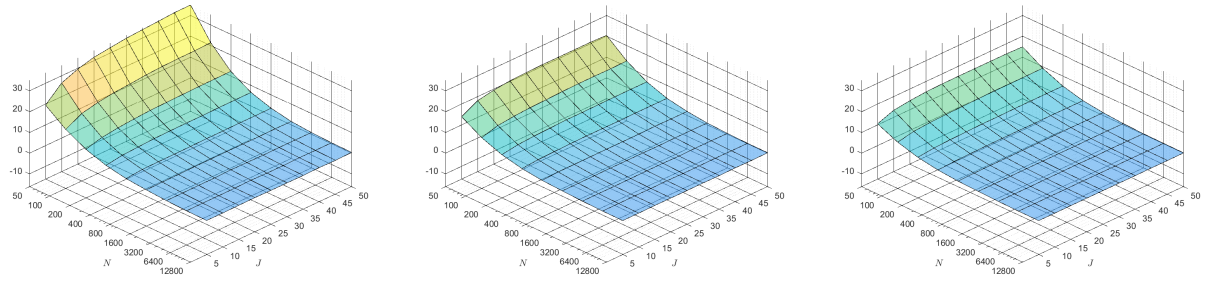
(a) OTM, $\sigma = 0.1$ (b) ATM, $\sigma = 0.1$ (c) ITM, $\sigma = 0.1$

(d) OTM, $\sigma = 0.2$ (e) ATM, $\sigma = 0.2$ (f) ITM, $\sigma = 0.2$

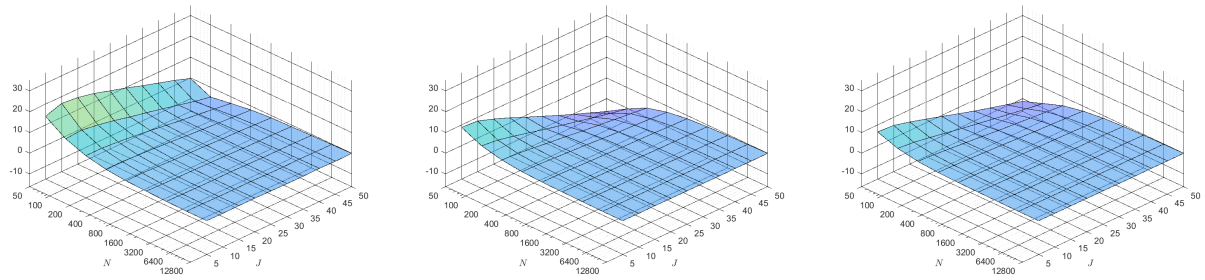
(g) OTM, $\sigma = 0.4$ (h) ATM, $\sigma = 0.4$ (i) ITM, $\sigma = 0.4$

Figure 3.D.1: Relative bias of put option with GBM underlying asset, $T = 1$.

Put option bias as a percentage of the benchmark price for single-asset American call LSM, F-LSM, and FS-LSM price estimators. Results for 1-year OTM, ATM, and ITM options with weekly exercise opportunities (i.e. $J/T = 50$) are reported across asset volatilities $\sigma \in \{0.1, 0.2, 0.4\}$. The left column shows results for OTM options, the center column for ATM options, and the right column for ITM options. The top row shows results for $\sigma = 0.1$, the center row for $\sigma = 0.2$, and the bottom row for $\sigma = 0.4$. A cubic approximation of the exercise strategy (i.e. $L = 3$) is used to compute estimates and benchmark prices.



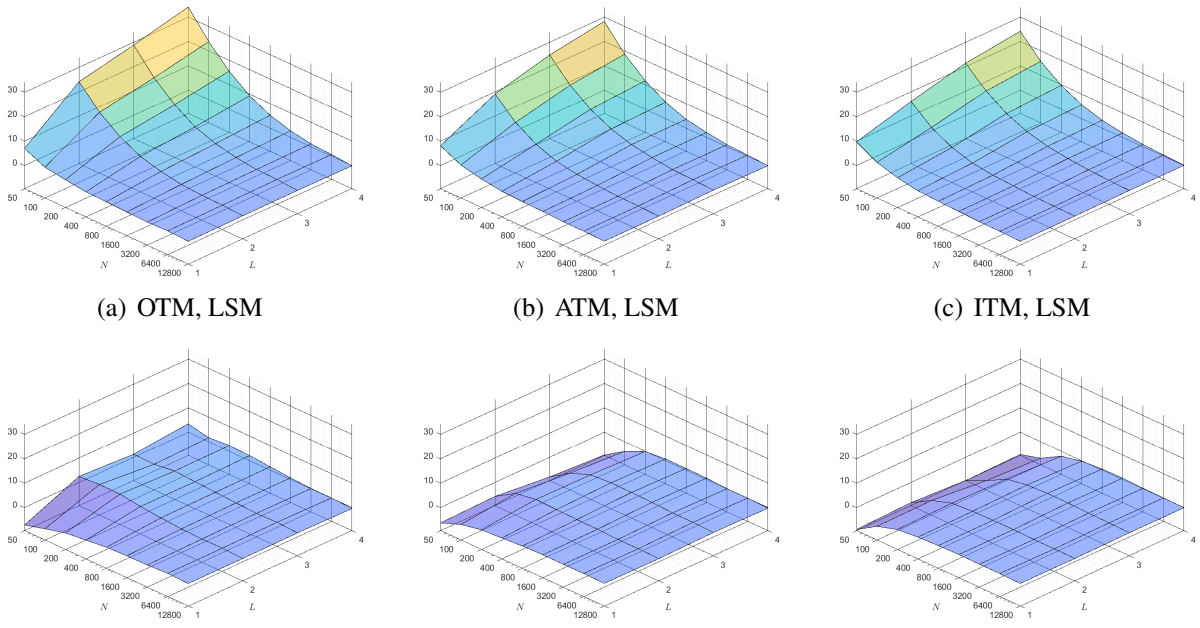
(a) OTM, LSM (b) ATM, LSM (c) ITM, LSM



(d) OTM, FS-LSM (e) ATM, FS-LSM (f) ITM, FS-LSM

Figure 3.D.2: Relative bias of five-asset max-call for varying N and J/T.

Bias as a percentage of the benchmark price for five-asset American max-call LSM and FS-LSM price estimators. Results for one-year OTM, ATM, and ITM options with a cubic approximation of the early-exercise strategy ($L = 3$), and asset price volatilities of $\sigma = 0.2$, are reported across the number of simulated paths $N \in \{50, 100, 200, 400, 800, 1600, 3200, 6400, 12800\}$ and exercise opportunity frequency $J/T \in \{5, 10, 15, \ldots, 50\}$. The left column shows results for OTM options, the center column for ATM options, and the right column for ITM options. The top row shows results for the LSM estimator and the bottom row for the FS-LSM. For every option, the benchmark is computed with the corresponding exercise opportunity frequency J/T.



(a) OTM, LSM (b) ATM, LSM (c) ITM, LSM

(d) OTM, FS-LSM (e) ATM, FS-LSM (f) ITM, FS-LSM

Figure 3.D.3: Relative bias of single-asset call with JDM underlying for varying N and L.

Bias as a percentage of the benchmark price for single-asset American call LSM and FS-LSM price estimators. Results for one-year OTM, ATM, and ITM options with weekly exercise opportunities (i.e. $J/T = 50$), and asset price volatilities of $\sigma = 0.2$, are reported across the number of simulated paths $N \in \{50, 100, 200, 400, 800, 1600, 3200, 6400, 12800\}$ and polynomial order $L \in \{1, \ldots, 4\}$. The compound Poisson process in parameterized by $\lambda = 1, a = 0, b = 0.2$. The left column shows results for OTM options, the center column for ATM options, and the right column for ITM options. The top row shows results for the LSM estimator and the bottom row for the FS-LSM. For every option, the benchmark is computed with the corresponding polynomial order L.

Chapter 4

Monte Carlo Variance Reduction and American Option Exercise Strategies

4.1 Introduction

Pricing American options in a Monte Carlo application involves the implementation of a dynamic program where a series of conditional expectations is estimated to determine an early-exercise strategy (Tilley 1993). Regression-based estimators approximate these conditional expectations with Ordinary Least-Squares (OLS). This class of estimators is flexible and easy to implement where deterministic numerical methods are not. Indeed, lattice- and tree-based solutions (e.g., finite differences, multinomial trees), though very accurate, quickly succumb to the curse of dimensionality when the diffusion of the underlying asset is governed by multiple stochastic factors or when the payoff is path-dependent. In contrast, regressions easily scale to multivariate settings and are well-adapted to general option payoff structures. Among other regression-based approaches like those of Carriere (1996) and Tsitsiklis and Van Roy (2001), one of the most popular is the Least-Squares Monte Carlo (LSM) algorithm introduced by Longstaff and Schwartz (2001) because of its robustness and accuracy (Stentoft 2004*b*, 2014). However, implementing the LSM algorithm can still be computationally costly, and its optimization remains an active area of research.

The main difficulty of implementing the LSM (or any dynamic program in a simulation framework) is that the option value process and the continuation values are jointly estimated, allowing pricing errors to be propagated and amplified throughout the backward recursion of the dynamic program. Indeed, in any actual application, a sub-optimal early-exercise strategy generates pricing errors that compromise the accuracy of subsequent continuation value estimates, and further exacerbate the issue of exercise errors. Hence, the bias and variance that result from this feedback effect can only be dealt with by focusing on exercise optimality via the accuracy of the estimated strategy.

To analyze the variance of LSM estimators, this paper discerns two variance factors: sampling variance, and exercise errors. First, sampling variance is the variance of discounted option cash flows when the exercise times are known. In other words, the LSM sampling variance is the variance of discounted option cash flows sampled at optimal stopping times. Second, exercise errors alter the variance of the estimator due to the uncertainty of estimated stopping times.

As discussed above, a sub-optimal strategy leads to random exercise errors, meaning that the option can be exercised prematurely or late compared to what is recommended by an optimal strategy. Note that exercise errors are absent in standard Monte Carlo integration problems like the valuation of European-style options because the exercise time is predetermined. It follows that the variance of European option price estimates is entirely composed of sampling variance. Indeed, exercise errors only occur in optimal stochastic control problems that involve the implementation of a dynamic program.

To measure sampling variance and the effect of exercise errors on variance in the LSM, we consider an alternative estimator that uses an optimal early-exercise strategy. To construct this optimal estimator, we replace fitted continuation values with a predetermined exercise boundary recovered from a finite difference (FD) algorithm. The optimal exercise boundary can then be approximated with arbitrary precision, thereby removing all sub-optimal exercise decisions. This method, here called the Finite Difference Monte Carlo (FDM), is obviously not practical because a very precise price estimate is already available from the calculations needed to compute the early-exercise boundary. That being said, the FDM estimator is still a useful benchmark because it eliminates the effect of exercise errors on variance while only retaining sampling variance. The LSM sampling variance is therefore equal to the FDM variance, and the effect of exercise errors on variance is simply measured as the difference between LSM and FDM variances.

Analyzing variance in this way allows us to discuss how different variance reduction techniques take effect. We consider two classes of variance reduction techniques in this paper, and combinations thereof. First, we present the LSM importance sampling technique of Moreni (2003), which operates a uniform change in the drift of asset prices to reduce sampling variance. We then briefly discuss an extension to this technique introduced in Boire et al. (2021*a*). Second, we consider the generalized control variates method of Rasmussen (2005), which can be implemented in two ways. The first control variate method consists in applying optimally sampled control variates to option cash flows. This method, here termed the Controlled Cash Flow (CCF) method, mainly reduces sampling variance because it does not affect the estimated stopping times. The second control variate method proposed by Rasmussen (2005) is to apply control variates to continuation values to reduce the variance of the strategy. This method, called the Controlled Continuation Value (CCV) method, thus focuses on stopping time optimality, and tackles both bias and the effect of exercise errors on variance. Control variates can therefore be used to reduce both sources of variance, depending on the way that they are implemented: the CCF method reduces sampling variance, and the CCV method reduces exercise errors.

The numerical results presented in this paper highlight the many benefits of using a CCV-corrected early-exercise strategy in the LSM. We first find that an LSM estimator with a corrected strategy replicates optimal exercise times with over 90% accuracy instead of 60%, roughly, virtually eliminating bias. Moreover, by suppressing stopping time errors, the CCV method yields estimators whose variance approaches the FDM variance lower bound, meaning that it successfully eliminates stopping time variance. These improvements are all the more remarkable as unbiased American price estimators can be obtained even with a very low number of paths. Therefore, instead of simulating a large number of sample paths and performing regressions to reduce the bias and variance caused by exercise errors, the CCV method permits to carry out several small simulations in parallel to speed up calculations and ease computational

memory management. The CCV approach thus adds a lot of flexibility to practitioners who need to accurately price American options in a timely fashion given a certain computational budget.

Next, we discuss how the CCV-corrected strategy affects the efficiency gains permitted by the CCF method. We find that using a corrected strategy in a standard LSM does not significantly change the variance of the estimator. However, using a corrected strategy with the CCF method reduces the standard error of the estimator by a factor reaching over 230 for short-term in-the-money (ITM) options, compared to 12 with the CCF approach only. The reason why a corrected strategy has such a trivial impact in the standard LSM and such an important one when combined with the CCF method is that the effect of exercise errors on variance is a small fraction of the total variance in a standard implementation of the LSM. However, when one or more variance reduction techniques are implemented to reduce sampling variance, the effect of exercise errors on variance becomes increasingly prominent. When the sampling variance has been reduced by several orders of magnitude by the CCF method, the extra variance due to exercise errors becomes the major source of variance. Efforts are then much better spent on improving the early-exercise strategy, and neglecting to do so severely stifles the efficiency of the CCF method.

Finally, we show that a combination of importance sampling and the CCF method is extremely efficient with a corrected strategy but inefficient with an uncorrected one. For the same reasons as discussed above, the efficiency gains achieved by combining variance reduction techniques with an uncorrected strategy are limited due to exercise errors. Such limitations are extensively documented in Boire et al. (2021*c*), where it is shown that combining antithetic sampling and importance sampling techniques with the CCF method only permits marginal efficiency gains, if any at all. If the practitioner ignores this, the combined implementation of variance reduction techniques can be computationally inefficient in the LSM. This issue is resolved when a corrected strategy is used, decreasing the standard error by a factor reaching 375 for short-term ITM options, compared to 32 for a similar estimator with an uncorrected strategy.

To sum up, this paper contributes to a strand of the literature on the optimization of continuation values in regression-based Monte Carlo estimators. Such optimization techniques include the leave-one-out LSM algorithm of Woo et al. (2018) and the bootstrapped early-exercise strategy of Létourneau and Stentoft (2019), which apply resampling methods (see Friedman et al. (2001)) to reduce the variance of continuation values and estimator bias. Alternatively, the sampling approach can be modified to improve the strategy. The initial state dispersion technique proposed by Rasmussen (2005) also significantly improves the regression goodness-of-fit by artificially dispersing asset price paths used to regress continuation values. This technique is combined to a piecewise linear regression model in Kan et al. (2010) to further improve continuation value estimates. Alternatively, the importance sampling technique of Boire et al. (2021*a*) improves continuation values by performing regressions on shifted asset price paths. To our knowledge, the most efficient and straightforward approach to optimize continuation values is the CCV approach of Rasmussen (2005), and is the one we focus on in this paper. First, we highlight important interactions between estimated stopping times and the performance of estimators that implement variance reduction techniques. Indeed, an accurate exercise strategy presents important advantages. In particular, stopping time optimality eliminates the effect of exercise errors on bias and variance, giving a strong boost to the effi-

ciency of control variates, and rendering the combination of importance sampling and control variates more efficient. Altogether, the reduction in both bias and variance leads to a substantial reduction of the root mean squared error (RMSE) in all cases considered. Second, we demonstrate that the CCV corrected strategy nearly achieves stopping time optimality and thus maximizes the efficiency of variance reduction techniques. Indeed, since the estimated strategy is practically optimal with a corrected strategy, we do not recommend further optimization of continuation values. Our final recommendation is therefore to use a corrected strategy when implementing multiple variance reduction techniques, and especially when implementing the CCF method.

The rest of the paper is organized as follows. Section 2 presents the American option pricing problem and the aforementioned LSM and FDM algorithms. Section 3 discusses LSM estimator variance and details individual and combined implementations of importance sampling and control variate techniques. Section 4 discusses the results. Section 5 concludes.

4.2 Pricing American options

In the following section, we first state the valuation problem associated with pricing American options. Next, we show how American option prices can be approximated in a discrete formulation with the LSM algorithm of Longstaff and Schwartz (2001). Finally, we provide details about the implementation of an FDM estimator that uses an arbitrarily precise early-exercise strategy recovered from a deterministic FD algorithm.

4.2.1 The valuation problem

On a continuous filtered probability space $(\Omega, \mathcal{F}, \mathbb{Q})$ equipped with the filtration $\mathbb{F} = \{\mathcal{F}_t\}_{0 \leq t \leq T}$, the value of an American option is the solution of an optimal stochastic control problem where the expected discounted payoff from exercising the option is estimated over the class of $\mathbb{F}$-adapted stopping times. Omitting discount factors for ease of exposition, the discounted value of an American option can be written as

$$V_t = \sup_{\tau \in \mathcal{T}(t,T)} \mathbb{E}_t \left[X_\tau \right], \tag{4.1}$$

where $\{X_t\}_{t \geq 0}$ is a non-negative payoff process related to the underlying asset $\{S_t\}_{t \geq 0}$ and $\mathcal{T}(t, T) \subseteq [t, T]$ is a class of adapted stopping times. Furthermore, we assume an adequately chosen numéraire satisfying the martingale property of discounted option prices under the risk-neutral measure $\mathbb{Q}$. An $\mathbb{F}$-adapted stopping time can be seen as the first time τ^* when the payoff process X_t is in the exercise region $\mathcal{E}_t$, written as

$$\tau^* = \inf_{\tau \in \mathcal{T}(t,T)} \{\tau : X_\tau \in \mathcal{E}_\tau\}, \tag{4.2}$$

where $\mathcal{E}_t \in \mathcal{F}_t$ for all $t \geq 0$.

Let us now consider a discrete-time formulation of the valuation problem. For an option with a maturity of T years, we consider an evenly spaced partition of time with J steps of length $\Delta t = T/J$. The discretized price process of the underlying asset $\{S_j : j = 0, \ldots, J\}$ is defined

on a discrete filtered probability space $(\Omega, \mathcal{G}, \mathbb{Q})$ and adapted to the filtration $\mathbb{G} = \{\mathcal{G}_j\}_{j=0}^{J}$. For simplicity, we use time j to refer to time $\tau_j = j\Delta t$. Option exercise is allowed only at the points defining the time discretization, hence the set of admissible exercise opportunities is $\{\tau_j = j\Delta t : j = 0, \ldots, J\}$, where $0 = \tau_0 < \tau_1 < \cdots < \tau_J = T$. We slightly adjust the above notation such that V_j denotes the time-j value of an unexercised option with corresponding underlying asset value S_j, the option price can be written as V_0, which solves the following dynamic program

$$\begin{cases} V_J = X_J \\ V_j = \max\left(X_j, \mathbb{E}_j[V_{j+1}]\right), \quad j = J-1, \ldots, 0. \end{cases} \tag{4.3}$$

The conditional expectation $\mathbb{E}_j[V_{j+1}]$ is called the continuation value and can be interpreted as the expected future value of the option conditional on keeping the option alive until the next period.

Finally, we note that options with a discrete early-exercise feature are technically termed Bermudan options. Because true Bermudan option prices approximate true American option prices by letting the exercise frequency J/T tend to infinity (Glasserman 2003), it is common to approximate American option prices with the LSM even though they are not equivalent. This approximation is not entirely faulty, however, since, in practice, option holders do not continuously consider exercising the option. Rather, they are more likely to reassess their position on a daily, or weekly basis. It is then reasonable to value American options as Bermudan options with a user-chosen frequency of admissible exercise times that reflects how the option will be managed. In this paper, as is common in the literature, we describe the options as American rather than Bermudan. The benchmark prices used are also Bermudan option prices to factor out discretization bias (i.e., the pricing gap between American and Bermudan options) in the reported bias results.

