Improved Greeks for American Options using Simulation*

Pascal Letourneau† and Lars Stentoft‡

September 19, 2016

Abstract

This paper considers the estimation of the so-called Greeks for American style options. This is a challenging task and we discuss in detail the shortcomings of existing methods. A new method is proposed which combines Initial State Dispersion with a value function iteration at the last step to obtain estimates of option prices and Greeks. Our method is benchmarked against various existing methods in terms of bias, convergence and overall performance and is shown to produce estimates which are less biased than what has been obtained previously. Based on the Local Polynomial Regression Literature, we also put forward recommendations on the optimal polynomial order to use.

JEL Classification: C15, G12, G13

Keywords: American options, Greeks, Simulation.

---

*The authors thank participants at the 2014 SIAM Conference on Financial Mathematics & Engineering for valuable comments and seminar participants at IIT Stuart School of Business and the University of Western Ontario. Lars Stentoft acknowledges support from CREATEs - Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation.

†Department of Finance and Business Law, University of Wisconsin-Whitewater, USA, letournp@uww.edu.

‡Department of Economics and Department of Statistical and Actuarial Sciences, University of Western Ontario, Canada, lars.stentoft@uwo.ca.
# Contents

1 Introduction 3

2 American Option Pricing using Simulation 8
   2.1 Simulation and Regression Methods 10
   2.2 Prices and Greeks with Initial State Dispersion 12
   2.3 Properties of the Price and Greeks Estimator 13

3 Assessment of the Basic Method 16
   3.1 Bias Study: Application to Known Data 16
   3.2 Variance Study: Application to Simulated Data 20
   3.3 Bias Study: Determination of the Optimal Exercise Strategy 22
   3.4 Basic Method Study: Application to Simulated Data with and without a known Exercise Strategy 26

4 Using the Value Function for Variance Reduction 28

5 Conclusion 31

A Finite Difference method for American Options 36
1 Introduction

Option pricing, in particular in high-dimensions and for options that have early exercise features, remains a challenge. In particular, this is the case when the goal is to find a flexible method that is generally applicable and can be used to price options in various settings. Monte Carlo simulation is essentially the only such methodology and has been used at least since Boyle (1977) to price European style derivatives in general and options in particular. Simulation methods are flexible and very easy to apply and if one can simulate the underlying dynamics it is essentially possible to price options. Moreover, simulation methods have nice properties since averages of random observations converge to the expected value under very mild assumptions. When pricing American style options the challenge is that one needs to simultaneously determine the optimal early exercise strategy. While it was for a long time believed that it would be difficult to price options with early exercise several methods are now available. Early attempts were made by Tilley (1993) and Barraquand and Martineau (1995), who used simulation to mimic the standard lattice method of determining the holding value function of the option. More recently methods that rely on cross-sectional regression to approximate the value function or to determine the optimal early exercise itself has been introduced (see Carriere (1996), Longstaff and Schwartz (2001), and Tsitsiklis and Van Roy (2001)) and this particular method has been analyzed in quite some details and the method has proven to be very flexible and has been applied in various different settings. Stentoft (2014) showed that among the various proposed numerical methods based on simulation and regressions, the Least-Squares Monte Carlo, or LSM, method, which approximates the stopping time, should be the one considered.

While option pricing is interesting in itself a much more important issue in finance is to calculate the various relevant hedging parameters or price sensitivities that market participants rely on for managing their positions. After all, you only need to price a derivative once, but once it is sold the risk exposures will generally need to be hedged through time. Thus, the price sensitivities,
or Greeks for short, are used on a daily basis by financial institutions for risk management and having these readily available in real time is a necessity for these firms to conduct their business. Note also that unlike prices which are observed in the market place, the Greeks are generally not observed in financial markets and will instead always have to be estimated. To complicate matters even further whereas there is only one single price for the option there are generally several Greeks. In particular, even in the simplest possible model, the constant volatility model of Black and Scholes (1973) and Merton (1973), the BSM model for short, one needs at least the Delta, the first derivative of the option price with respect to the underlying asset value, and likely also the Gamma, the corresponding second derivative or equivalently the first derivative of the Delta, to hedge the risk of changing prices of the underlying asset. In multivariate cases sensitivities to each of the underlying assets are needed along with cross-sensitivities. More generally, sensitivities towards all the stochastic or time varying factors determining an options price, which besides the value of the underlying asset are at least the volatility of this asset, the interest rate, and potentially also the dividend yield, are needed.

While Monte Carlo methods have been examined in detail for pricing of options their use for calculating the Greeks is less explored. For European style options (at least) three different simulation based methods exist and have been examined. The first of these is the so-called finite (forward, backward or central) difference method or approximation for calculating numerical derivatives. The finite difference method involves simulating at two or more values of the parameter of differentiation and approximating the derivative with the difference quotient. Though this method is "universally" applicable and easy to understand and implement, a drawback of the method is that it is biased and potentially very inefficient. The other two methods, the pathwise method and the likelihood ratio method, that have been proposed avoid the drawback of having to simulate several values and use instead information about the simulated stochastic process to replace numerical differentiation with exact calculations. When applicable these methods produce unbiased estimates.

\footnote{Most students who have taken an introductory course in derivatives pricing will have seen at least one example illustrating how option traders and market makers carry this out.}

\footnote{The finite difference method for approximating numerically derivatives of a function should not be confused with the (implicit or explicit) finite difference methods for option pricing of Hull and White (1990).}
though potentially with large variances. In the pathwise method each simulated outcome is differentiated with respect to the parameter of interest. This method works well if the discounted payoff is continuous in the parameter of differentiation which may not always be the case. For example, the method does not work for digital options, and by extension it cannot be used to get the Gamma of a regular option in the BSM model even though the method can be used to approximate the Delta of this option. In the likelihood ratio method on the other hand one differentiates the probability density rather than the outcome. This method thus relies on explicitly knowing the form of the probability densities which could be complicated to obtain for general diffusions. Moreover, an important drawback of this methodology is that the estimates obtained could have (and often does have very) large variance. For a general overview of the methods available for estimating the Greeks with simulation see the textbook of Glasserman (2004).

Much like it has been the case with American option pricing, the use of Monte Carlo methods for calculating the Greeks of options with early exercise is much less explored than for the European version and this is again caused be the fact that for these options we need precise methods for determining the optimal early exercise strategy. For this reason, the majority of the literature on the estimation of Greeks for options with early exercise simply assumes that a precise optimal stopping time strategy is known and examines the properties of methods for calculating the Greeks conditional on this. Conditional on knowing the optimal early exercise strategy calculating the Greeks for an American option is not significantly more complicated than it is for the European style options and one could in theory use the same three approaches mentioned previously. For example, given an early exercise strategy one could implement the finite difference approach mentioned above and estimate the Delta of an option by simulating at two different values of the underlying stock price and estimating the option value from applying the same early exercise strategy. When the difference between the two starting values is small enough by a smoothness argument the same early exercise strategy can be used and this method is viable. However, because the early exercise strategy is estimated, additional noise is introduced and even this simple method may not be appropriate. In particular, there is no guarantee that the option payoff is continuous in
the underlying asset, for example, which may result in poor or even non-existing estimates of the option Greeks. For applications, conditional on a known early exercise strategy, of the pathwise method see Piterbarg (2014) and for a recent method that essentially uses the likelihood ratio method see Kaniel, Tompaidis, and Zemliano (2008). The same caveat obviously applies to the use of a predetermined early exercise strategy for these methods.

Whereas most of the existing literature treats the stopping time as given, which essentially reduces the complexity of the problem to that for a European style option, in this paper we consider simulation methods using regression that can be used to jointly determine prices and sensitivities. There are several reasons to consider simulation algorithms/methods that can be used to jointly obtain the price estimate and estimates of the Greeks. First, by jointly estimating the stopping time strategy the issue of non-continuity in the state variables mentioned above may be mitigated. In the appendix we provide some evidence on this when using the simple finite difference method in the benchmark Black-Scholes-Merton setup. Second, a joint method is likely computationally more efficient as it does not require additional simulations to determine the Greeks. Finally and most importantly, the analysis of the numerical performance in general and the convergence in particular of any method that uses an exogenously given estimate of the stopping time is clearly conditional on this particular given stopping time. Thus, it is difficult if not impossible to make any argument about the actual performance of these methods that are generally and unconditionally applicable. Methods that jointly determine prices and sensitivities on the other hand are, at least in theory, easier to examine in terms of their numerical performance and convergence rates.