4.2.2 Least-Squares Monte Carlo

We now present the Least-Squares Monte Carlo (LSM) algorithm of Longstaff and Schwartz (2001), which approximates the dynamic program in Equation (4.3) in the discrete pricing framework. Using a risk-neutral valuation approach, we consider the discrete filtered probability space $(\Omega, \mathcal{G}, \mathbb{Q})$ equipped with the discrete filtration $\mathbb{G}$. Given a set of N simulated paths, the path-n asset price at time j is denoted $S_{n,j}$ for all $n = 1, \ldots, N$ and $j = 0, \ldots, J$. Corresponding option payoff and estimated option value are denoted $X_{n,j}$ and $\hat{V}_{n,j}$, respectively.

All least-squares estimators, and the LSM in particular, use OLS principles to estimate a cross-section of continuation values at each time step $j = 0, \ldots, J-1$ and form a criterion for early exercise. Continuation values are estimated by regressing future option values on a finite set of $\mathcal{G}$-measurable basis functions $\{\psi_\ell(\cdot) : \ell = 0, \ldots, L\}$ related to asset prices. That is, continuation values are modeled as

$$\hat{V}_{n,j+1} = \sum_{\ell=0}^{L} \psi_l(S_{n,j})\beta_{j,\ell} + \varepsilon_{n,j+1}, \tag{4.4}$$

where the path-n, time-j error term $\varepsilon_{n,j}$ satisfies the usual OLS assumptions. The path-n, time-j regression fitted value, denoted by $\hat{C}_{n,j}$, approximates the (risk-neutral) conditional expectation

$\mathbb{E}_j[\hat{V}_{n,j+1}]$ used to solve the exact dynamic program in Equation (4.3). Letting $\mathcal{N}_j = \{n : X_{n,j} > 0\}$ be the set of time-j ITM paths and $N_j = |\mathcal{N}_j|$ be the number of such paths, the LSM algorithm computes the $(L+1) \times 1$ vector of coefficient estimates $\hat{\beta}_j = \{\hat{\beta}_{j,0}, \ldots, \hat{\beta}_{j,L}\}$ by regressing discounted cashflows $\hat{V}_{n,j+1}$ against the $1 \times (L+1)$ basis functions $\psi(S_{n,j}) = \{\psi_0, \psi_1(S_{n,j}), \ldots, \psi_L(S_{n,j})\}$, with ψ_0 as a constant, and for $n \in \mathcal{N}_j$. One can then easily show that the least-squares solution is

$$\hat{\beta}_j = \left(\Psi_j' \Psi_j\right)^{-1} \Psi_j' \hat{\mathbf{V}}_{j+1}, \tag{4.5}$$

where the $N_j \times (L+1)$ matrix $\Psi_j = (\psi(S_{n,j}) : n \in \mathcal{N}_j)$ denotes the time-j cross-section of basis functions for ITM paths and the $N_j \times 1$ vector $\hat{\mathbf{V}}_{j+1} = (\hat{V}_{n,j+1} : n \in \mathcal{N}_j)$ is the corresponding sample of discounted option cashflows at time $j+1$. The path-n, time-j fitted continuation value then takes the form

$$\hat{C}_{n,j} = \psi(S_{n,j})\hat{\beta}_j, \tag{4.6}$$

and determines an exercise strategy in which the option is exercised if the payoff is greater than the fitted continuation value and the option is ITM. Using these exercise criteria, the LSM dynamic program is written as

$$\begin{cases} \hat{V}_{n,J} = X_{n,J} \\ \hat{V}_{n,j} = \begin{cases} X_{n,j} & \text{if } (X_{n,j} \geq \hat{C}_{n,j}) \text{ and } \left(X_{n,j} > 0\right) \\ \hat{V}_{n,j+1} & \text{if } (X_{n,j} < \hat{C}_{n,j}) \text{ or } \left(X_{n,j} = 0\right) \end{cases} \quad j = J-1, \ldots, 1. \end{cases} \tag{4.7}$$

The sample size-N LSM estimator is then defined as

$$\hat{V}_0^{(N)} = \frac{1}{N} \sum_{n=0}^{N} \hat{V}_{n,0} = \frac{1}{N} \sum_{n=0}^{N} X_{n,k_n} \tag{4.8}$$

where $\tau_{k_n} = k_n \Delta t$ is the path-n stopping time from the LSM exercise strategy and X_{n,k_n} is the corresponding discounted payoff. By convention, the stopping time determined by the LSM algorithm (and all other algorithms discussed hereafter) along an unexercised path is $\tau_J = T$.

The key element to obtain consistent American option prices is an accurate estimation of the continuation value. Estimating the continuation value is at the heart of any numerical approach to American option pricing. If the continuation value is known, the dynamic program can be solved exactly. However, true continuation values belong to the set of Hilbert spaces and are characterized by a countably infinite set of regression parameters (Royden 1988). Consequently, finite-order polynomial approximations are misspecified and thus inherently sub-optimal even when optimally parameterized. This negative bias, also called the approximation bias, is present in any actual application of the LSM, and can only be mitigated by increasing the polynomial order of the approximation. Furthermore, it is important to note that the LSM uses a strategy that is not restricted to being adapted to the filtration. Indeed, the optimal strategy at time j is $\mathcal{G}_j$-measurable, whereas the LSM strategy is $\mathcal{G}_J$-measurable for all $j = 0, \ldots, J-1$ and therefore has the potential to outperform the optimal strategy. This problem arises when continuation value estimates are overfitted, thereby providing foresight to the strategy and generating positive bias. Therefore, we discern two conflicting bias elements which manifest themselves to varying degrees in any LSM implementation: negative sub-optimality bias and positive foresight bias. Although it may seem desirable that two bias elements offset each other, this makes the sign of the LSM estimator bias ambiguous in finite samples.

4.2.3 Finite Difference Monte Carlo

The FDM algorithm is equivalent to the LSM algorithm in Equation (4.7), except it uses the strategy given by a FD algorithm. Instead of estimating continuation values, the FD strategy takes the form of a deterministic boundary for underlying asset prices, and is therefore referred to as the early-exercise boundary. The FD algorithm employs a dynamic programming representation over a deterministic partition of the state of the underlying asset price. The price and the strategy for the American option can then be made arbitrarily precise with a sufficiently fine grid of time and asset price. To compute the FD strategy, we consider an algorithm in which the set of admissible stopping times are the same times $\tau_j : j = 1, \dots, J$ as in the LSM. That is, if time is evenly partition over $mJ : m \in \mathbb{N}$ even time steps defined by $0 = t_0 \leq t_1 \leq \cdots \leq t_{mJ} = T$, exercise is only considered at times $t_{mj} : j = 0, \dots, J$. Indeed, instead of allowing exercise at each point of the time partition to mimic a continuously exercisable American option, the FD algorithm retains the Bermudan feature inherent to the aforementionned discrete simulation scheme. Denoting the time-j discounted option value process as $\hat{V}^*_j$, and the time-j FD boundary as b^*_j, the FDM dynamic program can be written

$$\begin{cases} \hat{V}^*_{n,J} = X_{n,J} \\ \hat{V}^*_{n,j} = \begin{cases} X_{n,j} & \text{if } S_{n,j} \geq b^*_j \\ \hat{V}^*_{n,j+1} & \text{if } S_{n,j} < b^*_j \end{cases} \quad j = J-1, \dots, 1. \end{cases} \tag{4.9}$$

Here, we note that since the optimal early-exercise stratetgy is known *a priori*, a dynamic program is not really needed, and the same FDM estimator can be obtained in a forward Monte Carlo simulation approach. In a sense, the predetermined FDM strategy removes the complicating issue of optimal stochastic control, and the estimator is simply obtained by averaging discounted option payoffs as in the European case. The sample-size N FDM estimator is then defined as

$$\hat{V}^{*(N)}_0 = \frac{1}{N} \sum_{n=0}^{N} \hat{V}^*_{n,0} = \frac{1}{N} \sum_{n=0}^{N} X_{n,k^*_n}, \tag{4.10}$$

where $\tau_{k^*_n} = k^*_n \Delta t$ is the path-n stopping time from the FD boundary and X_{n,k^*_n} is the corresponding discounted payoff.

The FDM is an important benchmark in the following analyses because it removes the bias and variance due to stopping time errors. Comparing price estimates to this benchmark therefore solely measures the pricing error. Furthermore, the variance of the FDM can be seen as an approximation of the variance of an estimator with an optimal exercise strategy. Comparing FDM and LSM variances therefore measures the effect of stopping time errors on the variance of LSM estimator. Furthermore, assuming that exercise errors lead to an increase in variance (which is the case in a great majority of the cases studied below), the FDM variance is a lower-bound for the LSM variance. By corollary, the efficiency factor of a given variance reduction technique in the FDM is a rough upper-bound to the efficiency factor in the LSM. Equivalently, we can then say that the effect of exercise errors on LSM variance is minimized when its efficiency factor coincides with the FDM efficiency factor.

4.3 Variance reduction techniques

In the following section we first briefly analyze the variance of the LSM by clearly defining sampling variance and the additional effect of exercise errors. Second, we discuss the implementation of variance reduction techniques and discuss their impact on variance. Specifically, we first present the generalized control variates of Rasmussen (2005). As mentioned above, control variates can be applied in two different ways in least-squares algorithms. They can be applied to the option cash flow at the estimated stopping time, or they can be applied to the continuation value estimates every time a regression is performed. Third, we present the American option importance sampling technique of Moreni (2003). Finally, we detail individual and combined implementations of all the variance reduction techniques presented herein.

4.3.1 Analysis of LSM variance

We now propose a decomposition of the variance of the LSM estimator to provide some intuition about the mechanisms of variance reduction techniques. We begin by introducing a bit of notation. Briefly dropping the path indices for notational simplicity, let the option value computed from the LSM program along a given path be denoted by the operator $v(\cdot)$, taking for argument the parameters that characterize the early-exercise strategy. The standard LSM estimator, for instance, is written $v(\hat{\boldsymbol{\beta}}) = X_k$ because it uses the regression parameters $\hat{\boldsymbol{\beta}} = \{\hat{\beta}_j : j = 0, \ldots, J-1\}$. Similarly, $v(\boldsymbol{\beta}) = X_{\bar{k}}$ is the estimated option value along that same path, but with an optimal finite parameterization of the early-exercise strategy $\boldsymbol{\beta} = \{\beta_j : j = 0, \ldots, J-1\}$. In this case, the LSM option value is estimated with the optimal stopping times given a finitely parameterized strategy, i.e., $\tau_{\bar{k}} = \bar{k}\Delta t$. The LSM estimator using the true optimal strategy is denoted $\mathcal{V}$. We then decompose the LSM estimated option value into

$$\begin{aligned} v(\hat{\boldsymbol{\beta}}) &= \mathcal{V} - \mathcal{V} + v(\boldsymbol{\beta}) - v(\boldsymbol{\beta}) + v(\hat{\boldsymbol{\beta}}) \\ &= \mathcal{V} + a(\boldsymbol{\beta}) + b(\hat{\boldsymbol{\beta}}, \boldsymbol{\beta}), \end{aligned} \tag{4.11}$$

where $a(\boldsymbol{\beta}) = v(\boldsymbol{\beta}) - \mathcal{V} = X_{\bar{k}} - X_{k^*}$ is the payoff difference between an LSM estimator using the optimal finite parameterization of the early-exercise strategy and an LSM estimator using the true optimal strategy. The expected value of $a(\boldsymbol{\beta})$ is then the negative approximation error due to the finite parameterization of true optimal strategy. Next, $b(\hat{\boldsymbol{\beta}}, \boldsymbol{\beta}) = v(\hat{\boldsymbol{\beta}}) - v(\boldsymbol{\beta}) = X_k - X_{\bar{k}}$ is the payoff difference between an LSM estimator using an estimated strategy and an LSM estimator using the optimal finite parameterization. It is easy to see that $b(\hat{\boldsymbol{\beta}}, \boldsymbol{\beta})$ is a random variable whose entire distribution depends on how closely $\hat{\boldsymbol{\beta}}$ approximates $\boldsymbol{\beta}$. The expected value of $b(\hat{\boldsymbol{\beta}}, \boldsymbol{\beta})$ is the estimation bias, or simply the bias of the LSM with respect to $v(\boldsymbol{\beta})$, and thus mixes both positive and negative elements. Indeed, a poor estimate of $\boldsymbol{\beta}$ leads to sub-optimal exercise decisions and exacerbates the negative bias of the LSM estimator. Furthermore, overfitted continuation values provides foresight to the early-exercise strategy, leading to supra-optimal exercise decisions and positive bias.

We then write the LSM variance in terms of $b(\hat{\boldsymbol{\beta}}, \boldsymbol{\beta})$ and $v(\boldsymbol{\beta})$ as

$$\mathrm{Var}[v(\hat{\boldsymbol{\beta}})] = \underbrace{\mathrm{Var}\left[v(\boldsymbol{\beta})\right]}_{\text{Sampling variance}} + \underbrace{\mathrm{Var}[b(\hat{\boldsymbol{\beta}}, \boldsymbol{\beta})] + 2\mathrm{Cov}[v(\boldsymbol{\beta}), b(\hat{\boldsymbol{\beta}}, \boldsymbol{\beta})]}_{\text{Effect of exercise errors on variance}}. \tag{4.12}$$

First, the variance of $v(\boldsymbol{\beta})$ is the sampling variance, which is present in all Monte Carlo applications. In the LSM algorithm the sampling distribution is parameterized by the strategy parameter $\boldsymbol{\beta}$ optimized over the user-chosen finite parametric space $\mathcal{B}$ adapted to $\mathbb{G}$. Therefore, we can say that sampling variance is determined by the parametric space $\mathcal{B}$ used to approximate the true optimal strategy. In the discussion to follow, the sampling variance is approximated by the variance of the FDM algorithm, though they are in fact slightly different. This is because the LSM price with regression coefficients $\{\hat{\beta}_j : j = 0 : J-1\}$ estimates the price with the true regression coefficients $\{\beta_j : j = 0 : J-1\}$. In contrast, the FDM is not restricted to a finitely parameterized strategy, but by a user-chosen lattice. The two methods thus exhibit different approximation biases such that the FDM variance slightly differs from the true sampling variance of a regression-based estimator.

Second, the effect of exercise errors on variance is the sum of the second and third terms on the right-hand side of Equation (4.12). The second term is the variance of $b(\hat{\boldsymbol{\beta}}, \boldsymbol{\beta})$. It is plain to see that $b(\hat{\boldsymbol{\beta}}, \boldsymbol{\beta})$ is affected by the error of the early-exercise strategy at every single time step. Indeed, due to the backward-recursive nature of the dynamic program, estimated cashflows $\{X_j : j = 0, \ldots, J\}$ and regression coefficient estimates $\{\hat{\beta}_j : j = 0, \ldots, J-1\}$ are all $\mathcal{G}_J$-dependent random variables. Therefore, the variance of $b(\hat{\boldsymbol{\beta}}, \boldsymbol{\beta})$ depends on the overall accuracy of the strategy. The third term is twice the covariance between $b(\hat{\boldsymbol{\beta}}, \boldsymbol{\beta})$ and $v(\boldsymbol{\beta})$. In the ideal case, the estimated strategy is nearly optimal such that $\hat{\boldsymbol{\beta}}$ approaches $\boldsymbol{\beta}$ and $b(\hat{\boldsymbol{\beta}}, \boldsymbol{\beta})$ tends to zero. If this assumption holds, the effect of exercise errors on variance is then simply equal to the variance of $b(\hat{\boldsymbol{\beta}}, \boldsymbol{\beta})$ and is positive.

In the following Sections 4.3.2 and 4.3.3, the variance decomposition in Equation (4.12) allows us to more precisely discuss the effect of importance sampling and control variates on the LSM variance. In LSM applications, sampling variance is commonly mitigated by standard variance reduction techniques, like antithetic sampling (Longstaff and Schwartz 2001), importance sampling (Moreni 2003) with change of measure optimization (Glasserman et al. 1999, Bolia and Juneja 2005, Morales 2006, Juneja and Kalra 2009), control variates (Tian and Burrage 2002, Rasmussen 2005), and low-discrepancy sequences (Lemieux and La 2005), to name only a few. All of these techniques affect the sampling distribution of option cash flows (and can be combined) to further reduce sampling variance. However, they do not significantly improve the early-exercise strategy and thus do nothing to mitigate the effects of exercise errors. The most straightforward way to tackle exercise errors is to increase the number of simulated paths to better approximate the optimal strategy. Another valid approach consists in imposing convexity constraints to the estimated continuation value models (Létourneau and Stentoft 2014). Alternatively, when several LSM programs are running in parallel, one can also employ bootstrap aggregation of regression coefficients (Létourneau and Stentoft 2019). Rasmussen (2005) also proposes to use initially state-dispersed simulation paths to improve the fit of the regression near the optimal exercise barrier. In the context of importance sampling Boire et al. (2021*a*) suggest performing regressions on the shifted paths to improve the regression fit and reduce estimator bias. Finally, Rasmussen (2005) applies control variates to fitted continuation values to obtain a nearly optimal early-exercise strategy. This CCV method works by improving the accuracy of exercise times and very effectively eliminates their effect on variance. To our knowledge, the CCV method is the most effective way to improve stopping times and is therefore the approach we use in this paper to boost the efficiency of other variance reduction techniques.

4.3.2 Control variates

A very effective and flexible Monte Carlo variance reduction technique is control variates. The goal of control variates is to estimate and correct the error of an estimator by exploiting our knowledge of the errors of another "nearby" estimator. Indeed, if two Monte Carlo variates are strongly correlated, then so are their errors. In turn, we can reduce the unknown error of an estimator by subtracting the known error of another unbiased estimator multiplied by an adequate coefficient, called the control parameter. The implementation of control variates therefore requires one to choose an unbiased Monte Carlo estimator that is strongly correlated with the estimator of interest, and whose expected value is known or can be computed efficiently. To implement control variates in the LSM, a natural approach is then to base control variates on the price process of a European option that replicates the payoff of the American option at maturity, and whose price is available in closed-form (Rasmussen 2005).

As stated above, we present two approaches to the implementation of control variates in the LSM. First, in the CCF approach, control variates are applied to the estimated option cash flows and can be implemented in any Monte Carlo option pricing application. Second, in the CCV approach, control variates are applied to the continuation values estimated in the dynamic program. This technique is therefore particular to least-squares estimators for options that feature early-exercise opportunities.

Control variates applied to option cash flows

We now focus on the first implementation, which involves a correction of the estimated option cash flows. The method of controlled cash flows, here termed the CCF method, is the well-known control variates method used in general simulation-based derivatives pricing applications and effectively corrects the sampling variance. The path-n, time-j control variate we consider here is defined as $(Y_{n,j} - Y_{n,0})$, where $Y_{n,j} = \mathbb{E}_j[X_{n,J}]$ is the path-n, time-j discounted value of a portfolio that replicates the payoff of the American option at maturity. Therefore, $Y_{n,j}$ is the value of a European option with initial asset price $S_{n,j}$ and remaining time to maturity equal to $T - \tau_j$ discounted back to time $\tau_0 = 0$. Similarly, $Y_{n,0} = \mathbb{E}[X_{n,J}]$ is the value of a European option with initial asset price $S_{n,0} = S_0$ and time to maturity equal to T. The control variates are then easy to compute using the Black-Scholes pricing formula. Finally, by virtue of the martingale property of discounted option prices, we can write $Y_{n,0} = \mathbb{E}[Y_{n,j}]$. It immediately follows that $\mathbb{E}[Y_{n,j} - Y_{n,0}] = 0$, and the CCF method does not alter the bias of the estimator.

To apply the CCF method optimally, one must first determine the stopping time estimated by the dynamic program along each path. Indeed, Rasmussen (2005) shows that a control variate sampled at the estimated stopping time Y_{n,k_n} minimizes the variance of the estimator. Next, assuming without loss of generality that the stopping times $\{\tau_{k_n} = k_n \Delta t : n = 1, \ldots, N\}$ were estimated by the dynamic program in Equation (4.7) with corresponding option cash flows $\{X_{n,k_n} : n = 1, \ldots, N\}$, we compute controlled discounted cash flows as $Z_{n,k_n}(\theta) = X_{n,k_n} + \theta(Y_{n,k_n} - Y_{n,0})$ for all $n = 1, \ldots, N$, where θ is the control parameter. It is easy to show that the optimal control parameter is $\theta^* = -\mathrm{Cov}[X_{n,k_n}, Y_{n,k_n}]/\mathrm{Var}[Y_{n,k_n}]$. We therefore use an approximation of the variance-minimizing control parameter in our experiments, which can be estimated with sample quantities as $\hat{\theta} = -\sum_{n=1}^{N}((X_{n,k_n} - \bar{X}_k)(Y_{n,k_n} - \bar{Y}_k))/\sum_{n=1}^{N}(Y_{n,k_n} - \bar{Y}_k)^2$, where $\bar{X}_k = N^{-1}\sum_{n=1}^{N} X_{n,k_n}$

and $\bar{Y}_k = N^{-1} \sum_{n=1}^{N} Y_{n,k_n}$.