There are to our knowledge only a few papers that consider the problem of option pricing and estimation of the Greeks jointly for American options. In Feng, Liu, and Sun (2013) an algorithm for determining the Greeks iteratively along with the value function is proposed. The paper though offers little evidence on the usefulness of the proposed method and very limited numerical results. Moreover, as shown in Stentoft (2014) methods that iterate on the value functions directly often lead to estimates with a significantly larger bias than that obtained with methods that instead iterate directly on the stopping time. The method of Feng, Liu, and Sun (2013) however does not
appear to be applicable to the type of algorithm that iterates on the stopping time directly and we conjecture that the same bias issues are present with this method. Jain and Oosterlee (2013) instead argue that the Delta can be approximated using a finite difference approach in which the regression coefficients from the first early exercise points are used. This method however requires that one uses a regress-later type approach (see Glasserman and Yu (2002)) and one therefore needs regressors that are martingales or for which the one step ahead conditional expectations are known in closed form or have analytical approximations. This restriction significantly limits the choice of potential regressors and for more complicated models it may be impossible to find regressors that satisfy this restriction and in this case the regress-later method would be infeasible. Finally, Wang and Caflisch (2010) suggest that the Greeks may be estimated by performing an additional regression using the values from an initial dispersed sample, a method that is close in spirit to starting the binomial model before the actual current time, the so-called extended tree, to obtain Greeks from this method (see Pelsser and Vorst (1994)).

In this paper we analyze the potential usefulness of the initial state dispersion method for estimating the Greeks in detail and as such our paper is closest in spirit to the work of Wang and Caflisch (2010). Though the idea behind using Initial State Dispersion, or ISD, for estimating Greeks is intuitive and simple, we document in detail that problems may arise unless care is taken even in the benchmark situation of a constant volatility Black-Scholes-Merton type model. In particular, using a simple polynomial approximation as suggested in, e.g., Wang and Caflisch (2010) can result in statistically as well as economically significantly biased results. This bias is not only a function of the order of the polynomial but is also related to the size of the initial dispersion, and it is impossible to propose a method that is generally applicable using this methodology. All hope is not lost however, and based on these findings, we suggest combining the ISD with a value function iteration at the initial time step. By doing so, we significantly reduce the variance of the estimates for a small ISD. Using a small ISD produces unbiased estimates of the Greeks, and the

---

4The authors incorrectly argue that the ability to estimate the Greeks is particular to their bundling algorithm. In fact, it is applicable to any algorithm that uses at the first early exercise point a regress-later style approach.

5In fact we show that for some settings it is impossible to estimate the Greeks precisely with a polynomial irrespective of the order of the approximating function.
value function iteration reduces the variance.

Our first contribution is to study the properties of the Greek estimator and to propose general guidelines to implement the method in practice. Instead of randomly picking an ISD, one has to be aware that the bias on a Greek estimate is proportional to the size of the ISD and that a small ISD is required. However, the variance of a Greek estimate is inversely proportional to the size of the ISD. Furthermore, the bias on the estimation of the optimal exercise strategy is a function of the ISD size. The selection of the ISD is thus a compromise between a biased estimate and a precise estimate.

Our main contribution is to propose an easy modification to the LSM algorithm that reduces the variance of the Greek estimates significantly. We call it the Last Step Value Function Iteration method. At the last iteration of the LSM method, we estimate the conditional expected payoff using all simulated paths. From it we find the approximated value function at $t = 1$. Finally, at $t = 0$, we discount the value function for all paths, which we use to approximate the final value function. By doing so, we significantly reduce the variance of the Greek estimates. We provide extensive numerical results which show that our proposed method improves on existing methods in terms of bias, convergence, and overall performance. Another contribution of our paper is to use results from the LPR literature to recommend the optimal polynomial order for the method.

The rest of the paper is structured as follows: In Section 2 we explain how American options can be priced using simulation and we discuss the idea behind obtaining the Greeks using initial state dispersion. In Section 3 we discuss in detail the numerical performance of the proposed method together with an analysis of this using local polynomial regression. In Section 4 we propose a variance reduction method for Greek estimates. Finally, Section 5 offers some concluding remarks and directions for future research. The Appendix contains additional numerical results.

2 American Option Pricing using Simulation

The first step in implementing any type of numerical algorithm to price American options is to assume that time can be discretized. Thus, we assume that the derivative considered may be
exercised at \( J \) early exercise points. We specify the potential exercise points as \( t_0 = 0 < t_1 \leq t_2 \leq \ldots \leq t_J = T \), with \( t_0 \) and \( T \) corresponding to the current time and maturity of the option, respectively, where it is implicitly assumed that the option cannot be exercised at time \( t_0 \). An American option can be approximated by increasing the number of early exercise points \( J \) and a European option can be valued by setting \( J = 1 \). We assume a complete probability space \((\Omega, \mathcal{F}, \mathbb{P})\) equipped with a discrete filtration \((\mathcal{F}(t_j))_{j=0}^J\). The derivative’s value depends on one or more underlying assets which are modeled using a Markovian process, with state variables \((X(t_j))_{j=0}^J\) adapted to the filtration. We denote by \((Z(t_j))_{j=0}^J\) an adapted payoff process for the derivative satisfying \(Z(t_j) = \pi(X(t_j), t_j)\) for a suitable function \(\pi(\cdot, \cdot)\), which is assumed to be square integrable. This notation is sufficiently general to allow for non constant interest rates through appropriate definition of the state variables \(X\) and the payoff function \(\pi\) (see, e.g., Glasserman (2004)). Following, e.g., Karatzas (1988) and Duffie (1996), in the absence of arbitrage we can specify the American option price as

\[
P(X(0)) = \max_{\tau(t_1) \in \mathcal{T}(t_1)} \mathbb{E}[Z(\tau)|X(0)],
\]

where \(\mathcal{T}(t_j)\) denotes the set of all stopping times with values in \(\{t_j, \ldots, t_J\}\).

In the literature the problem of calculating the American option price in (1) i.e. with \( J > 1 \) is referred to as a discrete time optimal stopping time problem. The preferred way to solve such problems is to use the dynamic programming principle. Intuitively this procedure can be motivated by considering the choice faced by the option holder at time \( t_j \): to exercise the option immediately or to continue to hold the option until the next period. Obviously, at any time the optimal choice will be to exercise immediately if the value of this is positive and larger than the expected payoff from holding the option until the next period and behaving optimally from hereon forward. To fix notation, in the following we let \(V(X(t_j))\) denote the value of the option for state variables \(X\) at a time \( t_j \) prior to expiration. We define \(F(X(t_j)) \equiv \mathbb{E}[Z(\tau(t_{j+1}))|X(t_j)]\) as the expected conditional payoff, where \(\tau(t_{j+1})\) is the optimal stopping time. It follows that

\[
V(X(t_j)) = \max(Z(t_j), F(X(t_j))),
\]

Because we want to consider situations with initial state dispersion we do not impose that \( X(0) = x \).
Thus, it is easily seen that it is possible to derive the optimal stopping time iteratively using the following algorithm:

\[
\begin{align*}
\tau(t_j) &= T \\
\tau(t_j) &= t_j 1_{\{Z(t_j) \geq F(X(t_j))\}} + \tau(t_{k+1}) 1_{\{Z(t_j) < F(X(t_j))\}}, \quad 1 < j \leq J - 1
\end{align*}
\]  

(3)

Based on this, the value of the option in (1) can be calculated as

\[
P(X(0)) = E[Z(\tau(t_1)) | X(0)],
\]

(4)

The backward induction theorem of Chow, Robbins, and Siegmund (1971) (Theorem 3.2) provides the theoretical foundation for the algorithm in (3) and establishes the optimality of the derived stopping time and the resulting price estimate in (4).

### 2.1 Simulation and Regression Methods

The idea behind using simulation for option pricing is quite simple and involves estimating the expected values and therefore the value of the option by an average of a number of random draws. This is easiest to illustrate in the case of a European option for which it is optimal to exercise at time \( T \), i.e. \( \tau(t_1) = T \) by definition. Substituting this into (1) we obtain the following formula

\[
p(X(0)) = E[Z(T) | X(0)],
\]

(5)

where we use lower case to denote that this is the European price and where \( Z(T) = \pi(X(T), T) \) is the payoff from exercising the option at time \( T \). From (5) it is clear, that all that is needed to price the option are the values of the state variables, \( X(T) \), on the day the option expires. Thus, if all the paths are started at the same known values of the state variable, i.e. \( X(0) = x \), an obvious estimate of the true price in (5) can be calculated using \( N \) simulated paths as

\[
\hat{p}^N(X(0) = x) = \frac{1}{N} \sum_{n=1}^{N} \pi(X(T, n), T),
\]

(6)

where \( X(T, n) \) is the value of the state variables at the time of expiration \( T \) along path number \( n \). That is, the price estimate is simply an average of discounted simulate payoffs, and if these are generated from independently simulated paths this estimate will have all the usual nice properties and will generally be unbiased.
When the option is American one needs to simultaneously determine the optimal early exercise strategy and this complicates matters. In particular, it is generally not possible to implement the exact algorithm in (2) because the conditional expectations are unknown and therefore the price estimate in (4) is not feasible. Instead an approximate algorithm is needed. Because conditional expectations can be represented as a countable linear combination of basis functions we write
\[ F(X(t_j)) = \sum_{m=0}^{\infty} \phi_m(X(t_j)) c_m(t_j), \]
where \( \{ \phi_m(\cdot) \}_{m=0}^{\infty} \) form a basis. In order to make this operational we assume that it is possible to approximate well the conditional expectation function by using the first \( M + 1 \) terms such that
\[ F(X(t_j)) \approx \hat{F}_M(X(t_j)) = \sum_{m=0}^{M} \phi_m(X(t_j)) c_m(t_j), \]
and that we can obtain an estimate of this function by
\[ \hat{F}_N(X(t_j)) = \sum_{m=0}^{M} \phi_m(X(t_j)) \hat{c}_m(t_j), \]
where \( \hat{c}_m(t_j) \) are approximated or estimated using \( N \geq M \) independent simulated paths. Based on the estimate in (7) we can derive an estimate of the optimal stopping time based on the estimated parameters as:
\[
\begin{cases}
\hat{\tau}_M^N(t_j) = T \\
\hat{\tau}_M^N(t_j) = t_j 1\{Z(t_j) \geq \hat{F}_M(X(t_j))\} + \hat{\tau}_M^1 1\{Z(t_j) < \hat{F}_M(X(t_j))\}, \quad 1 < j \leq J - 1
\end{cases}
\]
From the algorithm in (8) a natural estimate of the option value in (4) is given by
\[ \hat{P}_M^N(X(0)) = \mathbb{E}[Z(\hat{\tau}_M^N(1)) | X(0)] = \hat{F}_M^N(X(0)). \]
In the special case when all the paths are started at the same known values of the state variable, i.e. \( X(0) = x \), the estimate in (9) simplifies to
\[ \hat{P}_M^N(X(0) = x) = \frac{1}{N} \sum_{n=1}^{N} Z(n, \hat{\tau}_M^N(1, n)), \]
where \( Z(n, \hat{\tau}_M^N(1, n)) \) is the payoff from exercising the option at the optimal stopping time \( \hat{\tau}_M^N(1, n) \) determined for path \( n \) according to (8).

\(^7\text{This assumption is justified when approximating functions that are elements of the } L^2 \text{ space of square-integrable functions relative to some measure. Since } L^2 \text{ is a Hilbert space, it has a countable orthonormal basis (see, e.g., Royden (1988)).}\)
2.2 Prices and Greeks with Initial State Dispersion

In the general case with dispersed initial values of the state variables the algorithm above generates not only a price but a price function given by \( \hat{P}_M^N (X(0)) = \hat{F}_M^N (X(0)) \) and we make this explicit by specifying \( \hat{P}_M^N \) as a function of \( X(0) \). A natural estimate of the option price for a given value of the state variables \( X_0 \) is obtained by evaluating the approximation \( \hat{F}_M^N \) at this value and hence we have

\[
\hat{P}_M^N (X_0) = \hat{F}_M^N (X_0) .
\]  

(11)

In a similar way we can define the sensitivity of the option price at \( X_0 \) with respect to state variable \( X^i \) as

\[
\frac{\partial \hat{F}_M^N (X_0)}{\partial X^i} = \sum_{m=0}^{M} \frac{\partial \phi_m (X_0)}{\partial X^i} \hat{c}_m^N (0) .
\]  

(12)

Higher order derivatives or cross derivatives can be defined in a similar manner.

In the special case where the only state variable is the stock price, \( (S(t_j))_{j=0}^T \), the formulas for the first derivative, the \( \Delta \), and the second derivative, the \( \Gamma \), at \( S_0 \) are given by

\[
\hat{\Delta} (S_0) = \frac{\partial \hat{F}_M^N (S_0)}{\partial S_0} = \sum_{m=0}^{M} \frac{\partial \phi_m (S_0)}{\partial S_0} \hat{c}_m^N (0) = \sum_{m=0}^{M} \phi'_m (S_0) \hat{c}_m^N (0) ,
\]  

(13)

and

\[
\hat{\Gamma} (S_0) = \frac{\partial^2 \hat{F}_M^N (S_0)}{\partial S_0^2} = \sum_{m=0}^{M} \frac{\partial^2 \phi_m (S_0)}{\partial S_0^2} \hat{c}_m^N (0) = \sum_{m=0}^{M} \phi''_m (S_0) \hat{c}_m^N (0) ,
\]  

(14)

respectively. These formulas are equivalent to those from Wang and Caflisch (2010). Note that if the initial approximation \( \hat{F} \) is a simple polynomial the Greeks are particularly easy to calculate though it may be more complicated for other types of approximation functions. However, even if analytical derivatives are difficult to obtain one can always use numerical differentiation to obtain the required sensitivities.

Finally, it should be mentioned that though the initial approximation could be done using the same approximating functions that are used in the stopping time regressions, the regressions at time \( t_j > 0 \), it does not have to be so. In fact, an important lesson from our analysis is that this initial regression is very different from the other cross sectional regressions conducted as part of the LSM algorithm. For example, what is used to calculate the price at time \( t_j = 0 \) is the
actual approximation when using initial state dispersion. In the rest of the LSM algorithm the approximation is only used to make a decision about whether to exercise or not and the actual cash flows used for valuation are those realized along a particular path.\footnote{In some sense this is similar to the discussion about using value function iteration or stopping time iteration in Stentoft (2013), which shows that the bias from iterating on the value function directly and using the actual approximated values is much larger than when iterating on the stopping time. Unfortunately with ISD we do not have other choice than to use the value function approximation directly.}