Finally, the size-N LSM-CCF estimator with control variates is written as

$$\hat{V}_0^{C(N)} = \frac{1}{N} \sum_{n=1}^{N} Z_{n,k_n}(\hat{\theta}), \tag{4.13}$$

Applying control variates to the LSM option cash flows therefore boils down to averaging the corrected LSM cash flows $Z_{n,k_n}(\hat{\theta})$ as in Equation (4.13) instead of averaging the LSM cash flow X_{n,k_n} as in Equation (4.8). Similarly, the sample size-N FDM-CCF estimator with control variates is written as

$$\hat{V}_0^{*C(N)} = \frac{1}{N} \sum_{n=0}^{N} Z_{n,\tau_{k_n^*}}(\hat{\theta}^*), \tag{4.14}$$

where $\tau_{k_n^*} = k_n^* \Delta t$ is the path-n stopping time estimated by the FDM and $\hat{\theta}^* = -\sum_{n=1}^{N}((X_{n,k_n^*} - \bar{X}_{k^*})(Y_{n,k_n^*} - \bar{Y}_{k^*}))/\sum_{n=1}^{N}(Y_{n,k_n^*} - \bar{Y}_{k^*})^2$, where $\bar{X}_{k^*} = N^{-1}\sum_{n=1}^{N} X_{n,k_n^*}$ and $\bar{Y}_{k^*} = N^{-1}\sum_{n=1}^{N} Y_{n,k_n^*}$.

It is apparent that control variates applied to option cash flows do not affect the stopping times. In fact, CCF control variates only come into play after stopping times have been estimated when averaging the discounted option payoffs. This type of implementation of control variates therefore thus only affects the sampling distribution of option cash flows.

Control variates applied to continuation values

The second implementation of control variates applies a correction to the cross-section of estimated continuation values at every early-exercise opportunity. The method of controlled continuation values, here called the CCV method, reduces the variance of the continuation value, yielding a more accurate early-exercise strategy. Because we are modifying continuation values at each time step, we now introduce a different dynamic program called the LSM-CCV, written as

$$\begin{cases} \hat{W}_{n,J} = X_{n,J} \\ \hat{W}_{n,j} = \begin{cases} X_{n,J} & \text{if } (X_{n,J} \geq \hat{D}_{n,j}(\hat{\vartheta}_{n,j})) \text{ and } \left(X_{n,j} > 0\right) \\ \hat{W}_{n,j+1} & \text{if } (X_{n,J} < \hat{D}_{n,j}(\hat{\vartheta}_{n,j})) \text{ or } \left(X_{n,j} = 0\right) \end{cases} \quad j = J-1, \ldots, 1, \end{cases} \tag{4.15}$$

where the path-n, time-j estimated option value is expressed as $\hat{W}_{n,j}$. Notice that the continuation value $\hat{C}_{n,j}$ in the algorithm in Equation (4.7) is now replaced with the controlled continuation values $\hat{D}_{n,j}(\hat{\vartheta}_{n,j})$ written as

$$\hat{D}_{n,j}(\vartheta_{n,j}) = \psi(S_{n,j})\hat{\phi}_j + \vartheta_{n,j}\left(\psi(S_{n,j})\hat{\rho}_j - Y_{n,j}\right), \tag{4.16}$$

where $\vartheta_{n,j}$ is the path-n, time-j control parameter. The CCV parameters $\hat{\phi}_j$ and $\hat{\rho}_j$ are regression coefficients computed as

$$\hat{\phi}_j = \left(\Psi_j'\Psi_j\right)^{-1}\Psi_j'\hat{\mathbf{W}}_{j+1} \tag{4.17}$$

$$\hat{\rho}_j = \left(\Psi_j'\Psi_j\right)^{-1}\Psi_j'\mathbf{Y}_j, \tag{4.18}$$

where the $N_j \times 1$ vector $\hat{\mathbf{W}}_{j+1} = (X_{n,s_n} : n \in \mathcal{N}_j)$ is the time-$(j+1)$ cross-section of discounted option cashflows at time $j+1$ for all $n \in \mathcal{N}_j$, and $\mathbf{Y}_j = (Y_{n,s_n} : n \in \mathcal{N}_j)$ is the $N_j \times 1$ vector of European option prices where we let $\tau_{s_n} = s_n \Delta t : s_n > j$ be the first estimated stopping time after time-j along path n. Specifically, if the LSM algorithm has recorded an exercise time along path n, we set $s_n = \min\{j+1, \ldots, J : (X_{n,j} \geq \hat{D}_{n,j}(\hat{\vartheta}_{n,j}))$ and $\left(X_{n,j} > 0\right)\}$. Otherwise, we simply set $s_n = J$. The control variate $\psi(S_{n,j})\hat{\rho}_j$ is therefore the least-squares fit of the European option prices sampled at the next LSM stopping time, which approximates $\mathbb{E}[Y_{n,s_n}|\psi(S_{n,j})]$. It is easy to see that the control variate $\psi(S_{n,j})\hat{\rho}_j - Y_{n,j}$ once again has zero bias and therefore the continuation values $\hat{D}_{n,j}(\vartheta_{n,j})$ are unbiased estimators of $\hat{W}_{n,j+1}$. As outlined in Section 5 of Rasmussen (2005), the path-n, time-j variance-minimizing control parameter $\vartheta_{n,j}$ is estimated as

$$\hat{\vartheta}_{n,j} = \frac{\psi(S_{n,j})\left(\Psi_j'\Psi_j\right)^{-1}\Psi_j'(\hat{\mathbf{W}}_{j+1} \odot \mathbf{Y}_j) - (\psi(S_{n,j})\hat{\phi}_j \cdot \psi(S_{n,j})\hat{\rho}_j)}{\psi(S_{n,j})\left(\Psi_j'\Psi_j\right)^{-1}\Psi_j'(\mathbf{Y}_j \odot \mathbf{Y}_j) - (\psi(S_{n,j})\hat{\rho}_j)^2}, \tag{4.19}$$

where the operator $\odot$ is the element-wise product. Notice that the implementation of CCV therefore requires one to perform three additional regressions at each time step: one to compute the control variate, and two more to estimate the variance minimizing control parameter. Finally, the sample size-N LSM-CCV estimator is defined as

$$\hat{W}_0^{(N)} = \frac{1}{N}\sum_{n=0}^{N} \hat{W}_{n,0}^{(N)} = \frac{1}{N}\sum_{n=0}^{N} X_{n,k_n^C} \tag{4.20}$$

where $\tau_{k_n^C}$ is the path-n stopping time estimated in the LSM-CCV, and X_{n,k_n^C} is the corresponding discounted payoff. As suggested by Equation (4.20), the implementation of CCV does not change the optimal sampling distribution of cash flows because the estimator is still constructed by averaging draws from $X_{n,j}$. Rather, it reduces the effect of stopping time errors on the variance of the estimator. The purpose of CCV is to reduce the variance of continuation values, leading to a more accurate early-exercise strategy. As a result, CCV largely eliminates the effect of stopping time errors on variance, reducing both the size of the bias and the variance associated with stopping time uncertainty.

Finally, we note that CCF and CCV control variates are only unbiased when stopping times are estimated independently from future option values. If that is the case, the martingale property of discounted European option values ensures that control variates are unbiased. However, the assumption of independence does not hold true in a standard in-sample implementation of the LSM. As a result, the control variates employed in the CCF and CCV methods also suffer from foresight bias. We leave the question of implicit foresight of optimally sampled control variates for future research.

4.3.3 Importance sampling

Importance sampling is commonly used in Monte Carlo applications where relevant events are seldom represented in small samples. Importance sampling works by computing an equivalent expectation under an "importance" probability measure instead of the nominal probability measure to increase the sampling frequency of such events. The modified estimator is then constructed by averaging variates sampled under the importance measure and multiplied by

the corresponding likelihood ratios. In the context of pricing options with very rare or complex payoffs, importance sampling changes the diffusion of underlying assets. By virtue of the Girsanov theorem, the likelihood ratio is then defined as the Radon-Nikodym derivative related to the discretized asset price process. For example, re-directing simulated asset price paths toward the ITM region for an out-of-the-money (OTM) European option can significantly reduce the frequency of null payoffs that do not contribute to the computation of the expectation (Reider 1994). Alternatively, consider the example of a European knock-in option. The change of measure can re-orient asset price paths to increase the likelihood that they will meet the barrier and the strike price.[1] In both cases the role of importance sampling is to increase the sampling rate of rare and large option payoffs. For American options, the change of measure should not only reduce the number of unexercised paths but also increase the frequency of early exercises to provide a better estimate of the early-exercise premium.

In this paper, we focus on asset prices governed by a Geometric Brownian Motion (GBM) as in the Black and Scholes (1973) and Merton (1973) model with risk-free rate r, dividend yield q, and volatility σ. The path-n asset price can then be simulated exactly under the nominal risk-neutral measure $\mathbb{Q}$ as

$$S_{n,j} = S_{n,j-1} \exp\left((r - q - \frac{\sigma^2}{2})\Delta t + \sigma\sqrt{\Delta t} z_{n,j}\right), \tag{4.21}$$

where $Z_n = \{z_{n,j} : j = 1, \ldots, J\}$ is a vector of independent, identically distributed (iid) standard normal increments for all $n = 1, \ldots, N$. The class of importance measures we consider in this paper is parameterized by a uniform change of λ in the drift of the diffusion.[2] To be specific, we consider the situation in which a uniform drift term is added to the normal increments by posing $\tilde{Z}_n = \{z_{n,j} + \sqrt{\Delta t}\lambda : j = 1, \ldots, J\}$ and we denote the importance probability measure as $\tilde{\mathbb{Q}}$. The same normal variates are then used to simulate the GBM under the importance measure $\tilde{\mathbb{Q}}$ with $\tilde{S}_0 = S_0$ and we have

$$\tilde{S}_{n,j} = \tilde{S}_{n,j-1} \exp\left((r - q - \frac{\sigma^2}{2} + \sigma\lambda)\Delta t + \sigma\sqrt{\Delta t} z_{n,j}\right). \tag{4.22}$$

The k-step likelihood ratio for the discretized diffusion process then takes the form

$$\frac{d\mathbb{Q}}{d\tilde{\mathbb{Q}}}(\tilde{S}_{n,j}, k\Delta t) = \exp\left[\frac{\lambda}{\sigma}\left(r - q - \sigma^2/2\right)k\Delta t - \frac{\lambda}{\sigma}\log\left(\frac{\tilde{S}_{n,j}}{\tilde{S}_{n,j-k}}\right) + \frac{\lambda^2 k\Delta t}{2}\right], \quad k = 0, \ldots, j. \tag{4.23}$$

We denote the LSM estimator with importance sampling as LSM-I. The algorithm is then carried out by simulating shifted asset paths under the importance measure $\tilde{\mathbb{Q}}$. Let $\tilde{X}_{n,j}$ be the path-n, time-j discounted option payoff simulated under $\tilde{\mathbb{Q}}$. Furthermore, let $\tilde{\mathcal{N}}_j = \{n : \tilde{X}_{n,j} > 0\}$ be the set of time-j ITM paths simulated under $\tilde{\mathbb{Q}}$ and let $\tilde{N}_j = |\tilde{\mathcal{N}}_j|$. The path-$n$, time-$j$ fitted continuation value is computed with the same regression coefficients as those obtained in the

[1] See Boyle et al. (1997) or Example 4.6.4 of Glasserman (2003) for a detailed implementation with an optimal exponential change of measure

[2] Naturally, a more general class of distributions can be considered. However, optimizing the parameters of an importance measure in a dynamic program can be challenging, especially when importance sampling is combined with other variance redution techniques (see Boire et al. (2021*c*), Appendix A.3).

standard LSM algorithm and takes the form

$$\hat{\tilde{C}}_{n,j} = \psi(\tilde{S}_{n,j})\hat{\beta}_j, \tag{4.24}$$

and the dynamic program with importance sampling is then written

$$\begin{cases} \hat{\tilde{V}}_{n,J} = \tilde{X}_{n,J}\dfrac{d\mathbb{Q}}{d\tilde{\mathbb{Q}}}(\tilde{S}_{n,J}, \Delta t) \\ \hat{\tilde{V}}_{n,j} = \begin{cases} \tilde{X}_{n,j}\frac{d\mathbb{Q}}{d\tilde{\mathbb{Q}}}(\tilde{S}_{n,j}, \Delta t) & \text{if } (\tilde{X}_{n,j} \geq \hat{\tilde{C}}_{n,j}) \text{ and } (\tilde{X}_{n,j} > 0) \\ \hat{\tilde{V}}_{n,j+1}\frac{d\mathbb{Q}}{d\tilde{\mathbb{Q}}}(\tilde{S}_{n,j}, \Delta t) & \text{if } (\tilde{X}_{n,j} < \hat{\tilde{C}}_{n,j}) \text{ or } (\tilde{X}_{n,j} = 0) \end{cases} \quad j = J-1, \ldots, 1. \end{cases} \tag{4.25}$$

The resulting sample size-N LSM-I estimator is then written

$$\hat{\tilde{V}}_0^{(N)} = \frac{1}{N}\sum_{n=0}^{N} \hat{\tilde{V}}_{n,0} = \frac{1}{N}\sum_{n=0}^{N} \tilde{X}_{n,\tilde{k}_n}\frac{d\mathbb{Q}}{d\tilde{\mathbb{Q}}}(\tilde{S}_{n,\tilde{k}_n}, \tau_{\tilde{k}_n}), \tag{4.26}$$

where $\tau_{\tilde{k}_n} = \tilde{k}_n \Delta t$ is the path-n stopping time estimated in the LSM-I algorithm, and $\tilde{X}_{n,\tilde{k}_n}$ is the corresponding discounted payoff. Importance sampling modifies the standard LSM in many ways. First, the change of measure increases the frequency of non-null payoffs. Second, shifting the paths toward the money also changes the estimated stopping times because the option is typically exercised earlier than under the nominal measure. The change of drift therefore achieves variance reduction efficiency gains by reducing the sampling variance, which is the primary purpose of importance sampling methods in general applications.

It is worth noting here that further improvements are possible when regressions are performed on the shifted paths directly, as proposed by Boire et al. (2021*a*). This alternative approach to importance sampling is here termed LSM-s where the "-s" stands for "shifted regressions", because the regressions are performed under the importance measure instead of the nominal measure. This has many desirable effects. First, shifting the paths toward the money generate larger proportions of ITM paths, which in turn generates larger regression samples and more accurate continuation value estimates. Second, the strategy is estimated from paths that are in the vicinity of the optimal exercise boundary, which is crucial to achieving exercise optimality. The authors show that directly performing regressions on the importance measure indeed reduces estimator bias. Third, the LSM-s entails a less cumbersome implementation than the LSM-I of Moreni (2003). The LSM-I uses paths simulated under the nominal measure to estimate regression coefficients as in Equation (4.5), and then uses paths simulated under the importance measure to estimate option cash flows. In contrast, the LSM-s only uses paths that are simulated under the importance measure and is thus more convenient. In this article, we mainly focus on the importance sampling technique of Moreni (2003), which is the standard, and briefly return to the Boire et al. (2021*a*) method when discussing results for combined variance reduction techniques in Section 4.4.3.

A similar estimator that implements importance sampling for FDM prices, called the FDM-I, is obtained by replacing the continuation values $\hat{\tilde{C}}_{n,j}$ by the time-j FD boundary b_j^* in the LSM-I algorithm in Equation (4.25). Denoting the path-n, time-j option values as $\hat{\tilde{V}}_{n,0}^*$ we then write sample size-N FDM-I estimator as

$$\hat{\tilde{V}}_0^{*(N)} = \frac{1}{N}\sum_{n=0}^{N} \hat{\tilde{V}}_{n,0}^* = \frac{1}{N}\sum_{n=0}^{N} \tilde{X}_{n,\tilde{k}_n^*}\frac{d\mathbb{Q}}{d\tilde{\mathbb{Q}}}(\tilde{S}_{n,\tilde{k}_n^*}, \tau_{\tilde{k}_n^*}) \tag{4.27}$$

where $\tau_{\tilde{k}_n^*} = \tilde{k}_n^* \Delta t$ is the path-n (optimal) stopping time determined by the FD exercise boundary.

Finally, it is worth noting that, unlike other variance reduction techniques, the implementation of importance sampling requires some care on the part of the user because its efficiency largely depends on an adequate choice of a change of measure. Indeed, a poor choice may increase estimator variance. To provide some intuition about what constitutes an optimal importance measure, it is useful to consider a stratified sampling approach where the optimal sampling weights are inversely proportional to the variance in the strata. This ensures that less effort is spent on sampling from zero-variance regions that do not contribute to the estimation of an expected value. The analogous rationale applies to importance sampling. The optimal importance measure should sample more heavily in regions where the variance of the random variable is high, and less so in regions where the variance is low. Due to the computational cost of optimizing a drift parameter for American-style options, we set the drift parameter λ to the variance-minimizing drift for a European option computed with the saddle point approximation of Glasserman et al. (1999). This choice of drift is the one favoured by Moreni (2003), Lemieux and La (2005), and Boire et al. (2021*a*) because it is easily computed and consistently reduces variance.

4.3.4 Combining variance reduction techniques

In the following, we provide a detailed outline of the implementation of estimators that combine importance sampling and control variates. To compare the efficiency gains of variance reduction techniques implemented with and without a corrected strategy, we present four new estimators: 1) the LSM-CCF-CCV that combines CCF and CCV methods, 2) the LSM-I-CCF estimator that combines importance sampling and CCF methods, 3) the LSM-I-CCV estimator that combines importance sampling and CCV methods, and 4) the LSM-I-CCF-CCV that combines all three. We also show how to implement importance sampling and CCF methods in the context of an FDM algorithm. The variance of this FDM-I-CCF estimator will then serve as a lower bound for the analogous LSM-I-CCF and LSM-I-CCF-CCV estimators.

LSM-CCF-CCV: The sample size-N LSM-CCF-CCV estimator is obtained by controlling the payoffs an LSM-CCV estimator simply by setting

$$\hat{W}_0^{C(N)} = \frac{1}{N} \sum_{n=0}^{N} Z_{n,k_n^C}(\hat{\theta}^C), \tag{4.28}$$

where the optimal control parameter is now estimated with the LSM-CCV stopping times as $\hat{\theta}^C = -\sum_{n=1}^{N}((X_{n,k_n^C} - \bar{X}_{k^C})(Y_{n,k_n^C} - \bar{Y}_{k^C}))/\sum_{n=1}^{N}(Y_{n,k_n^C} - \bar{Y}_{k^C})^2$, where $\bar{X}_{k^C} = N^{-1}\sum_{n=1}^{N} X_{n,k_n^C}$ and $\bar{Y}_{k^C} = N^{-1}\sum_{n=1}^{N} Y_{n,k_n^C}$.

LSM-I-CCF: With importance sampling, the European option prices serving as control variates are also sampled under the importance measure $\tilde{\mathbb{Q}}$. That is, the control variate $\tilde{Y}_{n,j} = \mathbb{E}_j\left[\tilde{X}_{n,J}\right]$ is the path-n, time-j discounted value of a portfolio that replicates the shifted payoff of the American option at maturity $\tilde{X}_{n,J}$. The sample size-N LSM-I-CCF estimator is then

written

$$\hat{\tilde{V}}_0^{C(N)} = \frac{1}{N}\sum_{n=0}^{N}\left(\tilde{X}_{n,\tilde{k}_n} + \tilde{\theta}\tilde{Y}_{n,\tilde{k}_n}\right)\frac{d\mathbb{Q}}{d\tilde{\mathbb{Q}}}(\tilde{S}_{n,\tilde{k}_n}, \tau_{\tilde{k}_n}) - \tilde{\theta}Y_{n,0}, \tag{4.29}$$

where $\tau_{\tilde{k}_n} = \tilde{k}_n\Delta t$ is the same path-n stopping time from as the LSM-I algorithm, and $\tilde{X}_{n,\tilde{k}_n}$ is the corresponding discounted payoff. The optimal control parameter is now estimated with the LSM-I stopping times as $\hat{\tilde{\theta}} = -\sum_{n=1}^{N}((X_{n,\tilde{k}_n} - \bar{X}_{\tilde{k}})(Y_{n,\tilde{k}_n} - \bar{Y}_{\tilde{k}}))/\sum_{n=1}^{N}(Y_{n,\tilde{k}_n} - \bar{Y}_{\tilde{k}})^2$, where $\bar{X}_{\tilde{k}} = N^{-1}\sum_{n=1}^{N} X_{n,\tilde{k}_n}$ and $\bar{Y}_{\tilde{k}} = N^{-1}\sum_{n=1}^{N} Y_{n,\tilde{k}_n}$.