The idea of using dispersion of the initial value in a simulation context has been used in several papers. In fact, it is difficult to pinpoint exactly who came up with this idea. Rasmussen (2005) is one of the first to use the idea and this is done together with other techniques for variance reduction. He proposes initial state dispersion as an alternative method to importance sampling or stratification to increase the number of paths that are in the money. Proper use of initial state dispersion ensures that a certain number of paths should be exercised at each time step between the initial time $t_j = 0$ and maturity $t_j = T$ and this improves the estimation of the optimal exercise strategy. Ideally, the ISD would be large enough to cover the exercise frontier from $t_j = 0$ to $t_j = T$, i.e., the ISD should be large. Wang and Caflisch (2010), however, is to our knowledge the first to propose that ISD can be used together with an additional initial regression to approximate numerically the Delta and Gamma of American options.\footnote{Rasmussen (2005) instead uses ISD on one set of simulated paths to determine the early exercise strategy but uses a new set of paths to price the option. Doing so essentially avoids having to perform the extra initial regression and involves an element of the out of sample pricing method suggested in Longstaff and Schwartz (2001) that allows one to sign the bias of the price estimate. While it is possible that the improved early exercise strategy from using ISD could mitigate the issues of non-continuity mentioned above, the other arguments for using a joint method remain valid and such improvements should actually further improve our proposed method.} That said, initial dispersion could equally well be used to obtain the Greeks of European style options in situations where existing methods are either inapplicable or would simply be too time consuming.\footnote{Besides the above uses of ISD there are several other important benefits. For example, when using initial dispersion we can let $\Delta_t$, the time between two early exercise points, tend to zero and still do the regressions. Otherwise, the small dispersion of $X(t_j)$ near $t_0$ lead to numerical issues.}

2.3 Properties of the Price and Greeks Estimator

In the stopping time iteration algorithm regressions are used to approximate $\hat{F}_M^N(X(\hat{t}_j))$. These approximations are used exclusively to solve \footnote{In some sense this is similar to the discussion about using value function iteration or stopping time iteration in Stentoft (2013), which shows that the bias from iterating on the value function directly and using the actual approximated values is much larger than when iterating on the stopping time. Unfortunately with ISD we do not have other choice than to use the value function approximation directly.}, i.e., to compare the holding value of the option to the exercise value. The values from $\hat{F}_M^N(X(\hat{t}_j))$ are not reused. This is different from the regression
at $t = 0$ where $\hat{F}^N_M (X (t_0))$ is evaluated to obtain the price, and derived to obtain the Greeks. Here, we wish to study the properties of this regression to better understand the estimator and provide guidance into how to implement the method in practice.

To do so, we use the fact that an ordinary least squares regression is a special case of a Local Polynomial Regression where the kernel would be a uniform distribution and the bandwidth would be large enough to include all simulated paths. Local Polynomial Regression (LPR hereafter) have been introduced by Stone (1977) and extensively studied (see, e.g., Fan and Gijbels (1996) and the references therein). LPR are generalized to multivariate problems (see, e.g., Masry (1996)) and thus applicable to approximating (11) in general. In what follows, we use the univariate notation for simplicity.

Let $X_i$ be variables on the support $\mathcal{X}$ and let $Y_i$ be scalar responses. Consider an approximation to $\Phi (x) = \mathbb{E} [Y | X = x]$ at a point $x$. The data can be viewed as being generated by:

$$Y_i = \Phi (X_i) + \sigma (X_i) e_i, \quad i = 1, \ldots, N$$

where $\mathbb{E} [e] = 0$, $\sigma^2 (X_i) < \infty$, $X$ and $e$ are independent, and $\Phi (\cdot)$ is twice continuously differentiable on the support $\mathcal{X}$. The function $\Phi (\cdot)$ can be approximated locally by a polynomial of order $M$:

$$\Phi (x) \simeq \Phi (X_0) + \Phi^{(1)} (X_0) (X - X_0) + \ldots + \Phi^{(M)} (X_0) (X - X_0)^M / M!, \quad (15)$$

for $x$ in the neighborhood of $x_0$, where $\Phi^{(q)}$ is the $q$th derivative of $\Phi$ with respect to $x$. Let $K$ be a probability density function assigning weights to local data points and let $h$ be a bandwidth parameter controlling the size of the neighborhood.

A LPR is performed by the following weighted least squares regression:

$$\min_{\beta_j} \sum_{i=1}^{N} \left\{ Y_i - \sum_{j=0}^{M} \beta_j (x_i - x_0)^j \right\}^2 K_h (x_i - x_0), \quad (16)$$

where $\beta_j$ are the solutions to the regressions, $N$ the number of sample points (or the number of simulated paths in our current case), $M$ the polynomial order, $K_h (t) = K (\frac{t}{h}) / h$ is the kernel function assigning weights to each path, and $h$ is the bandwidth parameter. The whole function $\hat{\Phi} (x)$ is obtained by running (16) with $x_0$ varying in an appropriate domain.
The estimation of the qth order derivative with LPR is straight forward. One can estimate $\Phi^{(q)}(x)$ via the intercept coefficient of the qth derivative of the local polynomial being fitted at $X$, assuming $M > q$, $\hat{\Phi}^{(q)}(x) = q!\hat{\beta}_q$. For the problem at hand, this means regressing the conditional discounted cash flows for each path on the distance of each path to the initial state, $S(0)$, using a weighted least squares regression. Asymptotic properties of estimating the derivatives of $\Phi$ as well as asymptotic normality were established by Fan and Gijbels (1996). Strong uniform consistency properties were shown by Delecroix and Rosa (1996).

The asymptotic conditional bias of $\hat{\Phi}^{(q)}(x)$ is proportional to $q!$, proportional to $h$, and inversely proportional to $M$. The higher the derivative to estimate, the higher the potential bias. The higher the polynomial order used to approximate $\hat{\Phi}^{(q)}(x)$, the lower the bias. This suggest using the highest possible polynomial order. The larger the bandwidth, the larger the ISD, the larger the bias will be. Thus, to avoid bias, one should aim for the lowest bandwidth possible and the highest polynomial order possible. Note that using a small bandwidth (i.e., a small ISD) to get unbiased Greek estimates is potential conflicting with using a large ISD to improve the estimation of the optimal exercise strategy.

One as to also consider the variance of the estimates. The asymptotic conditional variance of $\hat{\Phi}^{(q)}(x)$ is proportional to $q!^2$ and $M$, but inversely proportional to $h$. The higher the derivative to estimate, the higher the variance will be. The higher the polynomial used to approximate $\hat{\Phi}^{(q)}(x)$, the higher the variance. Suggesting limiting the order of the polynomial. Finally, the larger bandwidth, the larger the ISD, the lower the variance will be. To limit the variance one needs to use a larger bandwidth. In practice, one has to compromise between bias and variance in the selection of the bandwidth and the polynomial order. (For more details on the asymptotic bias and variance, we refer the reader to Fan and Gijbels (1996) Chapter 3.)

LPR literature as useful results for selecting the polynomial order. Fan and Gijbels (1995) show that for going from $M - q$ even to odd reduces the bias, but has no effect on the variance. However, going from $M - q$ odd to even reduces the bias, but increases the variance of the estimates. For that reason, it is preferable to use $M - q$ odd. In our application, we are interested in $q = 0, 1, 2,$
which are the price, Delta, and Gamma, respectively. For the remainder on this study, we will select $M \geq 3$ odd such that $M - 0$ and $M - 2$ are odd.

3 Assessment of the Basic Method

The idea is simple but often simple ideas are really complicated!

In this section we provide numerical results for different setups. First, we study a best case scenario to get insights on the potential bias of the method at $t = 0$. Second, we study the variance of the estimates using simulated paths. Third, we review the approximation of the optimal stopping time. Finally, we apply the method to a large sample of options to assess the bias and variance and the estimates. Throughout, we use simple monomials and a regular Ordinary Least Squares regression (OLS) to estimate $\hat{F}_N^M(X(t_j)), \forall t_j$ and $\hat{F}_N^M(X(0))$.

3.1 Bias Study: Application to Known Data

To analyze the proposed methodology further we estimate (7) at time $t_0$ using true values instead of simulated data. That is, we are considering the "best case" scenario in which a polynomial is fit directly on the option price function. We considered a set of options in a BSM world with volatility $\sigma \in \{10\%, 20\%, 40\%\}$, strike $K \in \{36, 40, 44\}$, and with maturities of $T \in \{0.5, 1, 2\}$. For all cases, we consider an initial asset price is $S_0 = 40$ and a risk free rate is $r = 6\%$. To generate the $d$th value, $S^d_{t_0}$, of a total of $N$ initially dispersed values we use a deterministic version of the method proposed in Wang and Caflisch (2010) given by

$$S^d_{t_0} = S_0 \exp \left(\alpha \sigma \sqrt{T} \epsilon_d \right), \quad (17)$$

where $\epsilon_d$ is the $d/N$th percentile of the standard normal distribution. In (17), $\alpha$ is the parameter driving the dispersion and in particular when $\alpha = 0$ the standard LSM method is obtained. We calculate benchmark prices for a total of $N = 100,000$ dispersed values of the initial stock price using the Binomial model with 25 steps per trading day for a total of 6,300 annual steps. We also, for reference, calculate the Delta and Gamma for each of these options using simple one and two step ahead finite differences as described in, for example, Hull (2006).
This figure plots the price, Delta, and Gamma from the Binomial Model for $N = 100,000$ initially spread values of the stock price using $\alpha = 0.50$ to determine the ISD with an initial stock price of $S_0 = 40$. Results for both European style and American style options with $J = 50$ exercise possibilities are shown. The strike price is $K = 44$, the volatility is $\sigma = 20\%$, the interest rate is $r = 6\%$, and the maturity of the option is $T = 1$ year.

To get an idea about the data used in the initial regression we start by plotting the resulting pathwise payoffs, essentially the option prices, for both the in the money, or ITM, European and American option with 50 exercise points in Figure 1 using a value of $\alpha = 0.50$ as suggested in Wang and Caflisch (2010) for the ISD. The plot shows that though European and American option prices look similar (the difference being the early exercise premium) there are major differences in the Greeks, particularly when it comes to the Gamma. In particular, Figure 1 clearly shows that methods that approximate the price function and use first and second order derivatives of these to approximate the Delta and the Gamma, respectively, are likely much more difficult to implement for the American option than for the European option. The reason for the problems arising is

---

Plots for options that are at the money and out of the money are similar though shifted to the left because of the lower exercise price.
Table 1: Benchmark values for the sample of options used in this section.

<table>
<thead>
<tr>
<th></th>
<th>Price</th>
<th>Delta</th>
<th>Gamma</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K = 36$</td>
<td>0.9166</td>
<td>-0.1979</td>
<td>0.0381</td>
</tr>
<tr>
<td>$K = 40$</td>
<td>2.3141</td>
<td>-0.4040</td>
<td>0.0597</td>
</tr>
<tr>
<td>$K = 44$</td>
<td>4.6535</td>
<td>-0.6648</td>
<td>0.0765</td>
</tr>
</tbody>
</table>

Note: We considered a set of options in a BSM world with volatility $\sigma = 20\%$, strike $K \in \{36, 40, 44\}$, and with maturities of 1 year. For all cases, we consider an initial asset price is $S_0 = 40$ and a risk free rate is $r = 6\%$. We calculate benchmark prices using the Binomial model with 25 steps per trading day for a total of 6,300 annual steps.

that as the stock price becomes very low and the option is deep in the money the price essentially becomes linear in the underlying asset and the Delta approaches a value of $-1$ very quickly and much faster than for the European option. Because of this, the Gamma for the American option approaches zero much faster than for the European option. A simple polynomial in the underlying asset cannot easily approximate this type of function. In fact, it should be noted that when the stock price is around $36.7957$ the option should actually be exercised immediately. This would actually introduce a discontinuity in the option price which would make the approximation even more difficult.

Let us first illustrate the effect of the size of the ISD on the bias of the price, Delta, and Gamma estimates. For space consideration, we show here results for the 1 year option with $\sigma = 20\%$ and $K \in \{36, 40, 44\}$. For reference, table 1 show the benchmark price, Delta, and Gamma for options in a BSM world with volatility $\sigma = 20\%$, strike $K \in \{36, 40, 44\}$, and with a maturity of 1 year. For all cases, we consider an initial asset price is $S_0 = 40$ and a risk free rate is $r = 6\%$. The conclusions drawn are robust to other volatilities and maturities.

Figure 2 shows the effect of the size of the ISD on the bias of the estimates. First, observe that in general, as the ISD becomes larger, the bias is larger. Second, observe that the bias is not monotone. In some cases, it is possible to obtain spurious good results as the bias can go from an upper bias to a lower bias as the ISD increase. Third, as the order of the derivative increases, the bias also increases. Finally, observe that when the ISD is small, Gamma estimates become unstable. This is due to numerical issues. The data used in the regression is, at the numerical precision of

\[12\] All other results are available upon request.
Figure 2: Effect of the size of the ISD on the bias of price, Delta, and Gamma estimates.

A sample of 100,000 benchmark prices are generated for $S_0^d = S_0 \exp \left( \alpha \sigma \sqrt{T} \epsilon_d \right)$ for 100 different $\alpha \in [0.01; 1]$. For each alpha, a polynomial of order 3 is fit on the sample data. The estimation of the option price is obtained by evaluating the polynomial at $S(0)$, while Delta and Gamma est obtained by evaluating the first and second derivative of the polynomial, respectively. The graphs show the relative bias as a function of the ISD size determined by the factor $\alpha$. Relative bias = (Estimate − benchmark) / benchmark.

Matlab, practically a straight line and the higher derivatives become impossible to estimate.

Next, we show the effect of the polynomial order on the bias. Figure [3] shows the effect of the polynomial order on the bias for a large range of ISD size for an OTM option. First, observe that increasing the polynomial order from 3 to 5 reduces the bias significantly for this OTM option.
It allows for a much large ISD size without producing biased results. Second, observe that larger polynomial order do reduce the bias, but the marginal effect is smaller and smaller. The general results hold for other moneyness, other maturities and other volatilities. In light of these results, we suggest to use a polynomial of order 5 to estimate the value function and a small ISD to avoid getting biased estimates.

3.2 Variance Study: Application to Simulated Data

We gather insights on the variance of the estimate by studying results when the optimal stopping time and the initial value function are approximated using simulated paths. For a 1 year option, with \( \sigma = 0.20 \).

Figure 4 shows the standard deviation of the price, Delta, and Gamma estimates as a function of the size of the ISD for a 1year option with \( \sigma = 0.20 \). For each \( \alpha \in \{0.010; \ldots; 0.500\} \), a sample of 100,000 paths is simulated from an ISD using (17). The optimal stopping time is approximated from the simulated paths, and the value function is estimated from the discounted cash flows at \( t = 0 \) using \( M = 3 \). This is repeated 100 times and the standard deviation of the 100 estimates is reported. Observe that the standard deviation of the price estimates is barely affected by the size of the ISD. However, for Delta and Gamma, the standard deviation quickly diminishes as the ISD size increases. Or, equivalently, the standard deviation quickly becomes very large as the ISD size decreases. Note, the range of standard deviation increases for Delta and Gamma.

Figure 5 shows the effect of the size of the ISD for different polynomial orders. We use the standard deviation of the estimation with \( M = 3 \) as reference to show how much the standard deviation is affected when \( M \) is increased. The standard deviation on the price estimates is affected, but the effect is small. This is interesting because in the estimation of the optimal stopping time, the prices from the approximated function are the only information used in determining the optimal stopping time. Thus a large polynomial could be used without affecting the estimation. Which is not the case for the Greeks. In general, the standard deviation on Greeks estimates is larger for higher order polynomial. This is especially true for the estimation of Gamma. Clearly, \( M = 9 \) is not appropriate to estimate Gamma.
Figure 3: Effect of the polynomial order on the bias of price, Delta, and Gamma estimates.

A sample of 100,000 benchmark prices are generated using $S_0^d = S_0 e^{\alpha \sigma \sqrt{T} \epsilon_d}$ for 100 different $\alpha \in [0.01; 1]$. For each alpha, we regress the benchmark price of the option on the asset price using a polynomial of order $M \in \{3; 5; 7; 9; 11; 13; 15\}$. The estimation of the option price is obtained by evaluating the polynomial at $S(0)$, while Delta and Gamma est obtained by evaluating the first and second derivative of the polynomial, respectively. The graphs show the relative bias as a function of the ISD size determined by the factor $\alpha$. For this option, $S(0) = 40$, $K = 36$, $\sigma = 0.20$, $r = 0.06$, and $T = 1$. 
Figure 4: Effect of the size of the ISD on the std.dev. of price, Delta, and Gamma estimates.