LSM-I-CCV: Implementing importance sampling with a corrected strategy involves a new dynamic program. We therefore write the LSM-I algorithm with controlled continuation values, or LSM-I-CCV, as

$$\begin{cases} \hat{\tilde{W}}_{n,J} = \tilde{X}_{n,J}\dfrac{d\mathbb{Q}}{d\tilde{\mathbb{Q}}}(\tilde{S}_{n,J}, \Delta t) \\ \hat{\tilde{W}}_{n,j} = \begin{cases} \tilde{X}_{n,j}\frac{d\mathbb{Q}}{d\tilde{\mathbb{Q}}}(\tilde{S}_{n,j}, \Delta t) & \text{if } \left(\tilde{X}_{n,j} \geq \hat{\tilde{D}}_{n,j}(\hat{\tilde{\vartheta}}_{n,j})\right) \text{ and } \left(\tilde{X}_{n,j} > 0\right) \\ \hat{\tilde{W}}_{n,j+1}\frac{d\mathbb{Q}}{d\tilde{\mathbb{Q}}}(\tilde{S}_{n,j}, \Delta t) & \text{if } \left(\tilde{X}_{n,j} < \hat{\tilde{D}}_{n,j}(\hat{\tilde{\vartheta}}_{n,j})\right) \text{ or } \left(\tilde{X}_{n,j} = 0\right) \end{cases} \quad j = J-1, \ldots, 1, \end{cases} \tag{4.30}$$

where the path-n, time-j cross-section of controlled continuation values in the LSM-I-CCV are

$$\hat{\tilde{D}}_{n,j}(\tilde{\vartheta}_{n,j}) = \psi(\tilde{S}_{n,j})\hat{\phi}_j + \tilde{\vartheta}_{n,j}\left(\psi(\tilde{S}_{n,j})\hat{\rho}_j - \tilde{Y}_{n,j}\right). \tag{4.31}$$

The CCV parameters $\hat{\phi}_j$ and $\hat{\rho}_j$ are the same coefficients that are used in the LSM-CCV algorithm. Furthermore, the path-n, time-j variance-minimizing control parameter $\tilde{\vartheta}_{n,j}$ is now estimated as

$$\hat{\tilde{\vartheta}}_{n,j} = \frac{\psi(\tilde{S}_{n,j})\left(\Psi_j'\Psi_j\right)^{-1}\Psi_j'(\hat{\mathbf{W}}_{j+1}\odot\mathbf{Y}_j) - (\psi(\tilde{S}_{n,j})\hat{\phi}_j)(\psi(\tilde{S}_{n,j})\hat{\rho}_j)}{\psi(\tilde{S}_{n,j})\left(\Psi_j'\Psi_j\right)^{-1}\Psi_j'(\mathbf{Y}_j\odot\mathbf{Y}_j) - (\psi(\tilde{S}_{n,j})\hat{\rho}_j)^2}, \tag{4.32}$$

for all $n \in \tilde{\mathcal{N}}_j$. Finally, the sample size-N LSM-I-CCV estimator is defined as

$$\hat{W}_0^{(N)} = \frac{1}{N}\sum_{n=0}^{N}\hat{W}_{n,0}^{(N)} = \frac{1}{N}\sum_{n=0}^{N}\tilde{X}_{n,\tilde{k}_n^C}\frac{d\mathbb{Q}}{d\tilde{\mathbb{Q}}}(\tilde{S}_{n,\tilde{k}_n^C}, \tau_{\tilde{k}_n^C}) \tag{4.33}$$

where $\tilde{\tau}_{k_n^C} = k_n^C\Delta t$ is the path-n stopping time from the LSM-I-CCV, and $X_{n,\tilde{k}_n^C}$ is the corresponding discounted payoff.

LSM-I-CCF-CCV: When we combine importance sampling, CCF and CCV methods, we obtain the LSM-I-CCF-CCV program, which implements the LSM-I-CCV algorithm and applies control variates to the estimated cash flows. The sample size-N LSM-I-CCF-CCV estimator is then written

$$\hat{\tilde{W}}_0^{C(N)} = \frac{1}{N}\sum_{n=0}^{N}\left(\tilde{X}_{n,\tilde{k}_n^C} + \tilde{\theta}^C\tilde{Y}_{n,\tilde{k}_n^C}\right)\frac{d\mathbb{Q}}{d\tilde{\mathbb{Q}}}(\tilde{S}_{n,\tilde{k}_n^C}, \tau_{\tilde{k}_n^C}) - \tilde{\theta}^C Y_{n,0}, \tag{4.34}$$

where $\tau_{\tilde{k}_n^C} = \tilde{k}_n^C \Delta t$ is the same path-n stopping time from the LSM-I-CCV algorithm, and $\tilde{X}_{n,\tilde{k}_n^C}$ is the corresponding discounted payoff. The optimal control parameter is now estimated with the LSM-I-CCV stopping times as $\hat{\tilde{\theta}}^C = -\sum_{n=1}^N ((X_{n,\tilde{k}_n^C} - \bar{X}_{\tilde{k}^C})(Y_{n,\tilde{k}_n^C} - \bar{Y}_{\tilde{k}^C}))/\sum_{n=1}^N (Y_{n,\tilde{k}_n^C} - \bar{Y}_{\tilde{k}^C})^2$, where $\bar{X}_{\tilde{k}^C} = N^{-1}\sum_{n=1}^N X_{n,\tilde{k}_n^C}$ and $\bar{Y}_{\tilde{k}^C} = N^{-1}\sum_{n=1}^N Y_{n,\tilde{k}_n^C}$.

FDM-I-CCF: Finally, we can apply a combination of importance sampling and CCF methods in the FDM estimator. The CCV approach if of no use in this particular algorithm, as the stopping times are already optimal. The sample size-N FDM-I-CCF is expressed as

$$\hat{\tilde{V}}_0^{*C(N)} = \frac{1}{N}\sum_{n=0}^{N}\left(\tilde{X}_{n,k_n^*} + \theta^*\tilde{Y}_j\right)\frac{d\mathbb{Q}}{d\tilde{\mathbb{Q}}}(\tilde{S}_{n,k_n^*}, \tau_{k_n^*}) - \theta_j^* Y_{n,j}, \tag{4.35}$$

and uses the same (optimal) stopping times as those determined in the FDM-I algorithm.

4.4 Results

Three sets of results are reported in this section to illustrate the benefits of a corrected strategy in the LSM with a GBM asset price diffusion. First, we compare how closely corrected and uncorrected strategies replicate the optimal stopping times recommended by the FD exercise boundary. We then show that a corrected strategy effectively removes the bias. Second, we demonstrate that removing exercise errors greatly amplifies the efficiency gains permitted by the CCF method. Third, we report that the joint implementation of importance sampling and CCF methods is made efficient by using a corrected strategy. Indeed, with an uncorrected strategy, this specific combination of variance reduction techniques does not significantly reduce the variance of an estimator that uses CCF control variates only. In other words, we find that the LSM-I-CCF and the LSM-CCF have similar variances, whereas the LSM-I-CCF-CCV has a much smaller variance than the LSM-I-CCF. Finally, we show how the efficiency of a combined implementation of importance sampling and CCF methods is further increased by implementing the shifted regression approach to importance sampling presented in Boire et al. (2021*b*).

We test put option price estimators across a wide range of maturities $T \in \{0.5, 1, 1.5, 2\}$ and levels of moneyness. In all cases, we set $J/T = 50$ (i.e., 50 exercise opportunities per year), $r = q = 0.06$, $\sigma = 0.4$, $K = 40$, and $S_0 \in \{34, 36, 38, 40, 42, 44, 46\}$ for a total of 28 options. We use a cubic approximation of the continuation value (i.e., $L = 3$) and set the basis functions used in the regression to $(S_j/K)^\ell : \ell = 0, \ldots, L$. Unless specified otherwise, our analyses systematically compare the LSM, FDM and LSM-CCV methods outlined above with $N \in \{10^4, 10^5\}$ simulated paths per repeated valuation to estimate the decision boundary. For each option, we use a total budget of 10^8 sample paths, such that we use $R = 10^8/N$ independently repeated option valuations to compute five statistics: bias, standard error, RMSE, variance reduction efficiency, and the ratio of the estimator variance compared to the analogous FDM variance. The FD boundary is obtained by pricing an ATM put option with a maturity of $T = 2$ years, 20,000 time steps per year, $J + 1$ admissible exercise opportunities $\{\tau_j : j = 0, \ldots, J\}$, 80,000 equal stock price increments over the interval $[0, S_{\max}]$, where $S_{\max} = 200$. The same boundary can then be used to compute FDM estimators for options with shorter maturities and

varying initial asset prices. The bias is computed as the difference between the average of R estimator valuations and the benchmark price. Benchmark prices are recovered from the same FD solution used to approximate the optimal boundary by selecting the option value at the node that corresponds to the option maturity and initial asset price. The bias of the FDM approach does not depend on a regression design as in the LSM, but on the refinement of the lattice used to approximate a solution. By using very fine partitions of time and stock price ($dt = 5 \times 10^{-5}$ and $dS = 5 \times 10^{-4}$, respectively, in the present experiments), the benchmark can be made arbitrarily accurate, such that, as $N \to \infty$, the LSM estimators should be asymptotically negatively biased compared to our selected benchmark. Next, we compute estimator standard errors as the standard deviation of the sample of R values. The bias and standard errors are then used to compute the RMSE. RMSE values are computed as the square root of the sum of the squared bias and the variance. We also use the standard errors to compute the efficiency of variance reduction techniques. Considering a general estimator $\hat{V}_0$ whose variance serves as the basis of comparison for variance reduction, the standard error efficiency of a modified estimator $\hat{U}_0$ is defined as

$$\text{Eff}(\hat{U}_0) = \frac{\text{SE}(\hat{V}_0)}{\text{SE}(\hat{U}_0)} - 1, \tag{4.36}$$

where $\text{SE}(V)$ denotes the estimated standard error of a given estimator V. A positive efficiency thus indicates that the $\hat{U}_0$ has a lower variance than $\hat{V}_0$. To measure the efficiency, both standard error measures used in the calculation of efficiency $\hat{V}_0$ and $\hat{U}_0$ need to be the results obtained with the same simulated paths and the same number of independently repeated valuations. Finally, we also compute the ratio of the variances of analogous LSM and FDM estimators to measure how closely the LSM estimators approach the lower-bound variance of the analogous FDM estimator. A ratio greater than one then indicates that the LSM estimator has a larger variance than the FDM estimator.

4.4.1 Stopping time optimality

In the following, we discuss the relation between stopping time optimality, bias and variance in the standard LSM with by using the CCV approach. To do so, we compare the stopping times of LSM and LSM-CCV estimators to those of an FDM estimator to see how closely estimated stopping times replicate the optimal stopping times. When estimated stopping times approach optimality, we find that the bias is significantly reduced. Alternatively, when estimated stopping times are sub-optimally estimated, the bias is negative. Furthermore, when the estimated stopping times are guided by foresight via regression overfitting, the bias is positive. Finally, we find that exercise errors have little to no impact on the variance of estimators when there are no variance reduction techniques.

To measure stopping time optimality, we compare the proportion of estimated stopping times that coincide with the "optimal" stopping times recommended by the FD exercise boundary. To provide some intuition about how stopping time errors may generate positive and negative bias, we categorize stopping times as "sub-optimal" if the estimated option cash flow is lower than the optimal cash flow and as "supra-optimal" if it is greater. Here, we assume that a sub-optimal strategy causes sub-optimal stopping times and leads to negative bias. Conversely, a strategy that benefits from overfitting leads to positive bias.

Table 4.4.1: Proportions of exercise times (%).

	Number of simulated paths (N)							
LSM	100	500	1,000	2,000	3,000	5,000	10,000	100,000
Sub-optimal	24.61	23.87	23.06	22.16	21.39	20.39	18.98	16.71
Optimal	42.89	54.19	57.37	60.37	62.47	65.03	68.52	74.74
Supra-optimal	32.50	21.95	19.56	17.47	16.13	14.58	12.49	8.54
CCV	100	500	1,000	2,000	3,000	5,000	10,000	100,000
Sub-optimal	7.32	4.29	3.58	3.10	2.91	2.76	2.67	2.55
Optimal	87.11	93.17	94.50	95.36	95.70	95.96	96.15	96.34
Supra-optimal	5.56	2.55	1.92	1.54	1.39	1.27	1.19	1.11

Proportions of sub-optimal, optimal, and supra-optimal exercise times for LSM and LSM-CCV estimators. Proportions are expressed in percentage and measured as the average proportion across $R = 10^8/N$ replications when using N simulation paths. Stopping times are classified as optimal when they coincide with the FDM stopping times. Otherwise, stopping times are classified as sub-optimal if they generate smaller cash flows than the optimal strategy, and supra-optimal if they generate larger discounted cash flows. Results are reported for a one-year ATM American put option written on a GBM underlying asset with diffusion parameters $S_0 = K = 40$, $r = q = 0.06$, $\sigma = 0.4$, and $J/T = 50$.

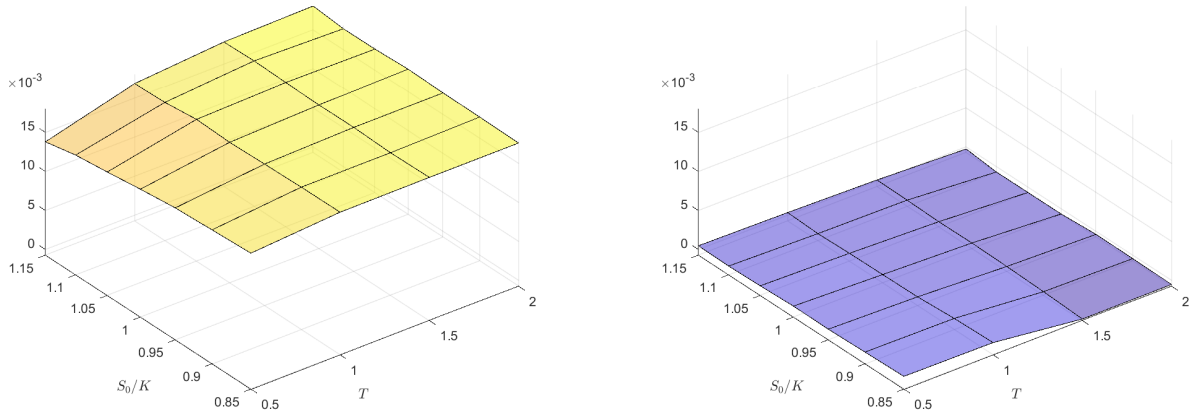
Table 4.4.1 shows the proportions of stopping time categories for the standard LSM estimator and the LSM-CCV estimator. Results indicate that the corrected strategy significantly increases the frequency of optimally exercised paths. We observe that an LSM-CCV estimator with only $N = 100$ simulated paths generates 87.16% of optimal stopping times, compared to 42.89% for the LSM. Moreover, with $N = 100,000$ simulated paths, the LSM-CCV optimally exercises 96.34% of paths, compared to 74.74% for the LSM. Indeed, the LSM-CCV requires much fewer simulation paths to obtain the same frequency of optimal stopping times as the LSM. Moreover, we notice that the LSM tends to have a higher frequency of supra-optimal stopping times than sub-optimal stopping times when the number of paths is low, which indicates that foresight bias is indeed an important source of bias in those cases. As the number of paths increases, the LSM generates more sub-optimal stopping times than it does supra-optimal ones, yet it still has over over 8% of supra-optimal stopping times when $N = 100,000$. In contrast, the LSM-CCV estimator always has a lower frequency of supra-optimal than sub-optimal stopping times. Moreover, we report about 1% of supra-optimal stopping times in the LSM-CCV when $N = 100,000$, suggesting that a corrected strategy effectively deals with both positive and negative bias elements. Consistent with the results in Table 4.4.1, Table 4.4.2 shows that the LSM is positively biased across all numbers of paths, suggesting that overfitting is indeed the cause of the positive bias reported across all sample sizes. In contrast, the LSM-CCV estimator exhibits nearly zero bias even for $N = 100$.

To compare how the biases of LSM and LSM-CCV estimators vary across option characteristics like moneyness and time to maturity, Figure 4.4.1 illustrates the bias of LSM and LSM-CCV estimators across levels of moneyness K/S_0 and maturities T for $N = 1,000$. As expected from the high frequency of supra-optimal exercise times, it is clear from panel 4.1(a)

Table 4.4.2: Bias comparison.

	Number of simulated paths (N)							
Strategy	100	500	1,000	2,000	3,000	5,000	10,000	100,000
LSM	0.6408	0.2138	0.1305	0.0768	0.0534	0.0374	0.0197	0.0005
CCV	-0.0131	-0.0008	0.0005	0.0012	0.0002	0.0030	0.0021	-0.0000
FDM	-0.0012	0.0022	0.0020	0.0019	0.0007	0.0037	0.0025	0.0001

Bias results of LSM, LSM-CCV, and FDM estimators. Bias is measured as the average bias across $R = 10^8/N$ replications when using N simulation paths. Results are reported for a one-year ATM American put option written on a GBM underlying asset with diffusion parameters $S_0 = K = 40$, $r = q = 0.06$, $\sigma = 0.4$, and $J/T = 50$.



(a) LSM (b) LSM-CCV

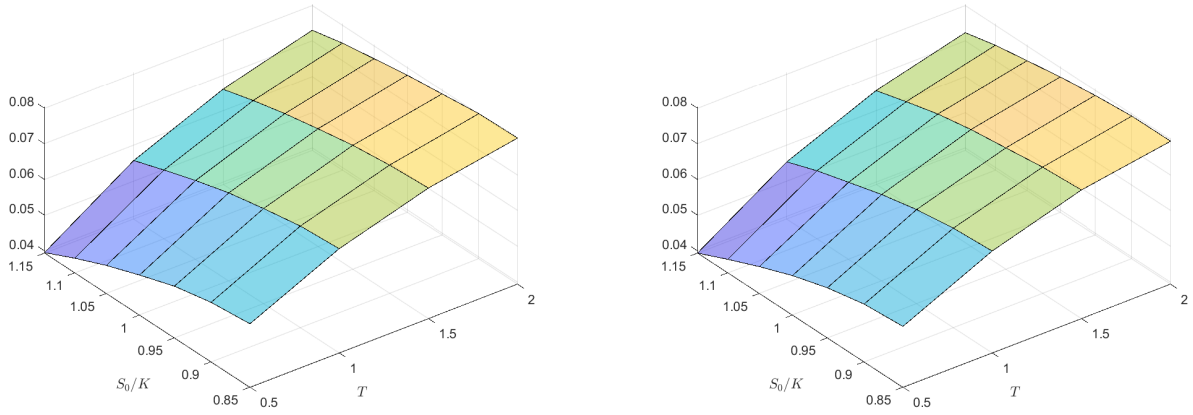
Figure 4.4.1: LSM and LSM-CCV bias.

Bias results are calculated by comparing the FD solution to the LSM and LSM-CCV estimators. Results are reported for an American put option estimators using $N = 10,000$ paths and $R = 10,000$ repetitions with $K = 40$, $S_0 \in \{34, 36, 38, 40, 42, 44, 46\}$, $r = q = 0.06$, $T \in \{0.5, 1, 1.5, 2\}$, and $\sigma = 0.4$.

that in the LSM algorithm, foresight bias dominates sub-optimality bias leading to a net positive bias. In contrast, the LSM-CCV results illustrated in panel 4.1(b) show that an estimator with a corrected strategy exhibits little to no bias. In fact, detailed results shown in Tables 4.A.1 and 4.A.2 of Appendix 4.A for $N = 10,000$ and $N = 100,0000$, respectively, indicate that the "unbiased" FDM and LSM-CCV biases are very close in all cases considered. Referring to the same tables, similar conclusions are reached when comparing the biases of the LSM, LSM-CCV and FDM estimators when they are supplemented with variance reduction techniques.