Estimates for a 1 year option with $\sigma = 0.20$ and $r = 0.06$. For each $\alpha \in \{0.010; \ldots; 0.500\}$, a sample of 100,000 paths is simulated from an ISD using $S^d_0 = S_0 \exp \left( \alpha \sigma \sqrt{T} \epsilon_d \right)$. For each alpha, we regress the approximated payoff on the asset price using a polynomial of order 3. The estimation of the option price is obtained by evaluating the polynomial at $S(0)$, while Delta and Gamma are obtained by evaluating the first and second derivative of the polynomial, respectively. The graphs show the relative variance on the estimates as a function of the ISD size determined by the factor $\alpha$. The simulation is repeated 100 times, and the standard deviation is computed over those 100 repetitions. The optimal stopping time is estimated using $M = 9$ in the LSM algorithm. This is our base case for the standard deviation.

3.3 Bias Study: Determination of the Optimal Exercise Strategy

In the stopping time iteration algorithm regressions are used to approximate $\hat{F}_M^N (X(t_j))$. These approximations are used exclusively to solve $[S]$, e.i. to compare the holding value to the exercise
Figure 5: Effect of the size of polynomial order on the std.dev. of price, Delta, and Gamma estimates.

For each $\alpha \in \{0.010; \ldots; 0.500\}$, a sample of 100,000 paths is simulated from an ISD using $S_0^d = S_0 \exp \left( \alpha \sigma \sqrt{T} \epsilon_d \right)$. For each simulation, we regress the approximated payoff on the asset price using a polynomial of order $M \in \{3; 5; 7; 9\}$. The estimation of the option price is obtained by evaluating the polynomial at $S(0)$, while Delta and Gamma est obtained by evaluating the first and second derivative of the polynomial, respectively. The experiment is repeated 100 times and we compute the standard deviation over the 100 repetitions. Finally, we compute the ratio of the standard deviation for each polynomial order to the base case with $M = 3$. The optimal stopping time is estimated using $M = 9$.

value. The estimation of $\hat{\tau}_M^N(t_0)$ using $\hat{F}_M^N(X(t_j))$ relies on determining the intersection of $\hat{F}_M^N(X(t_j))$ with $Z(t_j)$. As mentioned above, the bias on $\hat{F}_M^N(X(t_j))$ is proportional to the range of $X(t_j)$.
(or the bandwidth if the OLS regression is modeled as a LPR) and inversely proportional to the polynomial order. The difficulty of estimating $\hat{F}_M^N(X(t_j))$ is addressed, e.g., in Rasmussen (2005) where he proposes to use an ISD to improve the estimation.

There are four potential problems in estimating $\hat{F}$ (shorthand for $\hat{F}_M^N(X(t_j))$). First, $\hat{F}$ will be biased if the polynomial order is too low, leading to under-fitting. Second, there can be over-fitting problems with a high order polynomial. Third, $\hat{F}$ will be biased if the range of $X(t_j)$ is too large, leading under-fitting. Fourth, there can be over-fitting problems with a small range of $X(t_j)$. Under-fitting produces sub-optimal exercise strategies and price estimates which are biased low. Over-fitting produces exercise strategies which are adapted to the sample of simulated paths and price estimates which are biased high.

Our recommendation is to use a relatively large ISD and a relatively large polynomial order to better approximate the exercise strategy. We present here a summary of the situation.

Figure 6 shows the effect of both the polynomial order and the size of the ISD on the approximation of the optimal exercise strategy. When the polynomial order is low, the approximation of $\hat{F}$ is biased close to maturity because the polynomial lacks flexibility to match the sharp change in slope near the exercise price $K$. When the polynomial order is low and the ISD is small, the approximation of $\hat{F}$ is erratic away from maturity because there are no paths in the exercise region. The intersection of $\hat{F}$ and $Z$ cannot be determined adequately. The bias near maturity leads to under-fitting and price estimates which are biased low. The erratic estimation away from maturity is prone to over-fitting and leads to price estimates which are biased high. Both bias can potentially balanced one another. When the polynomial order is low and the ISD is large, the approximation of $\hat{F}$ for all time steps. When the polynomial order is high and the ISD is small, the exercise strategy is approximated well close to maturity, but becomes erratic away from maturity. When the polynomial order is high and the ISD is large, the exercise strategy is approximated well. If there is a bias, it appears to be small. In light of these results, we propose to use a large ISD combined with a large polynomial order to approximate the exercise strategy.

This proposition conflicts with the estimation of the Greeks at $t = 0$. The solution is simple and
Figure 6: Effect of Polynomial Order and ISD Size on the Approximation of the Exercise Strategy.

This figure shows the effect of the polynomial order and ISD size of the approximation of the optimal exercise strategy. Results are shown for an ATM option with $S(0) = 40$, and $K = 40$ with a maturity of $T = 1$, a volatility of $\sigma = 0.20$, and a risk free rate of $r = 0.06$. The simulation uses 100,000 paths. At each time steps, the approximated values function from the regression is compared to the exercise value function to determine the exercise frontier. The simulations are repeated 100 times. The mean over 100 exercise frontiers is shown in the figures. The top figures uses $M = 3$, while the bottom figures uses $M = 9$. The approximated exercise strategy is contrasted with a benchmark exercise strategy obtained using a binomial tree.

consist in using one set of paths to estimate the exercise strategy and one set of paths to estimate the price and Greeks. This "Out-Of-Sample" pricing method can be implement in one algorithm. First, simulate two sets of paths, one with a large ISD and one with a small ISD. Apply the LSM algorithm. Estimate $\hat{F}$ using the ITM paths from the first set of paths. Apply $\hat{F}$ to both sets
of paths to determine the optimal exercise strategy for both sets. Use the second set of paths to
estimate the price and Greeks at $t = 0$.

### 3.4 Basic Method Study: Application to Simulated Data with and without a known Exercise Strategy

In this section, we show results for the basic method for a large set of options. In order to approximate the exercise strategy and estimate price and Greeks, we propose to use the out-of-sample algorithm. We simulate a first set of 100,000 path using a large ISD by setting $\alpha = 1$ in (17). This ISD should be large enough to encompass the exercise frontier. To estimate $\hat{F}$, we use $M = 9$ as this should be flexible enough to avoid large bias. We simulate a second set of 100,000 paths using a small ISD by setting $\alpha = 0.05$.

In the Least Squares Monte Carlo algorithm we estimate $\hat{F}$ using the first set of path. The approximate $\hat{F}$ is then apply to both sets of paths. At $t = 0$, we use the second set of paths to estimate $\hat{F}_M^N (X(t_0))$ using a polynomial of order $M = 5$. $\hat{F}_M^N (X(t_0))$ is used to estimate the price of the option and derivated to estimate the Greeks.

Here we present the results for large sample of options. We considered a set of options in a BSM world with volatility $\sigma \in \{10\%, \ 20\%, \ 40\%\}$, strike $K \in \{36, \ 40, \ 44\}$, and with maturities of $T \in \{0.5, \ 1, \ 2\}$. For all cases, we consider an initial asset price is $S_0 = 40$ and a risk free rate is $r = 6\%$. To generate an initial state dispersion, we use (17), where $\alpha$ drives the size of the ISD. We use the out-of-sample algorithm and estimate the price and Greeks using the method describe previously. We use 100,000 simulated paths to estimate the exercise frontier, and another set of 100,000 simulated paths to estimate the price and Greeks. The simulations are repeated 1,000 times. We then report the mean and standard deviation over 1,000 repetitions.