This is interesting because a controlled strategy corrects not only the negative bias elements caused by sub-optimal decisions but also fixes the positive bias caused by foresight. To see why the LSM-CCV is well-adapted to the task of reducing both bias elements, we consider the work of Whitehead et al. (2012), Kan et al. (2009), and Boire et al. (2021*b*). Assuming that the estimated strategy is independent of future option values (i.e., assuming that there is

no positive foresight bias), Whitehead et al. (2012) derive explicit approximations to the sub-optimality bias of American option prices in the context of the easy-to-analyze stochastic tree and mesh of Broadie and Glasserman (1997, 2004), while Kan and Reesor (2012) provide analogous derivations for LSM estimators. Boire et al. (2021*b*) extend this by deriving an explicit approximation to the foresight bias of LSM prices. They find that both sub-optimality and foresight biases (locally defined along path n and at time j) depend on the conditional variance of the fitted continuation value (i.e., $\text{Var}_j[\hat{C}_{n,j}]$ in the LSM algorithm). Both sources of bias go to zero as the conditional variance of the continuation value approaches zero. This happens when the sample size increases, for instance, because (local) bias is $O(N^{-1})$ (Kan et al. 2009). This also happens when we apply a variance reduction technique to continuation value estimates as we do with the CCV method. Hence, the CCV method leads to a reduction of all sources of bias, except for the approximation bias caused by the finite parameterization of the early-exercise strategy.

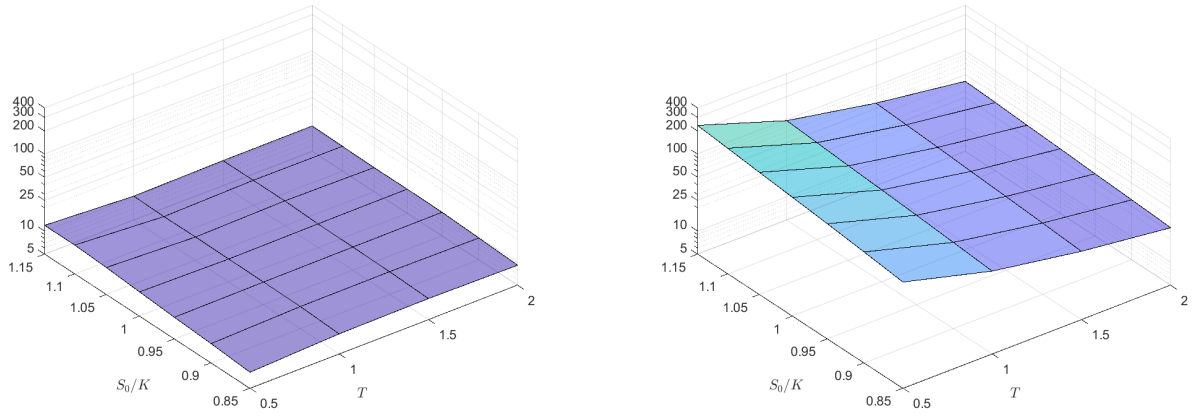


(a) LSM (b) LSM-CCV

Figure 4.4.2: LSM and LSM-CCV standard error.

Standard error results for LSM and LSM-CCV estimators. Results are reported for an American put option estimators using $N = 10,000$ paths and $R = 10,000$ repetitions with $K = 40$, $S_0 \in \{34, 36, 38, 40, 42, 44, 46\}$, $r = q = 0.06\}$, $T \in \{0.5, 1, 1.5, 2\}$, and $\sigma = 0.4$.

To conclude our comparison of LSM and LSM-CCV estimators, we observe in Figure 4.4.2 that the standard error of LSM and LSM-CCV estimators are very similar across all cases considered. As expected, the variance increases with option maturity T and decreases with option moneyness (defined as K/S_0) in both cases. Therefore, we conclude that the effect of exercise errors on the variance is dwarfed by the large variance of the payoffs themselves. Detailed results presented in Tables 4.A.1 and 4.A.2 of Appendix 4.A for $N = 10,000$ and $N = 100,0000$, respectively, indicate that the standard errors of LSM, LSM-CCV, and FDM estimators are indeed very close. Therefore, using a corrected strategy here will not significantly change the variance of the estimator, but significantly decreases the bias. We will see in the following sections that the effect of exercise errors does have a significant impact on the variance of the LSM estimator when control variates are involved. In those cases, the corrected strategy operates both a reduction in both bias and variance, yielding significant efficiency gains.



(a) LSM-CCF

(b) LSM-CCF-CCV

Figure 4.4.3: Standard error efficiency of control variates.

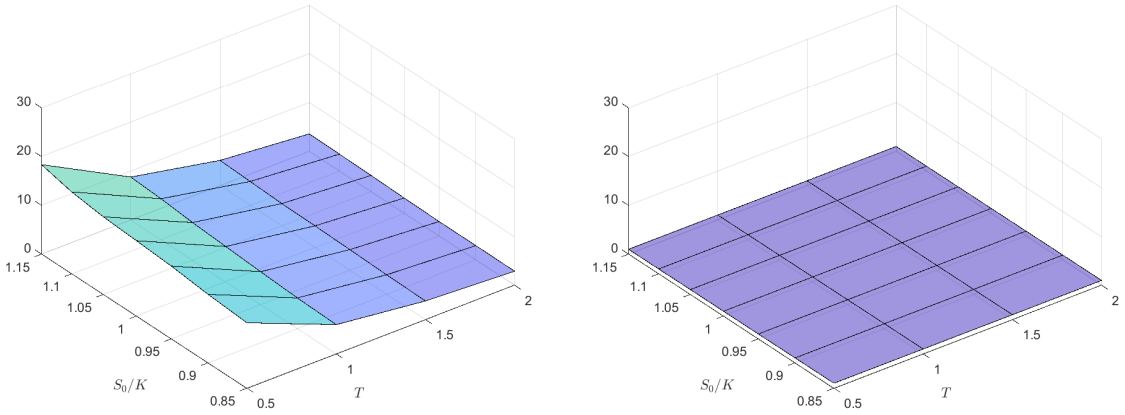
Standard error efficiencies of control variates for LSM and LSM-CCV estimators. Results are reported for an American put option estimators using $N = 10,000$ paths and $R = 10,000$ repetitions with $K = 40$, $S_0 \in \{34, 36, 38, 40, 42, 44, 46\}$, $r = q = 0.06$, $T \in \{0.5, 1, 1.5, 2\}$, and $\sigma = 0.4$.

4.4.2 Improved efficiency of control variates

Seeing as the LSM-CCV method reduces the bias, but does not significantly change the variance, one might consider implementing variance reduction techniques for this estimator. A natural choice is that of the CCF method because the computation of control variates is already required by the CCV approach and thus comes at no additional cost. Furthermore, we now show that the CCF method is extremely effective at variance reduction when the CCV strategy correction is implemented. A combined implementation of both CCV and CCF methods therefore effectively reduces both bias and variance, yielding a drastic reduction in the RMSE.

The left panel of Figure 4.4.3 shows that the LSM-CCF-CCV method is much more efficient than the LSM-CCF method, presenting standard error efficiency gains of a much greater order of magnitude across a range of moneyness levels and maturity. Indeed, detailed results shown in Table 4.A.7 of Appendix 4.A show that the standard error reduction factor is between 8 and 12 in the LSM-CCF compared to factors between 27 and 230 in the LSM-CCF-CCV. This huge gain in efficiency is due to the fact that the variance caused by exercise errors is a significant portion of the estimator variance when control variates are used. In turn, when the sampling variance is already minimized, it stands to reason that the most potent technique should focus on the remaining variance related to exercise errors. The same conclusions can be drawn by observing the results for estimators using $N = 100,000$ simulated paths, as shown in Table 4.A.8 of Appendix 4.A, where similar efficiency gains are reported.

Furthermore, we can see in Figure 4.4.4 that the LSM-CCF-CCV nearly achieves the optimal standard error, as suggested by the ratio of its variance and that of an FDM-CCF estimator approaching one. As argued at the end of Section 4.2.3, the standard error efficiency of the CCF approach is essentially maximized because it is equal to the standard error efficiency obtained by using the optimal strategy. Indeed, by eliminating the effect of exercise errors on the variance of the estimator, we get rid of the extra variance component, yielding optimal variance



(a) LSM-CCF (b) LSM-CCF-CCV

Figure 4.4.4: Standard error ratio with respect to the FDM-CCF.

Ratio of the standard error of implementations with the CCF method and the standard error of the FDM-CCF. Results are reported for an American put option estimators using $N = 10,000$ paths and $R = 10,000$ repetitions with $K = 40$, $S_0 \in \{34, 36, 38, 40, 42, 44, 46\}$, $r = q = 0.06$, $T \in \{0.5, 1, 1.5, 2\}$, and $\sigma = 0.4$.

reduction. Results show that the LSM-CCF-CCV efficiency is optimal across various maturities, levels of moneyness, and sample sizes, as seen in Tables 4.A.11 and 4.A.12 of Appendix 4.A for $N = 10,000$ and $N = 100,0000$, respectively.

The net effect of a reduction in bias and variance is therefore a substantial reduction in the RMSE. Detailed RMSE results reported in Table 4.A.5 indicate that a corrected strategy reduces the RMSE of the LSM-CCF method by 91.61% on average when the sample size is $N = 10,000$. The relative importance of these RMSE improvements decreases as the moneyness increases, and decreases as the time to maturity increases, suggesting that the LSM-CCF-CCV is most efficient for short-term, deep OTM options. The same conclusion can be drawn for a sample size of $N = 100,000$, where we report an average RMSE reduction of 88.23% across the cases considered.

4.4.3 Efficient combination of control variates and importance sampling

Finally, we demonstrate that using an accurate strategy permits an efficient combination of importance sampling and CCF methods. Indeed, it has recently been shown in Boire et al. (2021*c*) that combining the Moreni (2003) importance sampling approach to the CCF method does not improve the variance of an estimator that uses the CCF method only. The inefficiency of importance sampling when combined with control variates only occurs with an uncorrected strategy, which leads us to believe that inaccurate stopping times may be at fault. Indeed, detailed standard error efficiency results in Table 4.A.7 in Appendix 4.A show that the efficiency of the LSM-I-CCF is not significantly different than the efficiency of the LSM-CCF, which calls into question the utility of using this specific combination of variance reduction techniques. However, when a corrected strategy is implemented the combination now seems to yield some efficiency gains. Indeed, comparing the efficiency gains obtained from the LSM-CCF-CCV and the LSM-I-CCF-CCV, it is apparent that some significant improvements are obtained by

adding importance sampling, with an average increase in efficiency of 34.46%. The issue of the inefficient combination of the Moreni (2003) importance sampling and the CCF methods is therefore resolved by correcting the strategy with the CCV method.

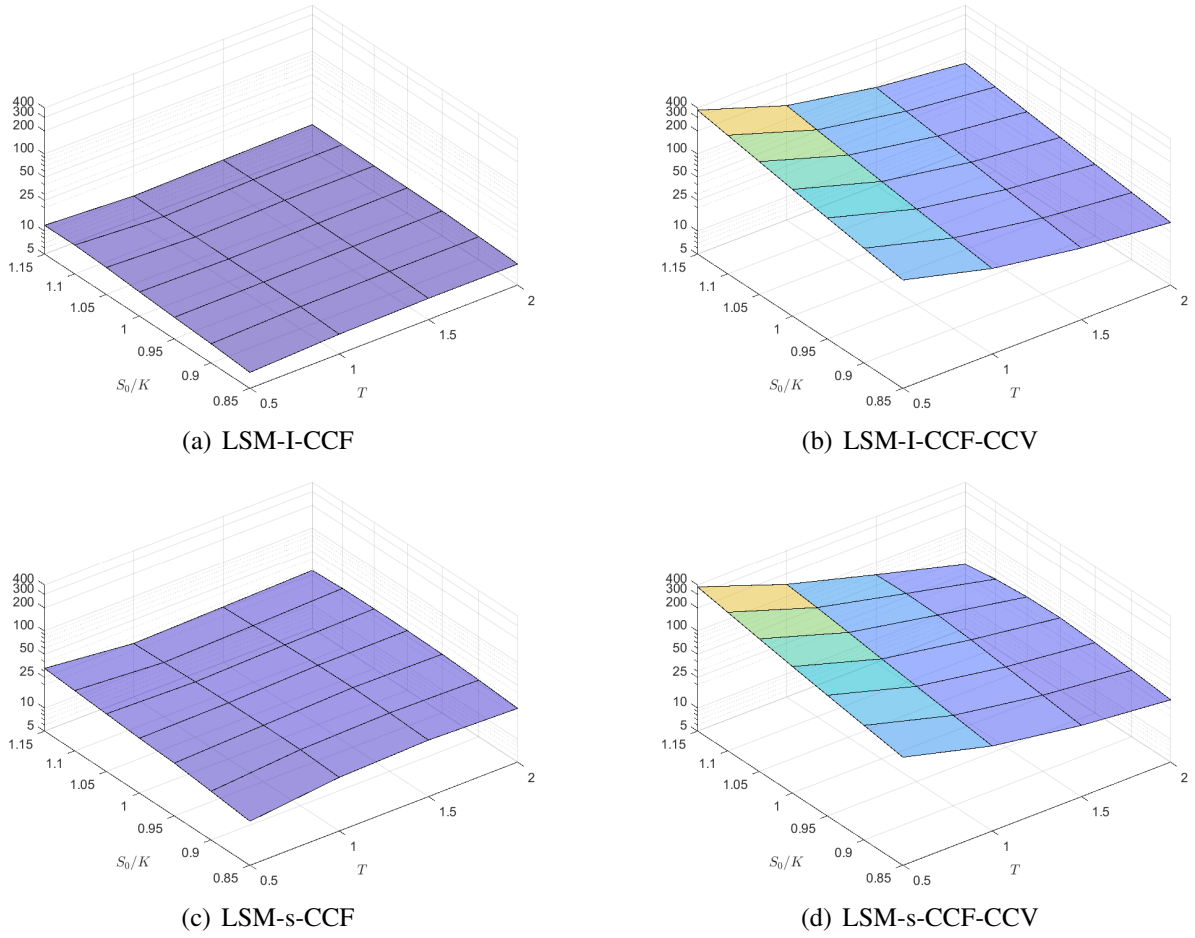
To further explore how stopping time optimality interacts with the efficiency of the importance sampling and CCF combination, we test the LSM-s importance sampling approach of Boire et al. (2021*a*) instead of the approach of Moreni (2003). This alternative approach performs regressions under the importance measure with the shifted paths instead of estimating the strategy under the nominal measure and has been shown to improve the fitted continuation values and reduce estimator bias. Figure 4.4.5 depicts efficiency results for the combinations of importance sampling and CCV methods, where both the LSM-I and the LSM-s importance sampling approaches are tested. Detailed results are shown in Table 4.A.9 in Appendix 4.A. With an uncorrected strategy, the slightly improved strategy of the LSM-s permits to add to the efficiency of the CCF method, which provides further evidence that exercise sub-optimality stifles the efficiency of control variates. Indeed, the combination with the Moreni (2003) approach yields efficiency gains between 8 and 13 and the combination with the Boire et al. (2021*a*) approach yields standard error efficiency gains between 19 and 33. With a corrected strategy, the combination of importance sampling and CCV methods yields efficiency gains between 32 and 375 for both importance sampling techniques.

Furthermore, the LSM-s-CCF-CCV method does not improve the variance of the LSM-I-CCF-CCV here, because the efficiency of importance sampling and control variates is maximized with the corrected strategy, irrespective of the regression design. Figure 4.4.6 indeed shows that the ratio of the variance of the LSM-I-CCF-CCV and the variance of the FDM-I-CCF approaches one, suggesting maximized efficiency gains. Detailed ratio results are presented in Table 4.A.11 in Appendix 4.A. The fact that the efficiency is maximized is further demonstrated by the fact that using the LSM-s approach in combination with the CCV approach does not improve the efficiency, as seen in Table 4.A.9. Indeed, the CCV already corrects most of the exercise errors, such that the shifted regressions of the LSM-s do not further improve stopping time optimality. Once again, the same conclusions hold when we consider the results reported in Tables 4.A.8 and 4.A.10 for the LSM and LSM-s estimators, respectively, with $N = 100,000$ sample paths.

Finally, the impressive reductions in bias and variance once again yield a large reduction in RMSE. Detailed RMSE results reported in Table 4.A.5 indicate that a corrected strategy reduces the RMSE of the LSM-I-CCF method by 93.50% on average when the sample size is $N = 10,000$. Again, the relative importance of these RMSE improvements decreases as the moneyness increases, and decreases as the time to maturity increases, suggesting that the LSM-I-CCF-CCV is most efficient for short-term, deep OTM options. The same conclusion can be drawn for a sample size of $N = 100,000$, where we report an average RMSE reduction of 89.72% across the cases considered.

4.5 Conclusion

To conclude, this paper examines the links between the Least-Squares Monte Carlo American option price estimates and the regression-based early-exercise strategy. The estimated strategy leads to stopping time errors, causing bias and variance. Indeed, a regression-based strategy



(a) LSM-I-CCF

(b) LSM-I-CCF-CCV

(c) LSM-s-CCF

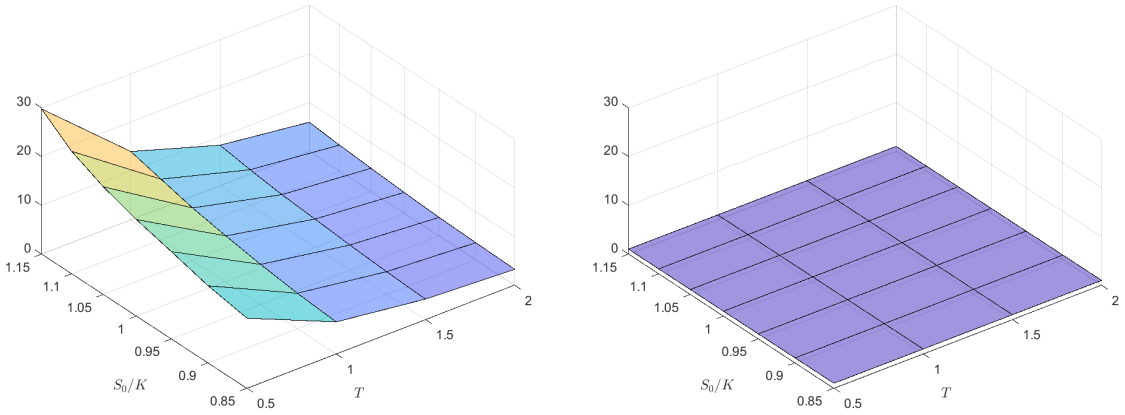
(d) LSM-s-CCF-CCV

Figure 4.4.5: Standard error efficiencies of control variates and importance sampling.

Standard error efficiencies of LSM-I-CCF, LSM-I-CCF-CCV, LSM-s-CCF, LSM-s-CCF-CCV estimators. Results are reported for an American put option estimators using $N = 10,000$ paths and $R = 10,000$ repetitions with $K = 40$, $S_0 \in \{34, 36, 38, 40, 42, 44, 46\}$, $r = q = 0.06$, $T \in \{0.5, 1, 1.5, 2\}$, and $\sigma = 0.4$.

based on a finite sample of simulated paths generates inaccurate exercise times. Whether they be sub-optimal due to model estimation errors or supra-optimal due to overfitting, it is well understood that these stopping time errors cause the bias of LSM estimators. However, the effect of random exercise errors on the variance of the estimator is not duly appreciated in the literature. Although negligible in standard LSM applications, the variance caused by stopping time errors is prominent when variance reduction techniques are implemented. In turn, neglecting to suppress this source of variance severely restricts the potential efficiency gains of variance reduction techniques. Indeed, combining multiple variance reduction techniques in the LSM has rapidly diminishing marginal efficiency gains with an uncorrected strategy. For example, adding importance sampling to an estimator that already implements control variates will not significantly change the variance.

This article first presents evidence that applying control variates to continuation values with the CCV method of Rasmussen (2005) nearly achieves stopping time optimality. First,



(a) LSM-I-CCF

(b) LSM-I-CCF-CCV

Figure 4.4.6: Standard error ratio with respect to the FDM-I-CCF.

Ratio of the standard error of implementations with importance sampling and CCF methods and the standard error of the FDM-I-CCF. Results are reported for an American put option estimators using $N = 10,000$ paths and $R = 10,000$ repetitions with $K = 40$, $S_0 \in \{34, 36, 38, 40, 42, 44, 46\}$, $r = q = 0.06$, $T \in \{0.5, 1, 1.5, 2\}$, and $\sigma = 0.4$.

focusing on exercise optimality yields a drastic reduction in bias. In particular, we show how CCV reduces both the negative sub-optimality bias and the positive foresight bias. Moreover, we demonstrate that the variance of an LSM estimator that uses a CCV-corrected strategy approaches the variance lower bound, suggesting that the CCV approach successfully suppresses the effects of exercise errors on variance. Next, we show that the variance reduction efficiency gains obtained with control variates improve substantially when the strategy is optimal. As a result, using a CCV-corrected strategy further reduces the variance by several orders of magnitude, again approaching the variance lower-bound. Finally, we show that the combined implementation of the importance sampling technique of Boire et al. (2021*a*) and the CCF of Rasmussen (2005) is extremely efficient when stopping times are optimal, resolving the issue of inefficient combinations of variance reduction techniques in the LSM documented by Boire et al. (2021*c*). Altogether, the corrected strategy, on average, corrects 92.85% of the absolute bias of American option estimates, increases the standard error efficiency gains of control variates by a factor of 7.19, and increases the standard error efficiency gains of a combined implementation of control variates and importance sampling by a factor of 9.95.

Appendix

4.A Tables

In this appendix, we present the detailed results discussed in the paper. We present bias, standard error, RMSE, standard error efficiency, and the variance ratio with respect to the FDM variance. For each statistic, we present results for put options with maturities $T \in \{0.5, 1, 1.5, 2\}$ and strike price $K = 40$. The underlying asset is governed by a GBM with an initial asset price $S_0 \in \{34, 36, 38, 40, 42, 44, 46\}$, volatility $\sigma = 0.4$, risk-free rate $r = 0.06$, and dividend yield $q = 0.06$. Prices are computed with sample sizes of $N \in \{10^4, 10^6\}$ and $R = 10^8/N$ repetitions.

Table 4.A.1: Detailed bias results ($\times 10^4$) ($N = 10,000$).

		No VR			**CCF**			**I-CCF**		
T	S_0	V^*	V	W	V^*	V	W	V^*	V	W
0.5	34	9.3	167.0	9.2	0.5	-123.0	-0.4	0.6	-122.5	-0.3
0.5	36	8.3	164.3	7.8	0.6	-114.3	-0.1	0.6	-114.1	-0.1
0.5	38	7.7	161.8	6.4	0.6	-104.4	0.1	0.7	-104.3	0.1
0.5	40	7.6	156.6	6.9	0.7	-93.6	0.2	0.7	-93.8	0.2
0.5	42	5.8	151.0	5.2	0.6	-82.3	0.3	0.7	-82.8	0.3
0.5	44	4.1	146.2	4.0	0.6	-72.1	0.3	0.6	-72.9	0.3
0.5	46	4.9	137.9	4.8	0.5	-63.4	0.3	0.5	-64.5	0.2
1	34	12.1	175.6	8.5	0.4	-136.6	-2.5	0.7	-135.6	-2.4
1	36	10.5	175.2	7.0	0.3	-131.7	-2.0	0.7	-131.6	-1.9
1	38	6.1	174.1	4.0	0.3	-126.8	-1.7	0.7	-127.5	-1.5
1	40	6.4	172.8	3.4	0.3	-122.1	-1.3	0.7	-123.4	-1.2
1	42	6.2	174.5	2.9	0.3	-117.0	-1.1	0.6	-118.8	-1.0
1	44	3.3	169.9	1.7	0.3	-112.5	-0.8	0.6	-114.7	-0.8
1	46	4.8	168.6	3.4	0.3	-106.8	-0.7	0.5	-109.7	-0.7
1.5	34	2.6	176.2	-5.1	0.2	-146.7	-5.4	0.8	-146.0	-5.5
1.5	36	4.8	175.4	-2.6	0.3	-142.5	-4.5	0.7	-143.3	-4.7
1.5	38	5.3	177.2	-1.0	0.3	-138.8	-3.8	0.7	-140.8	-4.0
1.5	40	4.9	176.8	-0.1	0.3	-134.7	-3.2	0.7	-137.7	-3.4
1.5	42	5.3	177.7	-0.5	0.4	-130.2	-2.6	0.6	-134.0	-2.9
1.5	44	5.1	179.0	0.7	0.3	-126.7	-2.2	0.6	-131.1	-2.5
1.5	46	5.0	178.0	1.3	0.3	-123.1	-2.0	0.6	-128.3	-2.3
2	34	5.5	176.3	-4.8	0.4	-154.6	-8.5	0.8	-154.8	-9.2
2	36	4.9	177.2	-4.2	0.3	-150.9	-7.3	0.8	-153.0	-8.1
2	38	4.8	177.4	-3.8	0.4	-146.4	-6.2	0.8	-149.9	-7.0
2	40	3.3	178.6	-6.8	0.3	-143.1	-5.5	0.7	-147.7	-6.1
2	42	0.0	176.7	-7.0	0.3	-140.3	-4.7	0.7	-146.0	-5.3
2	44	1.3	176.3	-7.1	0.3	-137.1	-4.2	0.7	-143.5	-4.7
2	46	4.9	180.2	-2.9	0.3	-133.7	-3.7	0.7	-141.0	-4.3

Detailed bias results obtained from $R = 10,000$ replications when using $N = 10,000$ simulation paths. Results are reported for FDM (V^*), LSM (V) and LSM-CCV (W) estimators across three algorithm configurations: no variance reduction technique ("No VR"), CCF control variates ("CCF"), and a combination of importance sampling and CCF control variates ("I-CCF"). Results are reported for put options with maturities $T \in \{0.5, 1, 1.5, 2\}$ and strike price $K = 40$. The underlying asset is governed by a GBM with an initial asset price $S_0 \in \{34, 36, 38, 40, 42, 44, 46\}$, volatility $\sigma = 0.4$, risk-free rate $r = 0.06$, and dividend yield $q = 0.06$.

Table 4.A.2: Detailed bias results ($\times 10^4$) ($N = 100,000$).

		No VR			**CCF**			**I-CCF**		
T	S_0	V^*	V	W	V^*	V	W	V^*	V	W
0.5	34	4.0	-1.2	2.6	0.5	-39.3	-0.3	0.5	-39.3	-0.3
0.5	36	2.3	-3.3	0.8	0.6	-39.2	-0.0	0.6	-39.2	-0.0
0.5	38	2.4	-3.0	2.3	0.7	-36.8	0.2	0.7	-36.8	0.2
0.5	40	3.4	-0.2	2.8	0.7	-33.9	0.4	0.7	-34.0	0.3
0.5	42	5.8	4.1	5.3	0.7	-30.7	0.4	0.7	-30.8	0.4
0.5	44	6.4	9.2	6.3	0.6	-27.2	0.4	0.6	-27.4	0.4
0.5	46	7.0	12.1	6.9	0.5	-24.0	0.4	0.5	-24.2	0.4
1	34	-11.5	-16.1	-14.2	0.4	-41.3	-2.4	0.3	-41.4	-2.5
1	36	-10.5	-13.9	-12.7	0.5	-40.5	-1.8	0.4	-40.6	-1.9
1	38	-10.6	-14.0	-12.7	0.6	-39.1	-1.3	0.5	-39.3	-1.4
1	40	-11.3	-13.7	-11.8	0.6	-38.1	-0.9	0.5	-38.4	-1.1
1	42	-9.4	-11.1	-11.5	0.5	-36.5	-0.7	0.5	-36.9	-0.8
1	44	-7.7	-8.5	-8.3	0.5	-34.9	-0.5	0.5	-35.4	-0.5
1	46	-5.2	-5.3	-5.8	0.4	-32.8	-0.4	0.5	-33.2	-0.4
1.5	34	-17.9	-22.4	-23.7	0.5	-40.0	-5.1	0.2	-40.2	-5.5
1.5	36	-18.7	-20.4	-21.7	0.5	-38.9	-4.1	0.3	-39.4	-4.5
1.5	38	-13.9	-17.2	-19.8	0.6	-38.2	-3.4	0.3	-38.7	-3.7
1.5	40	-11.2	-13.5	-15.0	0.6	-37.1	-2.8	0.3	-37.8	-3.1
1.5	42	-11.1	-10.9	-13.7	0.5	-36.2	-2.3	0.3	-37.0	-2.6
1.5	44	-9.7	-9.9	-12.4	0.5	-35.6	-1.9	0.3	-36.5	-2.1
1.5	46	-10.6	-4.6	-12.7	0.5	-34.3	-1.5	0.3	-35.2	-1.7
2	34	-15.3	-19.0	-24.3	0.3	-42.2	-8.8	0.0	-42.6	-9.2
2	36	-16.7	-19.0	-25.4	0.4	-40.6	-7.5	0.0	-41.2	-7.8
2	38	-16.3	-19.2	-23.6	0.4	-39.2	-6.3	0.1	-40.0	-6.7
2	40	-17.8	-19.3	-23.0	0.4	-37.8	-5.3	0.2	-38.7	-5.7
2	42	-20.3	-18.2	-25.9	0.5	-36.7	-4.5	0.2	-37.7	-4.8
2	44	-20.4	-18.0	-24.2	0.6	-35.6	-3.6	0.3	-36.8	-4.1
2	46	-17.7	-15.2	-22.3	0.6	-34.4	-3.1	0.3	-35.7	-3.5

Detailed bias results obtained from $R = 1,000$ replications when using $N = 100,000$ simulation paths. Bias results are measured for FDM (V^*), LSM (V) and LSM-CCV (W) estimators across three algorithm configurations: no variance reduction technique ("No VR"), CCF control variates ("CCF"), and a combination of importance sampling and CCF control variates ("I-CCF"). Results are reported for put options with maturities $T \in \{0.5, 1, 1.5, 2\}$ and strike price $K = 40$. The underlying asset is governed by a GBM with an initial asset price $S_0 \in \{34, 36, 38, 40, 42, 44, 46\}$, volatility $\sigma = 0.4$, risk-free rate $r = 0.06$, and dividend yield $q = 0.06$.

Table 4.A.3: Detailed standard error results ($\times 10^3$) ($N = 10,000$).

		No VR			**CCF**			**I-CCF**		
T	S_0	V^*	V	W	V^*	V	W	V^*	V	W
0.5	34	56.60	57.12	56.39	0.47	6.32	0.47	0.44	6.35	0.44
0.5	36	55.71	56.02	55.54	0.40	5.67	0.40	0.36	5.69	0.36
0.5	38	53.50	53.84	53.38	0.34	5.11	0.34	0.29	5.12	0.28
0.5	40	50.48	50.57	50.43	0.29	4.47	0.29	0.23	4.48	0.23
0.5	42	46.98	47.22	46.87	0.24	3.87	0.24	0.18	3.87	0.17
0.5	44	43.22	43.46	43.14	0.20	3.39	0.20	0.13	3.39	0.14
0.5	46	39.43	39.50	39.36	0.17	3.08	0.17	0.10	3.08	0.11
1	34	68.24	68.56	67.92	1.14	6.79	1.12	1.05	6.86	1.04
1	36	67.44	67.81	67.22	1.02	6.37	1.00	0.88	6.40	0.88
1	38	65.87	66.00	65.71	0.91	6.11	0.89	0.74	6.11	0.74
1	40	63.69	64.02	63.56	0.81	5.88	0.80	0.62	5.87	0.62
1	42	61.13	61.41	60.91	0.72	5.58	0.71	0.52	5.56	0.52
1	44	58.51	58.57	58.37	0.64	5.35	0.63	0.43	5.33	0.43
1	46	55.49	55.71	55.34	0.57	4.99	0.56	0.35	4.97	0.36
1.5	34	75.76	75.97	75.43	1.92	7.32	1.90	1.74	7.28	1.73
1.5	36	75.01	75.30	74.73	1.75	6.96	1.73	1.49	6.89	1.49
1.5	38	73.68	74.12	73.39	1.59	6.73	1.58	1.28	6.63	1.28
1.5	40	72.42	72.38	72.17	1.46	6.49	1.45	1.10	6.39	1.10
1.5	42	70.68	70.58	70.46	1.33	6.09	1.32	0.94	6.01	0.94
1.5	44	68.25	68.46	68.14	1.21	5.92	1.20	0.80	5.82	0.80
1.5	46	65.91	66.11	65.72	1.11	5.72	1.09	0.68	5.63	0.69
2	34	79.80	80.39	79.49	2.75	7.94	2.74	2.35	7.71	2.35
2	36	79.71	80.12	79.43	2.56	7.52	2.54	2.05	7.26	2.06
2	38	78.58	78.94	78.53	2.38	7.06	2.36	1.79	6.82	1.79
2	40	77.47	77.80	77.29	2.21	6.89	2.18	1.56	6.65	1.56
2	42	76.02	76.36	75.80	2.04	6.77	2.02	1.35	6.48	1.36
2	44	74.49	74.94	74.34	1.89	6.53	1.87	1.17	6.24	1.18
2	46	72.65	73.03	72.34	1.76	6.26	1.74	1.01	6.03	1.02

Detailed standard error results obtained from $R = 10,000$ replications when using $N = 10,000$ simulation paths. Standard error results are measured for FDM (V^*), LSM (V) and LSM-CCV (W) estimators across three algorithm configurations: no variance reduction technique ("No VR"), CCF control variates ("CCF"), and a combination of importance sampling and CCF control variates ("I-CCF"). Results are reported for put options with maturities $T \in \{0.5, 1, 1.5, 2\}$ and strike price $K = 40$. The underlying asset is governed by a GBM with an initial asset price $S_0 \in \{34, 36, 38, 40, 42, 44, 46\}$, volatility $\sigma = 0.4$, risk-free rate $r = 0.06$, and dividend yield $q = 0.06$.

Table 4.A.4: Detailed standard error results ($\times 10^3$) ($N = 100,000$).

		No VR			**CCF**			**I-CCF**		
T	S_0	V^*	V	W	V^*	V	W	V^*	V	W
0.5	34	17.23	17.35	17.16	0.15	1.37	0.15	0.14	1.39	0.14
0.5	36	17.08	17.02	17.08	0.13	1.34	0.12	0.11	1.37	0.11
0.5	38	16.30	16.45	16.29	0.11	1.21	0.11	0.09	1.22	0.09
0.5	40	15.43	15.50	15.42	0.09	1.10	0.09	0.07	1.10	0.07
0.5	42	14.26	14.24	14.20	0.08	1.03	0.08	0.05	1.04	0.05
0.5	44	12.95	13.01	12.89	0.06	0.96	0.06	0.04	0.95	0.04
0.5	46	11.76	11.71	11.75	0.05	0.84	0.05	0.03	0.84	0.03
1	34	21.59	21.70	21.43	0.35	1.72	0.34	0.33	1.71	0.33
1	36	21.26	21.40	21.18	0.30	1.60	0.30	0.28	1.57	0.28
1	38	20.79	20.81	20.77	0.28	1.49	0.27	0.24	1.47	0.24
1	40	19.76	19.94	19.69	0.25	1.38	0.25	0.20	1.37	0.20
1	42	18.72	18.93	18.66	0.22	1.33	0.22	0.17	1.31	0.16
1	44	17.93	17.93	17.90	0.20	1.27	0.20	0.14	1.25	0.14
1	46	17.06	17.13	17.01	0.18	1.19	0.18	0.11	1.18	0.11
1.5	34	23.94	24.15	23.92	0.59	1.80	0.58	0.55	1.75	0.55
1.5	36	23.55	23.61	23.50	0.54	1.70	0.54	0.47	1.60	0.47
1.5	38	22.91	23.19	22.78	0.50	1.62	0.49	0.40	1.52	0.40
1.5	40	22.24	22.45	22.18	0.45	1.54	0.45	0.35	1.45	0.35
1.5	42	21.62	21.79	21.58	0.42	1.43	0.42	0.30	1.36	0.30
1.5	44	20.88	20.88	20.76	0.38	1.42	0.37	0.25	1.36	0.25
1.5	46	20.12	20.15	20.03	0.35	1.36	0.34	0.21	1.29	0.22
2	34	24.29	24.71	24.30	0.88	1.98	0.86	0.70	1.80	0.71
2	36	23.86	24.22	23.81	0.81	1.94	0.81	0.62	1.73	0.63
2	38	23.50	23.65	23.41	0.75	1.86	0.73	0.54	1.64	0.55
2	40	23.04	23.24	22.88	0.69	1.73	0.68	0.47	1.54	0.47
2	42	22.61	22.58	22.45	0.63	1.56	0.63	0.41	1.40	0.41
2	44	22.27	22.15	22.19	0.60	1.50	0.59	0.36	1.37	0.36
2	46	21.68	21.88	21.51	0.55	1.43	0.54	0.31	1.31	0.31

Detailed standard error results obtained from $R = 1,000$ replications when using $N = 100,000$ simulation paths. Standard error results are measured for FDM (V^*), LSM (V) and LSM-CCV (W) estimators across three algorithm configurations: no variance reduction technique ("No VR"), CCF control variates ("CCF"), and a combination of importance sampling and CCF control variates ("I-CCF"). Results are reported for put options with maturities $T \in \{0.5, 1, 1.5, 2\}$ and strike price $K = 40$. The underlying asset is governed by a GBM with an initial asset price $S_0 \in \{34, 36, 38, 40, 42, 44, 46\}$, volatility $\sigma = 0.4$, risk-free rate $r = 0.06$, and dividend yield $q = 0.06$.

Table 4.A.5: Detailed RMSE results ($\times 10^4$) ($N = 10,000$).

		No VR			**CCF**			**I-CCF**		
T	S_0	V^*	V	W	V^*	V	W	V^*	V	W
0.5	34	566.0	595.1	564.0	4.7	138.3	4.7	4.5	138.0	4.4
0.5	36	557.2	583.8	555.4	4.1	127.7	4.0	3.7	127.5	3.6
0.5	38	535.0	562.2	533.9	3.5	116.2	3.4	2.9	116.2	2.8
0.5	40	504.9	529.4	504.3	3.0	103.7	2.9	2.4	104.0	2.3
0.5	42	469.9	495.7	468.8	2.5	90.9	2.4	1.9	91.3	1.8
0.5	44	432.3	458.5	431.4	2.1	79.7	2.0	1.5	80.4	1.4
0.5	46	394.3	418.4	393.6	1.7	70.5	1.7	1.2	71.5	1.1
1	34	682.5	707.7	679.2	11.4	152.5	11.5	10.5	152.0	10.7
1	36	674.5	700.4	672.2	10.2	146.3	10.2	8.9	146.4	9.0
1	38	658.7	682.6	657.1	9.1	140.8	9.1	7.5	141.4	7.6
1	40	636.9	663.1	635.6	8.1	135.5	8.1	6.3	136.6	6.3
1	42	611.3	638.5	609.1	7.2	129.6	7.2	5.2	131.2	5.3
1	44	585.1	609.8	583.7	6.4	124.5	6.4	4.3	126.5	4.4
1	46	554.9	582.1	553.4	5.7	117.9	5.7	3.6	120.4	3.7
1.5	34	757.6	779.9	754.3	19.2	163.9	19.8	17.4	163.2	18.1
1.5	36	750.1	773.1	747.3	17.5	158.6	17.9	14.9	159.0	15.6
1.5	38	736.8	762.1	733.9	15.9	154.2	16.2	12.8	155.6	13.4
1.5	40	724.2	745.1	721.7	14.6	149.5	14.8	11.0	151.8	11.5
1.5	42	706.8	727.8	704.6	13.3	143.8	13.4	9.4	146.9	9.9
1.5	44	682.5	707.7	681.4	12.1	139.9	12.2	8.0	143.5	8.4
1.5	46	659.1	684.6	657.2	11.1	135.8	11.1	6.8	140.1	7.3
2	34	798.0	823.0	794.9	27.5	173.8	28.7	23.5	173.0	25.2
2	36	797.1	820.6	794.3	25.6	168.6	26.5	20.6	169.4	22.1
2	38	785.8	809.1	785.3	23.8	162.5	24.4	17.9	164.7	19.2
2	40	774.7	798.2	773.0	22.1	158.8	22.5	15.6	161.9	16.8
2	42	760.2	783.8	758.0	20.4	155.8	20.8	13.5	159.7	14.6
2	44	744.9	769.9	743.4	18.9	151.9	19.1	11.7	156.5	12.7
2	46	726.5	752.2	723.4	17.6	147.6	17.8	10.2	153.3	11.1

Detailed RMSE results obtained from $R = 10,000$ replications when using $N = 10,000$ simulation paths. standard error results are measured for FDM (V^*), LSM (V) and LSM-CCV (W) estimators across three algorithm configurations: no variance reduction technique ("No VR"), CCF control variates ("CCF"), and a combination of importance sampling and CCF control variates ("I-CCF"). Results are reported for put options with maturities $T \in \{0.5, 1, 1.5, 2\}$ and strike price $K = 40$. The underlying asset is governed by a GBM with an initial asset price $S_0 \in \{34, 36, 38, 40, 42, 44, 46\}$, volatility $\sigma = 0.4$, risk-free rate $r = 0.06$, and dividend yield $q = 0.06$.