Table 2 shows the results when applying the basic method using the out of sample algorithm. Most of the price estimates are statistically different from the benchmark values. However, the bias is generally very small. The maximum absolute error is 0.18%. Most of the Greek estimates are not statistically different from the benchmark values. Note, however, that the standard deviation on the estimates is very large.
Table 2: Estimation of Price and Greeks using the Basic Method and the Out-Of-Sample Algorithm.

<table>
<thead>
<tr>
<th>k</th>
<th>σ</th>
<th>T</th>
<th>Price BM Estimate (std)</th>
<th>Delta BM Estimate (std)</th>
<th>Gamma BM Estimate (std)</th>
</tr>
</thead>
<tbody>
<tr>
<td>36</td>
<td>10%</td>
<td>0.5</td>
<td>0.0304 (0.0008)</td>
<td>-0.0281 (0.0070)</td>
<td>0.0236 (0.0861)</td>
</tr>
<tr>
<td>36</td>
<td>20%</td>
<td>0.5</td>
<td>0.4978 (0.0051)</td>
<td>-0.1607 (0.0210)</td>
<td>0.0449 (0.1297)</td>
</tr>
<tr>
<td>36</td>
<td>40%</td>
<td>0.5</td>
<td>2.1993 (0.0152)</td>
<td>-0.2759 (0.0403)</td>
<td>0.0305 (0.1004)</td>
</tr>
<tr>
<td>36</td>
<td>10%</td>
<td>1.0</td>
<td>0.0895 (0.0016)</td>
<td>-0.0545 (0.0094)</td>
<td>0.0305 (0.0872)</td>
</tr>
<tr>
<td>36</td>
<td>20%</td>
<td>1.0</td>
<td>0.9166 (0.0080)</td>
<td>-0.1979 (0.0291)</td>
<td>0.0381 (0.1021)</td>
</tr>
<tr>
<td>36</td>
<td>40%</td>
<td>1.0</td>
<td>3.4366 (0.0207)</td>
<td>-0.2863 (0.0384)</td>
<td>0.0227 (0.0673)</td>
</tr>
<tr>
<td>36</td>
<td>10%</td>
<td>2.0</td>
<td>0.1713 (0.0025)</td>
<td>-0.0751 (0.0101)</td>
<td>0.0313 (0.0640)</td>
</tr>
<tr>
<td>36</td>
<td>20%</td>
<td>2.0</td>
<td>1.4317 (0.0104)</td>
<td>-0.2165 (0.0283)</td>
<td>0.0311 (0.0325)</td>
</tr>
<tr>
<td>36</td>
<td>40%</td>
<td>2.0</td>
<td>4.9643 (0.0253)</td>
<td>-0.2786 (0.0346)</td>
<td>0.0168 (0.0443)</td>
</tr>
<tr>
<td>40</td>
<td>10%</td>
<td>0.5</td>
<td>0.7347 (0.0042)</td>
<td>-0.4088 (0.0326)</td>
<td>0.1846 (0.2002)</td>
</tr>
<tr>
<td>40</td>
<td>20%</td>
<td>0.5</td>
<td>1.7915 (0.0099)</td>
<td>-0.4256 (0.0386)</td>
<td>0.0790 (0.2525)</td>
</tr>
<tr>
<td>40</td>
<td>40%</td>
<td>0.5</td>
<td>3.9718 (0.0209)</td>
<td>-0.4186 (0.0549)</td>
<td>0.0367 (0.1333)</td>
</tr>
<tr>
<td>40</td>
<td>10%</td>
<td>1.0</td>
<td>0.8893 (0.0050)</td>
<td>-0.3901 (0.0269)</td>
<td>0.1505 (0.1369)</td>
</tr>
<tr>
<td>40</td>
<td>20%</td>
<td>1.0</td>
<td>2.3141 (0.0123)</td>
<td>-0.4040 (0.0458)</td>
<td>0.0597 (0.1558)</td>
</tr>
<tr>
<td>40</td>
<td>40%</td>
<td>1.0</td>
<td>5.3120 (0.0253)</td>
<td>-0.3903 (0.0478)</td>
<td>0.0265 (0.0819)</td>
</tr>
<tr>
<td>40</td>
<td>10%</td>
<td>2.0</td>
<td>1.0241 (0.0056)</td>
<td>-0.3729 (0.0233)</td>
<td>0.1301 (0.1456)</td>
</tr>
<tr>
<td>40</td>
<td>20%</td>
<td>2.0</td>
<td>2.8846 (0.0144)</td>
<td>-0.3796 (0.0399)</td>
<td>0.0468 (0.0980)</td>
</tr>
<tr>
<td>40</td>
<td>40%</td>
<td>2.0</td>
<td>6.9171 (0.0298)</td>
<td>-0.3552 (0.0409)</td>
<td>0.0195 (0.0519)</td>
</tr>
<tr>
<td>44</td>
<td>10%</td>
<td>0.5</td>
<td>3.9473 (0.0025)</td>
<td>-0.9998 (0.0195)</td>
<td>0.0010 (0.2596)</td>
</tr>
<tr>
<td>44</td>
<td>20%</td>
<td>0.5</td>
<td>4.3091 (0.0119)</td>
<td>-0.7563 (0.0468)</td>
<td>0.0907 (0.1052)</td>
</tr>
<tr>
<td>44</td>
<td>40%</td>
<td>0.5</td>
<td>6.3262 (0.0251)</td>
<td>-0.5637 (0.0664)</td>
<td>0.0389 (0.1588)</td>
</tr>
<tr>
<td>44</td>
<td>10%</td>
<td>1.0</td>
<td>3.9473 (0.0025)</td>
<td>-0.9998 (0.0152)</td>
<td>0.0054 (0.1347)</td>
</tr>
<tr>
<td>44</td>
<td>20%</td>
<td>1.0</td>
<td>4.6535 (0.0144)</td>
<td>-0.6648 (0.0520)</td>
<td>0.0765 (0.1812)</td>
</tr>
<tr>
<td>44</td>
<td>40%</td>
<td>1.0</td>
<td>7.6104 (0.0288)</td>
<td>-0.4966 (0.0544)</td>
<td>0.0291 (0.0932)</td>
</tr>
<tr>
<td>44</td>
<td>10%</td>
<td>2.0</td>
<td>3.9480 (0.0028)</td>
<td>-0.9963 (0.0118)</td>
<td>0.0161 (0.0810)</td>
</tr>
<tr>
<td>44</td>
<td>20%</td>
<td>2.0</td>
<td>5.0832 (0.0174)</td>
<td>-0.5897 (0.0478)</td>
<td>0.0639 (0.1098)</td>
</tr>
<tr>
<td>44</td>
<td>40%</td>
<td>2.0</td>
<td>9.1820 (0.0340)</td>
<td>-0.4342 (0.0472)</td>
<td>0.0219 (0.0587)</td>
</tr>
</tbody>
</table>

An ISD is created using $\alpha = 1$ and 100,000 paths are simulated. An optimal exercise strategy is approximated using the simulated paths and $M = 9$. An new ISD is created using $\alpha = \ldots$ and a new set of 100,000 paths are simulated. The optimal exercise strategy obtained from the previous step is applied. The value function at $t = 0$ is approximated using $M = 5$. The approximation is used to get the price, and derivated to get the Greeks. This is repeated 1000 times and we report the mean and standard deviation of the estimates. A * represent an estimate which is within a 90% confidence interval of the benchmark value based on 1000 repetitions.
To solve the problem of the large variance on the Greek estimates, one could use a larger ISD. In some cases, good results can be obtained. However, in other cases, it is not possible to get a good compromise between low bias and low variance. Furthermore, there are no indication to determine the largest ISD which does not produce biased results in general. In the next section, we propose a solution that reduces the variance significantly, while keeping the ISD small to avoid significant bias. As shown in Figure 2 and 3, a larger ISD may lead to biased results.