Table 4.A.6: Detailed RMSE results ($\times 10^4$) ($N = 100,000$).

		No VR			**CCF**			**I-CCF**		
T	S_0	V^*	V	W	V^*	V	W	V^*	V	W
0.5	34	172.4	173.5	171.6	1.5	41.6	1.5	1.5	41.7	1.4
0.5	36	170.8	170.3	170.8	1.4	41.4	1.2	1.3	41.5	1.1
0.5	38	163.0	164.5	162.9	1.3	38.8	1.1	1.1	38.8	0.9
0.5	40	154.4	155.0	154.2	1.2	35.6	1.0	1.0	35.7	0.8
0.5	42	142.7	142.4	142.1	1.0	32.4	0.9	0.9	32.5	0.7
0.5	44	129.7	130.4	129.1	0.9	28.9	0.8	0.7	29.0	0.6
0.5	46	117.8	117.7	117.7	0.7	25.4	0.6	0.6	25.6	0.5
1	34	216.2	217.6	214.8	3.5	44.8	4.2	3.3	44.8	4.1
1	36	212.8	214.4	212.2	3.1	43.5	3.5	2.8	43.6	3.4
1	38	208.2	208.5	208.1	2.8	41.9	3.0	2.4	42.0	2.8
1	40	197.9	199.9	197.2	2.6	40.5	2.6	2.1	40.8	2.3
1	42	187.4	189.7	187.0	2.3	38.8	2.3	1.7	39.1	1.8
1	44	179.4	179.5	179.2	2.1	37.2	2.1	1.4	37.5	1.5
1	46	170.7	171.4	170.2	1.9	34.9	1.8	1.2	35.3	1.2
1.5	34	240.1	242.6	240.4	5.9	43.9	7.7	5.5	43.9	7.7
1.5	36	236.2	237.0	236.0	5.4	42.4	6.8	4.7	42.5	6.5
1.5	38	229.5	232.6	228.7	5.0	41.5	6.0	4.0	41.6	5.5
1.5	40	222.7	224.9	222.3	4.5	40.2	5.3	3.5	40.5	4.7
1.5	42	216.5	218.1	216.3	4.2	39.0	4.8	3.0	39.5	3.9
1.5	44	209.0	209.0	208.0	3.8	38.3	4.2	2.5	38.9	3.3
1.5	46	201.5	201.6	200.7	3.5	36.9	3.7	2.2	37.5	2.8
2	34	243.4	247.8	244.2	8.8	46.6	12.3	7.0	46.3	11.6
2	36	239.2	242.9	239.5	8.1	45.0	11.0	6.2	44.7	10.0
2	38	235.6	237.3	235.3	7.5	43.4	9.7	5.4	43.3	8.7
2	40	231.1	233.2	230.0	6.9	41.6	8.7	4.7	41.7	7.4
2	42	227.0	226.5	226.0	6.3	39.9	7.7	4.1	40.2	6.3
2	44	223.6	222.2	223.2	6.0	38.6	6.9	3.6	39.3	5.4
2	46	217.5	219.4	216.2	5.5	37.2	6.3	3.1	38.0	4.7

Detailed RMSE results obtained from $R = 1,000$ replications when using $N = 100,000$ simulation paths. RMSE results are measured for FDM (V^*), LSM (V) and LSM-CCV (W) estimators across three algorithm configurations: no variance reduction technique ("No VR"), CCF control variates ("CCF"), and a combination of importance sampling and CCF control variates ("I-CCF"). Results are reported for put options with maturities $T \in \{0.5, 1, 1.5, 2\}$ and strike price $K = 40$. The underlying asset is governed by a GBM with an initial asset price $S_0 \in \{34, 36, 38, 40, 42, 44, 46\}$, volatility $\sigma = 0.4$, risk-free rate $r = 0.06$, and dividend yield $q = 0.06$.

Table 4.A.7: Detailed efficiency results ($N = 10,000$).

		CCF			**I**			**I-CCF**		
T	S_0	V^*	V	W	V^*	V	W	V^*	V	W
0.5	34	118.9	8.0	119.9	0.6	0.6	0.6	126.8	8.0	127.4
0.5	36	136.7	8.9	138.2	0.7	0.7	0.7	153.4	8.8	154.6
0.5	38	155.1	9.5	157.0	0.8	0.8	0.8	185.9	9.5	186.5
0.5	40	173.4	10.3	175.3	0.9	0.9	0.9	221.9	10.3	222.5
0.5	42	192.3	11.2	193.8	1.0	1.0	0.9	267.3	11.2	267.1
0.5	44	213.3	11.8	215.1	1.0	1.1	1.0	319.6	11.8	317.5
0.5	46	234.1	11.8	236.1	1.2	1.2	1.2	380.5	11.8	372.8
1	34	59.1	9.1	59.6	0.9	0.9	0.9	64.1	9.0	64.4
1	36	65.4	9.6	66.0	1.0	1.0	1.0	75.3	9.6	75.3
1	38	71.5	9.8	72.5	1.1	1.1	1.1	87.6	9.8	87.5
1	40	77.4	9.9	78.3	1.2	1.2	1.2	101.2	9.9	101.0
1	42	83.5	10.0	84.4	1.3	1.3	1.3	116.8	10.1	116.3
1	44	90.4	9.9	91.1	1.4	1.3	1.4	134.9	10.0	133.7
1	46	96.6	10.2	97.5	1.5	1.4	1.5	155.5	10.2	153.2
1.5	34	38.5	9.4	38.6	1.2	1.2	1.2	42.6	9.4	42.6
1.5	36	41.9	9.8	42.1	1.2	1.2	1.2	49.4	9.9	49.1
1.5	38	45.3	10.0	45.6	1.3	1.3	1.3	56.5	10.2	56.2
1.5	40	48.5	10.1	48.9	1.4	1.4	1.4	65.1	10.3	64.5
1.5	42	52.0	10.6	52.5	1.5	1.4	1.5	74.5	10.8	73.8
1.5	44	55.2	10.6	55.7	1.5	1.5	1.5	84.5	10.8	83.7
1.5	46	58.5	10.6	59.1	1.6	1.6	1.6	95.8	10.7	94.4
2	34	28.0	9.1	28.0	1.3	1.3	1.3	32.9	9.4	32.8
2	36	30.1	9.6	30.3	1.3	1.3	1.3	37.8	10.0	37.6
2	38	32.0	10.2	32.2	1.4	1.4	1.4	42.9	10.6	42.8
2	40	34.1	10.3	34.5	1.4	1.4	1.4	48.8	10.7	48.4
2	42	36.2	10.3	36.5	1.5	1.5	1.5	55.2	10.8	54.7
2	44	38.5	10.5	38.8	1.5	1.5	1.5	62.5	11.0	61.9
2	46	40.4	10.7	40.7	1.6	1.6	1.6	70.7	11.1	69.6

Detailed efficiency results for control variates, importance sampling, and control variates with importance sampling, obtained from $R = 10,000$ replications when using $N = 10,000$ simulation paths. Efficiency results are measured for FDM (V^*), LSM (V) and LSM-CCV (W) estimators across three algorithm configurations: CCF control variates ("CCF"), importance sampling ("I"), and a combination of importance sampling and CCF control variates ("I-CCF"). Results are reported for put options with maturities $T \in \{0.5, 1, 1.5, 2\}$ and strike price $K = 40$. The underlying asset is governed by a GBM with an initial asset price $S_0 \in \{34, 36, 38, 40, 42, 44, 46\}$, volatility $\sigma = 0.4$, risk-free rate $r = 0.06$, and dividend yield $q = 0.06$.

Table 4.A.8: Detailed efficiency results ($N = 100,000$).

		CCF			**I**			**I-CCF**		
T	S_0	V^*	V	W	V^*	V	W	V^*	V	W
0.5	34	116.7	11.7	115.2	0.6	0.6	0.6	122.2	11.5	123.6
0.5	36	134.6	11.7	136.7	0.7	0.7	0.7	152.4	11.5	153.7
0.5	38	150.0	12.6	151.2	0.8	0.8	0.8	183.9	12.5	184.8
0.5	40	168.6	13.1	171.2	0.9	0.9	0.9	217.8	13.0	218.8
0.5	42	185.2	12.8	186.2	0.9	0.9	0.9	261.3	12.7	262.1
0.5	44	201.9	12.6	201.4	1.0	1.0	1.0	307.2	12.6	308.6
0.5	46	219.9	12.9	222.8	1.1	1.0	1.1	367.3	13.0	367.7
1	34	60.8	11.6	61.7	1.0	1.0	1.0	64.6	11.7	64.5
1	36	69.1	12.4	68.7	1.0	1.1	1.0	74.4	12.6	74.6
1	38	74.2	12.9	75.8	1.1	1.1	1.1	86.3	13.1	86.6
1	40	78.2	13.5	79.3	1.2	1.2	1.2	98.2	13.6	98.2
1	42	82.6	13.3	83.8	1.2	1.2	1.2	112.3	13.4	112.2
1	44	87.5	13.1	87.5	1.4	1.3	1.4	130.5	13.3	129.8
1	46	92.5	13.4	93.9	1.4	1.5	1.4	149.8	13.6	149.6
1.5	34	39.8	12.4	40.6	1.2	1.2	1.2	42.6	12.8	42.8
1.5	36	42.4	12.9	42.8	1.2	1.3	1.2	49.1	13.7	49.0
1.5	38	45.1	13.3	45.2	1.3	1.3	1.2	55.8	14.3	55.5
1.5	40	48.4	13.6	48.4	1.3	1.4	1.3	63.2	14.5	62.9
1.5	42	51.0	14.2	51.0	1.4	1.4	1.4	72.1	15.0	71.8
1.5	44	54.2	13.8	54.7	1.4	1.5	1.4	82.4	14.4	81.3
1.5	46	56.5	13.9	57.4	1.5	1.5	1.5	92.9	14.6	92.1
2	34	26.8	11.5	27.1	1.1	1.2	1.1	33.6	12.8	33.4
2	36	28.3	11.5	28.5	1.1	1.2	1.1	37.3	13.0	37.1
2	38	30.5	11.7	30.9	1.2	1.2	1.2	42.6	13.4	41.8
2	40	32.2	12.5	32.5	1.2	1.2	1.2	47.9	14.1	47.4
2	42	34.7	13.5	34.8	1.3	1.3	1.3	54.3	15.1	53.6
2	44	36.3	13.8	36.8	1.3	1.3	1.3	61.7	15.2	61.0
2	46	38.6	14.3	38.5	1.4	1.4	1.3	69.2	15.8	68.0

Detailed efficiency results for control variates, importance sampling, and control variates with importance sampling, obtained from $R = 1,000$ replications when using $N = 100,000$ simulation paths. Efficiency results are measured for FDM (V^*), LSM (V) and LSM-CCV (W) estimators across three algorithm configurations: CCF control variates ("CCF"), importance sampling ("I"), and a combination of importance sampling and CCF control variates ("I-CCF"). Results are reported for put options with maturities $T \in \{0.5, 1, 1.5, 2\}$ and strike price $K = 40$. The underlying asset is governed by a GBM with an initial asset price $S_0 \in \{34, 36, 38, 40, 42, 44, 46\}$, volatility $\sigma = 0.4$, risk-free rate $r = 0.06$, and dividend yield $q = 0.06$.

Table 4.A.9: Detailed efficiency results with LSM-s ($N = 10,000$).

		CCF			**s**			**s-CCF**		
T	S_0	V^*	V	W	V^*	V	W	V^*	V	W
0.5	34	118.9	8.0	119.9	0.6	0.6	0.6	126.8	18.6	126.7
0.5	36	136.7	8.9	138.2	0.7	0.7	0.7	153.4	21.4	153.7
0.5	38	155.1	9.5	157.0	0.8	0.8	0.8	185.9	24.3	184.9
0.5	40	173.4	10.3	175.3	0.9	0.9	0.9	221.9	25.8	221.1
0.5	42	192.3	11.2	193.8	1.0	0.9	0.9	267.3	28.1	265.3
0.5	44	213.3	11.8	215.1	1.0	1.0	1.0	319.6	30.3	317.4
0.5	46	234.1	11.8	236.1	1.2	1.1	1.2	380.5	32.8	375.9
1	34	59.1	9.1	59.6	0.9	0.9	0.9	64.1	24.8	63.5
1	36	65.4	9.6	66.0	1.0	1.0	1.0	75.3	26.7	74.4
1	38	71.5	9.8	72.5	1.1	1.1	1.1	87.6	27.0	86.6
1	40	77.4	9.9	78.3	1.2	1.2	1.2	101.2	26.8	99.7
1	42	83.5	10.0	84.4	1.3	1.3	1.3	116.8	26.2	115.0
1	44	90.4	9.9	91.1	1.4	1.3	1.4	134.9	25.3	130.7
1	46	96.6	10.2	97.5	1.5	1.4	1.5	155.5	24.7	145.4
1.5	34	38.5	9.4	38.6	1.2	1.1	1.2	42.6	27.1	42.1
1.5	36	41.9	9.8	42.1	1.2	1.2	1.2	49.4	28.6	48.7
1.5	38	45.3	10.0	45.6	1.3	1.3	1.3	56.5	29.8	55.6
1.5	40	48.5	10.1	48.9	1.4	1.4	1.4	65.1	29.5	63.3
1.5	42	52.0	10.6	52.5	1.5	1.4	1.5	74.5	28.4	69.5
1.5	44	55.2	10.6	55.7	1.5	1.5	1.5	84.5	27.3	71.4
1.5	46	58.5	10.6	59.1	1.6	1.6	1.6	95.8	26.4	69.5
2	34	28.0	9.1	28.0	1.3	1.3	1.3	32.9	25.1	32.5
2	36	30.1	9.6	30.3	1.3	1.3	1.3	37.8	27.4	36.8
2	38	32.0	10.2	32.2	1.4	1.4	1.4	42.9	29.1	41.1
2	40	34.1	10.3	34.5	1.4	1.4	1.4	48.8	29.8	43.7
2	42	36.2	10.3	36.5	1.5	1.5	1.5	55.2	30.0	42.8
2	44	38.5	10.5	38.8	1.5	1.5	1.5	62.5	29.8	39.4
2	46	40.4	10.7	40.7	1.6	1.6	1.6	70.7	28.4	34.2

Detailed efficiency results for control variates, importance sampling with shifted regressions, and a combination of the two, obtained from $R = 10,000$ replications when using $N = 10,000$ simulation paths. Efficiency results are measured for FDM (V^*), LSM (V) and LSM-CCV (W) estimators across three algorithm configurations: CCF control variates ("CCF"), Boire et al. (2021*a*) importance sampling ("s"), and a combination of Boire et al. (2021*a*) importance sampling and CCF control variates ("s-CCF"). Results are reported for put options with maturities $T \in \{0.5, 1, 1.5, 2\}$ and strike price $K = 40$. The underlying asset is governed by a GBM with an initial asset price $S_0 \in \{34, 36, 38, 40, 42, 44, 46\}$, volatility $\sigma = 0.4$, risk-free rate $r = 0.06$, and dividend yield $q = 0.06$.

Table 4.A.10: Detailed efficiency results with LSM-s ($N = 100,000$).

		CCF			**I**			**I-CCF**		
T	S_0	V^*	V	W	V^*	V	W	V^*	V	W
0.5	34	116.7	11.7	115.2	0.6	0.6	0.6	122.2	21.1	122.4
0.5	36	134.6	11.7	136.7	0.7	0.7	0.7	152.4	24.6	151.6
0.5	38	150.0	12.6	151.2	0.8	0.8	0.8	183.9	26.7	182.9
0.5	40	168.6	13.1	171.2	0.9	0.9	0.9	217.8	30.3	217.7
0.5	42	185.2	12.8	186.2	0.9	0.9	0.9	261.3	33.9	259.6
0.5	44	201.9	12.6	201.4	1.0	1.0	1.0	307.2	37.3	306.3
0.5	46	219.9	12.9	222.8	1.1	1.1	1.1	367.3	40.0	365.6
1	34	60.8	11.6	61.7	1.0	1.0	1.0	64.6	32.9	63.3
1	36	69.1	12.4	68.7	1.0	1.0	1.0	74.4	35.7	73.4
1	38	74.2	12.9	75.8	1.1	1.1	1.1	86.3	37.6	85.6
1	40	78.2	13.5	79.3	1.2	1.2	1.2	98.2	41.3	97.6
1	42	82.6	13.3	83.8	1.2	1.2	1.2	112.3	43.9	111.7
1	44	87.5	13.1	87.5	1.4	1.3	1.4	130.5	46.8	129.3
1	46	92.5	13.4	93.9	1.4	1.4	1.4	149.8	51.2	148.6
1.5	34	39.8	12.4	40.6	1.2	1.2	1.2	42.6	34.4	42.2
1.5	36	42.4	12.9	42.8	1.2	1.2	1.2	49.1	37.3	48.4
1.5	38	45.1	13.3	45.2	1.3	1.3	1.2	55.8	41.8	55.0
1.5	40	48.4	13.6	48.4	1.3	1.3	1.3	63.2	45.0	62.8
1.5	42	51.0	14.2	51.0	1.4	1.4	1.4	72.1	48.4	70.8
1.5	44	54.2	13.8	54.7	1.4	1.4	1.4	82.4	51.2	80.0
1.5	46	56.5	13.9	57.4	1.5	1.5	1.5	92.9	54.3	85.1
2	34	26.8	11.5	27.1	1.1	1.1	1.1	33.6	29.5	33.5
2	36	28.3	11.5	28.5	1.1	1.1	1.1	37.3	32.7	37.0
2	38	30.5	11.7	30.9	1.2	1.2	1.2	42.6	36.1	41.9
2	40	32.2	12.5	32.5	1.2	1.2	1.2	47.9	40.6	47.3
2	42	34.7	13.5	34.8	1.3	1.3	1.3	54.3	43.9	50.5
2	44	36.3	13.8	36.8	1.3	1.3	1.3	61.7	47.4	54.6
2	46	38.6	14.3	38.5	1.4	1.4	1.3	69.2	52.9	41.2

Detailed efficiency results for control variates, importance sampling with shifted regressions, and a combination of the two, obtained from $R = 1,000$ replications when using $N = 100,000$ simulation paths. Efficiency results are measured for FDM (V^*), LSM (V) and LSM-CCV (W) estimators across three algorithm configurations: CCF control variates ("CCF"), Boire et al. (2021*a*) importance sampling ("s"), and a combination of Boire et al. (2021*a*) importance sampling and CCF control variates ("s-CCF"). Results are reported for put options with maturities $T \in \{0.5, 1, 1.5, 2\}$ and strike price $K = 40$. The underlying asset is governed by a GBM with an initial asset price $S_0 \in \{34, 36, 38, 40, 42, 44, 46\}$, volatility $\sigma = 0.4$, risk-free rate $r = 0.06$, and dividend yield $q = 0.06$.

Table 4.A.11: Detailed results for the ratio of standard errors ($N = 10,000$).

		CCF		**I**		**I-CCF**	
T	S_0	V	W	V	W	V	W
0.5	34	1.009	0.996	13.390	0.989	14.338	0.991
0.5	36	1.006	0.997	14.030	0.986	15.785	0.990
0.5	38	1.006	0.998	14.912	0.986	17.880	0.994
0.5	40	1.002	0.999	15.446	0.988	19.783	0.996
0.5	42	1.005	0.998	15.916	0.990	22.076	0.998
0.5	44	1.005	0.998	16.795	0.990	25.118	1.004
0.5	46	1.002	0.998	18.350	0.990	29.774	1.019
1	34	1.005	0.995	5.980	0.986	6.549	0.990
1	36	1.006	0.997	6.269	0.988	7.237	0.996
1	38	1.002	0.998	6.728	0.984	8.227	0.999
1	40	1.005	0.998	7.236	0.986	9.417	1.000
1	42	1.005	0.996	7.709	0.986	10.708	1.001
1	44	1.001	0.998	8.363	0.991	12.385	1.006
1	46	1.004	0.997	8.773	0.988	14.008	1.012
1.5	34	1.003	0.996	3.814	0.992	4.194	0.996
1.5	36	1.004	0.996	3.983	0.991	4.625	1.001
1.5	38	1.006	0.996	4.229	0.990	5.176	1.003
1.5	40	0.999	0.997	4.435	0.988	5.836	1.007
1.5	42	0.999	0.997	4.572	0.988	6.414	1.006
1.5	44	1.003	0.998	4.877	0.989	7.286	1.008
1.5	46	1.003	0.997	5.163	0.987	8.270	1.011
2	34	1.007	0.996	2.884	0.994	3.281	0.999
2	36	1.005	0.996	2.940	0.993	3.535	1.002
2	38	1.005	0.999	2.964	0.992	3.812	1.001
2	40	1.004	0.998	3.122	0.987	4.272	1.005
2	42	1.004	0.997	3.319	0.991	4.788	1.006
2	44	1.006	0.998	3.458	0.988	5.323	1.008
2	46	1.005	0.996	3.564	0.988	5.951	1.010

Detailed standard error ratio results obtained from $R = 10,000$ replications when using $N = 10,000$ simulation paths. Ratio results are measured for LSM (V) and LSM-CCV (W) estimators across three algorithm configurations: CCF control variates ("CCF"), importance sampling ("I"), and a combination of importance sampling and CCF control variates ("I-CCF"). Results are reported for put options with maturities $T \in \{0.5, 1, 1.5, 2\}$ and strike price $K = 40$. The underlying asset is governed by a GBM with an initial asset price $S_0 \in \{34, 36, 38, 40, 42, 44, 46\}$, volatility $\sigma = 0.4$, risk-free rate $r = 0.06$, and dividend yield $q = 0.06$.

Table 4.A.12: Detailed results for the ratio of standard errors ($N = 100,000$).

		CCF		**I**		**I-CCF**	
T	S_0	V	W	V	W	V	W
0.5	34	1.007	0.996	9.337	1.009	9.943	0.984
0.5	36	0.997	1.000	10.640	0.984	12.258	0.991
0.5	38	1.009	1.000	11.243	0.991	13.875	0.995
0.5	40	1.004	0.999	12.041	0.984	15.641	0.994
0.5	42	0.999	0.996	13.492	0.990	19.053	0.993
0.5	44	1.004	0.995	14.958	0.998	22.714	0.991
0.5	46	0.996	1.000	15.783	0.987	26.203	0.999
1	34	1.005	0.993	4.914	0.978	5.202	0.994
1	36	1.007	0.996	5.269	1.003	5.581	0.994
1	38	1.001	0.999	5.402	0.979	6.176	0.996
1	40	1.010	0.997	5.525	0.983	6.870	0.997
1	42	1.012	0.997	5.929	0.983	7.943	0.998
1	44	1.000	0.999	6.287	0.998	9.195	1.004
1	46	1.004	0.997	6.535	0.982	10.400	0.998
1.5	34	1.009	0.999	3.068	0.979	3.179	0.994
1.5	36	1.003	0.998	3.126	0.989	3.411	0.999
1.5	38	1.013	0.995	3.254	0.992	3.767	1.000
1.5	40	1.009	0.997	3.415	0.997	4.178	1.001
1.5	42	1.008	0.998	3.445	0.998	4.596	1.002
1.5	44	1.000	0.995	3.742	0.986	5.427	1.008
1.5	46	1.002	0.996	3.879	0.982	6.026	1.005
2	34	1.017	1.000	2.258	0.987	2.557	1.006
2	36	1.015	0.998	2.390	0.993	2.783	1.004
2	38	1.006	0.996	2.490	0.983	3.054	1.016
2	40	1.009	0.993	2.488	0.984	3.276	1.003
2	42	0.999	0.993	2.459	0.990	3.420	1.005
2	44	0.995	0.996	2.511	0.983	3.845	1.007
2	46	1.009	0.992	2.610	0.997	4.231	1.009

Detailed standard error ratio results obtained from $R = 1,000$ replications when using $N = 100,000$ simulation paths. Ratio results are measured for (V) and LSM-CCV (W) estimators across three algorithm configurations: CCF control variates ("CCF"), importance sampling ("I"), and a combination of importance sampling and CCF control variates ("I-CCF"). Results are reported for put options with maturities $T \in \{0.5, 1, 1.5, 2\}$ and strike price $K = 40$. The underlying asset is governed by a GBM with an initial asset price $S_0 \in \{34, 36, 38, 40, 42, 44, 46\}$, volatility $\sigma = 0.4$, risk-free rate $r = 0.06$, and dividend yield $q = 0.06$.

Conclusion

In conclusion, the four essays presented in this thesis study bias and variance reduction techniques applied to the Least-Squares Monte Carlo (LSM) algorithm of Longstaff and Schwartz (2001) for pricing American options. We shed light on the mechanisms that de-stabilize the solutions estimated in the dynamic programming algorithm by emphasizing the crucial role of the continuation value estimation approach. These insights first allow us to make several recommendations about the efficient implementation of variance reduction techniques. Furthermore, we develop new LSM methods that successfully reduce bias and variance across a range of option characteristics and underlying asset dynamics. Every technique proposed in this thesis takes advantage in some way or other of the benefits of improving the early-exercise strategy.

In Chapter 1, call prices are replaced by symmetric put prices to improve the efficiency of variance reduction techniques. Indeed, the bounded payoff of the put option facilitates the estimation of an exercise rule, leading to reduced estimator bias and variance compared to calls. Moreover, the variance reduction efficiency gains resulting from the improved estimated strategy agree with the discussion in Chapter 4, according to which exercise errors stifle the efficiency gains of variance reduction techniques.

In Chapter 2, we propose a new importance sampling approach and provide a detailed discussion on the challenges related to optimizing the change of measure. By performing regressions on the shifted paths, the change of measure does more than just reduce the sampling variance. Indeed, our approach allows the change of measure to optimize the strategy too. Our new approach improves the strategy in many ways. First, regression coefficients are estimated with a larger sample. Second, the strategy is estimated with a higher frequency of paths near the optimal exercise boundary, thereby improving the fit in that crucial region. As a result, our proposed extension mitigates the bias plaguing estimators that use standard importance sampling.

In Chapter 3, we derive explicit approximations to the local LSM bias based on the statistical properties of estimated continuation values. Specifically, we discern two conflicting sources of bias: the negative sub-optimality bias caused by regression estimation errors, and the positive foresight bias stemming from the dependence between the estimated strategy and estimated option values. The bias expressions derived in this paper show that both sources of bias are increasing functions of the variance of continuation values, consistent with the idea that an accurate strategy is key to reducing bias. These approximations are then used to correct the bias of option values at every time step,

Finally, Chapter 4 emphasizes the important effect of random exercise errors on the variance of LSM estimators. Our findings suggest that eliminating these errors reduces the bias, boosts the efficiency of variance reduction techniques by orders of magnitude, and makes the combination of importance sampling and control variates efficient. To eliminate stopping time errors, we recommend using the corrected early-exercise strategy introduced by Rasmussen (2005) and demonstrate that this method indeed yields the same variance reduction efficiency that would be obtained with the optimal strategy.

Bibliography

Abramowitz, M. & Stegun, I. A. (1948), *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Vol. 55, US Government Printing Office.

Barone-Adesi, G. & Whaley, R. E. (1987), 'Efficient Analytic Approximation of American Option Values', *The Journal of Finance* **42**(2), 301–320.

Barraquand, J. & Martineau, D. (1995), 'Numerical Valuation of High Dimensional Multivariate American Securities', *Journal of Financial and Quantitative Analysis* pp. 383–405.

Black, F. & Scholes, M. (1973), 'The Pricing of Options and Corporate Liabilities', *Journal of Political Economy* **81**(3), 637–654.

Boire, F.-M., Reesor, R. M. & Stentoft, L. (2021*a*), 'American Option Pricing with Importance Sampling and Shifted Regressions', *Journal of Risk and Financial Management* **14**(8), 340.

Boire, F.-M., Reesor, R. M. & Stentoft, L. (2021*b*), 'Bias Correction in the Least-Squares Monte Carlo Algorithm', *Working paper* .

Boire, F.-M., Reesor, R. M. & Stentoft, L. (2021*c*), 'Efficient Variance Reduction for American Call Options Using Symmetry Arguments', *Journal of Risk and Financial Management* **14**(11), 504.

Bolia, N. & Juneja, S. (2005), 'Monte Carlo Methods for Pricing Financial Options', *Sadhana* **30**(2-3), 347–385.

Bolia, N., Juneja, S. & Glasserman, P. (2004), Function-Approximation-Based Importance Sampling for Pricing American Options, *in* 'Proceedings of the 2004 Winter Simulation Conference, 2004.', Vol. 1, IEEE.

Boyle, P., Broadie, M. & Glasserman, P. (1997), 'Monte Carlo Methods for Security Pricing', *Journal of Economic Dynamics and Control* **21**(8-9), 1267–1321.

Boyle, P. P. (1977), 'Options: A Monte Carlo Approach', *Journal of Financial Economics* **4**(3), 323–338.

Brealey, R. A., Myers, S. C., Allen, F. & Mohanty, P. (2018), *Principles of Corporate Finance, 12/e*, Vol. 12, McGraw-Hill Education.

Brennan, M. J. & Schwartz, E. S. (1978), 'Finite Difference Methods and Jump Processes Arising in the Pricing of Contingent Claims: A Synthesis', *Journal of Financial and Quantitative Analysis* **13**(3), 461–474.

Broadie, M. & Cao, M. (2008), 'Improved Lower and Upper Bound Algorithms for Pricing American Options by Simulation', *Quantitative Finance* **8**(8), 845–861.

Broadie, M. & Glasserman, P. (1995), A Pruned and Bootstrapped American Option Simulator, *in* 'Winter Simulation Conference Proceedings, 1995.', IEEE, pp. 229–235.

Broadie, M. & Glasserman, P. (1997), 'Pricing American-Style Securities using Simulation', *Journal of Economic Dynamics and Control* **21**(8-9), 1323–1352.

Broadie, M. & Glasserman, P. (2004), 'A Stochastic Mesh Method for Pricing High-Dimensional American Options', *Journal of Computational Finance* **7**, 35–72.

Broadie, M., Glasserman, P. & Jain, G. (1997), 'Enhanced Monte Carlo Estimates for American Option Prices', *Journal of Derivatives* **5**, 25–44.

Carriere, J. F. (1996), 'Valuation of the Early-Exercise Price for Options Using Simulations and Nonparametric Regression', *Insurance: Mathematics and Economics* **19**(1), 19–30.

Chau, K. W. & Oosterlee, C. W. (2019), 'Stochastic Grid Bundling Method for Backward Stochastic Differential Equations', *International Journal of Computer Mathematics* **96**(11), 2272–2301.

Clément, E., Lamberton, D. & Protter, P. (2002), 'An Analysis of a Least Squares Regression Method for American Option Pricing', *Finance and Stochastics* **6**(4), 449–471.

Cox, J. C., Ingersoll, J. E. & Ross, S. A. (1977), 'A Theory of the Term Structure of Interest Rates and the Valuation of Interest-Dependent Claims', *Journal of Financial and Quantitative Analysis* **12**(4), 661–661.

Cox, J. C., Ross, S. A. & Rubinstein, M. (1979), 'Option Pricing: A Simplified Approach', *Journal of Financial Economics* **7**(3), 229–263.

Detemple, J. (2001), 'American Options: Symmetry Properties', *Option Pricing, Interest Rates and Risk Management* pp. 67–104.

Duan, J.-C. & Simonato, J.-G. (1998), 'Empirical Martingale Simulation for Asset Prices', *Management Science* **44**(9), 1218–1233.

Efron, B. (1982), *The Jackknife, the Bootstrap and Other Resampling Plans*, Philadelphia: SIAM.

Fabozzi, F. J., Paletta, T. & Tunaru, R. (2017), 'An Improved Least Squares Monte Carlo Valuation Method Based on Heteroscedasticity', *European Journal of Operational Research* **263**(2), 698–706.

Friedman, J., Hastie, T. & Tibshirani, R. (2001), *The Elements of Statistical Learning*, Vol. 1, Springer Series in Statistics New York.

Geske, R. & Johnson, H. E. (1984), 'The American Put Option Valued Analytically', *The Journal of Finance* **39**(5), 1511–1524.

Glasserman, P. (2003), *Monte Carlo Methods in Financial Engineering*, Vol. 53, Springer Science & Business Media.

Glasserman, P., Heidelberger, P. & Shahabuddin, P. (1999), 'Asymptotically Optimal Importance Sampling and Stratification for Pricing Path-Dependent Options', *Mathematical Finance* **9**(2), 117–152.

Heath, D., Jarrow, R. & Morton, A. (1990), 'Bond Pricing and the Term Structure of Interest Rates: A Discrete Time Approximation', *Journal of Financial and Quantitative Analysis* **25**(4), 419–440.

Hinkley, D. V. (1977), 'Jackknifing in Unbalanced Situations', *Technometrics* **19**(3), 285–292.

Horn, S. D., Horn, R. A. & Duncan, D. B. (1975), 'Estimating Heteroscedastic Variances in Linear Models', *Journal of the American Statistical Association* **70**(350), 380–385.

Hull, J. & White, A. (1987), 'The Pricing of Options on Assets with Stochastic Volatilities', *The Journal of Finance* **42**(2), 281–300.

Juneja, S. & Kalra, H. (2009), 'Variance Reduction Techniques for Pricing American Options Using Function Approximations', *Journal of Computational Finance* **12**(3), 79.

Kan, K. F., Reesor, R. M., Whitehead, T. & Davison, M. (2009), Correcting the Bias in Monte Carlo Estimators of American-Style Option Values, *in* 'Monte Carlo and Quasi-Monte Carlo Methods 2008', Springer, pp. 439–454.

Kan, K. H. F., Frank, G., Mozgin, V. & Reesor, M. (2010), 'Optimized Least-squares Monte Carlo (OLSM) for Measuring Counterparty Credit Exposure of American-Style Options'.

Kan, K. H. & Reesor, R. M. (2012), 'Bias Reduction for Pricing American Options by Least-Squares Monte Carlo', *Applied Mathematical Finance* **19**(3), 195–217.

Lemieux, C. & La, J. (2005), A Study of Variance Reduction Techniques for American Option Pricing, *in* 'Proceedings of the Winter Simulation Conference, 2005.', IEEE, pp. 8–pp.

Létourneau, P. & Stentoft, L. (2014), 'Refining the Least Squares Monte Carlo Method by Imposing Structure', *Quantitative Finance* **14**(3), 495–507.

Létourneau, P. & Stentoft, L. (2019), 'Bootstrapping the Early Exercise Boundary in the Least-Squares Monte Carlo Method', *Journal of Risk and Financial Management* **12**(4), 190.

Longstaff, F. A. & Schwartz, E. S. (2001), 'Valuing American Options by Simulation: A Simple Least-Squares Approach', *The Review of Financial Studies* **14**(1), 113–147.

MacKinnon, J. G. & White, H. (1985), 'Some Heteroskedasticity-Consistent Covariance Matrix Estimators with Improved Finite Sample Properties', *Journal of Econometrics* **29**(3), 305–325.

MacMillan, L. W. (1986), 'Analytic Approximation for the American Put Option', *Advances in futures and options research* **1**(1), 119–139.

McDonald, R. & Schroder, M. (1998), 'A Parity Result for American Options', *Journal of Computational Finance* **1**(3), 5–13.

Merton, R. C. (1973), 'Theory of Rational Option Pricing', *The Bell Journal of Economics and Management Science* pp. 141–183.

Merton, R. C. (1976), 'Option Pricing when Underlying Stock Returns are Discontinuous', *Journal of Financial Economics* **3**(1-2), 125–144.

Morales, M. (2006), 'Implementing Importance Sampling in the Least-Squares Monte Carlo Approach for American options', *University of Montreal, Montreal* .

Moreni, N. (2003), 'Pricing American Options: A Variance Reduction Technique for the Longstaff-Schwartz Algorithm', *Research report 2003-256* .

Moreni, N. (2004), 'A Variance Reduction Technique for American Option Pricing', *Physica A: Statistical Mechanics and its Applications* **338**(1-2), 292–295.

Omberg, E. (1987), 'The Valuation of American Put Options with Exponential Exercise Policies', *Advances in Futures and Options Research* **2**(1), 117–142.

Rasmussen, N. S. (2005), 'Control Variates for Monte Carlo Valuation of American Options', *Journal of Computational Finance* **9**(1).

Reider, R. L. (1994), Two Applications of Monte Carlo Techniques to Finance, PhD thesis, University of Pennsylvania.

Royden, H. L. (1988), *Real analysis*, Vol. 32, Macmillan New York.

Stentoft, L. (2004*a*), 'Assessing the Least Squares Monte-Carlo Approach to American Option Valuation', *Review of Derivatives Research* **7**(2), 129–168.

Stentoft, L. (2004*b*), 'Convergence of the Least Squares Monte Carlo Approach to American Option Valuation', *Management Science* **50**(9), 1193–1203.

Stentoft, L. (2014), 'Value Function Approximation or Stopping Time Approximation: A Comparison of Two Recent Numerical Methods for American Option Pricing using Simulation and Regression', *Journal of Computational Finance* **18**(1).

Stentoft, L. (2019), 'Efficient Numerical Pricing of American Call Options Using Symmetry Arguments', *Journal of Risk and Financial Management* **12**(2), 59.

Su, Y. & Fu, M. C. (2000), Importance Sampling in Derivative Securities Pricing, *in* '2000 Winter Simulation Conference Proceedings (Cat. No. 00CH37165)', Vol. 1, IEEE, pp. 587–596.

Tian, T. & Burrage, K. (2002), 'Accuracy Issues of Monte-Carlo Methods for Valuing American Options', *ANZIAM Journal* **44**, C739–C758.

Tilley, J. A. (1993), Valuing American Options in a Path Simulation Model, *in* 'Transactions of the Society of Actuaries', Citeseer.

Tsitsiklis, J. N. & Van Roy, B. (2001), 'Regression Methods for Pricing Complex American-Style Options', *IEEE Transactions on Neural Networks* **12**(4), 694–703.

White, H. (1980), 'A Heteroskedasticity-Consistent Covariance Matrix and a Direct Test for Heteroskedasticity', *Econometrica* **48**(4), 721–746.

White, H. (2001), *Asymptotic Theory for Econometricians: Revised Edition*, New York: Academic press.

Whitehead, T., Mark Reesor, R. & Davison, M. (2012), 'A Bias-Reduction Technique for Monte Carlo Pricing of Early-Exercise Options', *Journal of Computational Finance* **15**(3), 33.

Woo, J., Liu, C. & Choi, J. (2018), 'Leave-One-Out Least Square Monte Carlo Algorithm for Pricing American Options', *arXiv preprint arXiv:1810.02071* .
1.251

Curriculum Vitae

Name: François-Michel Boire

Degrees:
The University of Western Ontario
London, Ontario, Canada
2017 - 2022 Ph.D. in Financial Modelling

HEC Montréal
Montreal, Quebec, Canada
2015 - 2017 M.Sc. in Financial Engineering

Université de Montréal
Montreal, Quebec, Canada
2012 - 2015 B.Sc. in Mathematics and Economics

Honours and Awards:
Graduate Fellowship
2017 - 2022

Ontario Graduate Scholarship
2019, 2020, 2021

Graduate Student Teaching Award
2019, 2021

Related Work Experience:
Teaching Assistant
The University of Western Ontario
2017 - 2022

Ph.D. Trainee
Bank of Canada
2019

Publications:

Boire, François-Michel, R. Mark Reesor, and Lars Stentoft. 2021. "Efficient Variance Reduction for American Call Options Using Symmetry Arguments" *Journal of Risk and Financial Management* 14, no. 11: 504.
https://doi.org/10.3390/jrfm14110504

Boire, François-Michel, R. Mark Reesor, and Lars Stentoft. 2021. "American Option Pricing with Importance Sampling and Shifted Regressions" *Journal of Risk and Financial Management* 14, no. 8: 340.
https://doi.org/10.3390/jrfm14080340