4 Using the Value Function for Variance Reduction

To get unbiased estimates of price and Greeks, one needs to use a small ISD, but that results in estimates with large variance. Increasing the order of the polynomial will decrease the bias, but the variance will increase and numerical issues will occur. Without prior knowledge of the best ISD it is very difficult to get satisfactory results in general. To verify whether the method can give good results in practice, we gathered results when repeating the simulation 100,000 times and took the average and standard deviation. When doing that many simulation, results are good, but it is not useful in practical applications. The problem is that estimates are either biased, or have a variance which is too large. Here we propose a new method to reduce the variance of Greek estimates when the ISD is small.

In the original Least Squares Monte Carlo method, at \( t = 0 \), the payoff for each path is discounted from the optimal exercise time determined by \( \tau \), the optimal stopping time. Figure 7 shows the sample of data available at \( t = 0 \) for the regression. Sub-figure 7(a) shows the data for a large ISD. The data displays structure. When a function is fit through the data it will approximate the true function relatively well overall, though there might be some small bias on the price and large bias on the Greeks. Sub-figure 7(b) shows the data for a small ISD. The data displays no structure. When a function is fit through the data it will approximate the true function relatively well locally, though there will be a lot of variability for the Greeks.

Our solution is to bring more structure in the data at \( t = 0 \) even when using a small ISD. To do so, we use the information from \( t = 1 \). In the out-of-sample algorithm, at \( t = 1 \), \( \hat{F}_N^M(X(t_1)) \) is

28
Sample data used in the regression at $t = 0$ for a large ISD with $\alpha = 0.50$ and a small ISD with $\alpha = 0.05$, for a 1 year to maturity option with $S(0) = 40$, volatility of 20%, a strike $K = 40$, a risk free rate of $r = 0.06$ and 50 early exercises.

estimate from the first set of paths, and then applied to the second set of paths. At this point, we estimate the holding value function by regressing using the the paths from the second set which should not be exercised (and we include OTM paths). Next, the value function is build using the exercise value function for executed paths and the holding value function for held paths. Next, at $t = 0$, instead of discounting the payoffs for each paths from when it is optimal to exercise, we discount the value function from $t = 1$ over one period. This additional simple step removes a lot of variability in the sample data at $t = 0$. Figure 7(c) shows the sample data when discounting the value function from $t = 1$ for $\alpha = 0.05$. Note how compact the data is compared to Sub-figure 7(b).

Note, throughout the LSM algorithm, only in the money paths are used in the regressions. That is because adding OTM paths increases the size of the ISD, i.e. the bandwidth for the regression and increases the bias. However, to implement our solution, one needs to use all paths at $t = 1$ as well as at $t = 0$.

Figure 8 compares the standard deviation from the normal method to the standard deviation of the method using the value function from $t = 1$. The standard deviation of the normal method using $M = 3$ is used as a reference. We can observe that for the price estimates, the reduction is relatively small. However, for the Greeks, the reduction in standard deviation is dramatic, cutting the standard deviation in half for a small ISD. Since a small ISD is required to avoid a biased
Figure 8: Reduction of the standard deviation by using the Value Function from $t = 1$. Compares the standard deviation relative to the standard deviation when using $M = 3$. A sample of 100,000 paths are simulated from an ISD using $S_0^d = S_0 \exp(\alpha \sigma \sqrt{T} \epsilon_d)$ for $\alpha \in \{0.010; \ldots; 0.500\}$. For each alpha, we regress the approximated payoff on the asset price using a polynomial of order $M \in \{3; 5; 7; 9\}$. The estimation of the option price is obtained by evaluating the polynomial at $S(0)$, while Delta and Gamma are obtained by evaluating the first and second derivative of the polynomial, respectively. The graphs show the relative variance on the estimates as a function of the ISD size determined by the factor $\alpha$. The optimal stopping time is estimated using $M = 9$. Solid lines represent the standard deviations for the normal method, while the dashed line represents the standard deviations for the Value Function method.

Estimates, this method will be crucial to keep the standard deviation to an acceptable level.

Table 3 shows the results when applying the Value Function method using the out of sam-
ple algorithm. Most of the price estimates are statistically different from the benchmark values. However, the bias is generally very small. The maximum absolute error is 0.16%, slightly smaller than for the basic method. Most of the Greek estimates are not statistically different from the benchmark values. The only problematic Greek estimates and the Gammas for three options which should be exercised at time. If we discard the results for these three options, the maximum absolute error of Greek estimates is 2.85%. Furthermore, the standard deviation of the estimates is much smaller than for the basic method.

5 Conclusion

Simulation techniques have gained importance for option pricing because of their flexibility. By now efficient methods exist for pricing European as well as American style derivatives. However, as important as option pricing is in real applications, calculating option price sensitivities is equally if not even more important. These risk sensitivities are also called the Greeks and are used by financial institutions that will need these not only once (when the derivative is priced and traded) but continuously through the life of the option for hedging and risk assessment purposes.

Though several methods have been developed for calculating the Greeks of European style options less research has dealt with the issue when options have American style features. One reason is the fact that one needs to simultaneously determine the optimal early exercise strategy significantly complicates matters.

This paper examines the value of using initially dispersed paths, i.e. paths that are started at slightly different values, together with an initial cross sectional regression for estimating prices and Greeks. We analyze carefully a benchmark case and shows that one needs to carefully pick the amount of initial dispersion. The optimal choice depends on the option being priced and the number of paths being used.

To provide a method which is generally applicable we use local polynomial regression, or LPR, which provides an optimal bandwidth and weighs the paths accordingly. We modify this method such that the initial dispersed paths are rescaled accordingly and this method which we call the
LPR-OLS improves on the bias of the OLS and reduces the variance of the LPR method.

References


A Finite Difference method for American Options

In this appendix we provide numerical results on the performance of the Finite Difference method to numerical differentiation for estimating the Greeks with simulation. As explained in the text this is the simplest possible of the three standard methods to implement and the only one which is generally applicable. We use the central Finite Difference method so this involve pricing 3 options. We consider here two applications of the methodology: 1) using the same estimated stopping time for all evaluations and 2) re-estimating the stopping time for each evaluation. Irrespective of which method is used the same random numbers are used to calculate simulated stock prices irrespective of the initial stock price as proposed by Boyle, Broadie, and Glasserman (1997).

In Figure 9 we plot the estimated Delta and Gamma for the two methods as a function of $\Delta S$, that is the difference in the initial stock price used in the Finite Difference approximation, along with the benchmark values obtained from the Binomial Model. We consider an option with a strike price of $K = 44$ in a BSM world with a volatility of $\sigma = 20\%$. The option is assumed to mature in one year and has 50 early exercise points. The risk free rate is $r = 6\%$. The initial asset price is $S_0 = 40$. The reported estimates are averages of 100 independent simulations with 100,000 paths using the standard LSM method in which the cross sectional regressions uses a 3rd order polynomial fitted on the in the money paths only.

The figure shows that for this option the second method that uses re-estimated stopping times generally produces Greeks that are closer to the benchmark values than the method that uses the same stopping time for all price calculations. This is particularly so for the estimate of Delta and for this sensitivity it is not a viable strategy to decrease nor to increase $\Delta S$ as estimates do not converge as Figure 9(a) clearly shows. The estimated Gamma, however, does seem to converge to the true value as $\Delta S$ increases. However, for this particular Greek the method that uses re-estimated stopping times converges faster and is almost always less biased.
Figure 9: Finite Difference estimates of Delta and Gamma as a function of $\Delta S$

This figure plots the Delta and Gamma as a function of $\Delta S$ calculated with the (central) finite difference method using a method that uses the same estimated stopping time for all evaluations (LSM no resim) and a method that re-estimates the stopping time for each evaluation (LSM resim). We consider an option with a strike price of $K = 44$ in a BSM world with a volatility of $\sigma = 20\%$. The option is assumed to mature in one year and has 50 early exercise points. The risk free rate is $r = 6\%$. The initial asset price is $S_0 = 40$. The reported estimates are averages of 100 independent simulations with 100,000 paths using the standard LSM method in which the cross section regressions uses a 3rd order polynomial fitted on the in the money paths only.
Table 3: Estimation of Price and Greeks using the Last Step Value Function Method and the Out-Of-Sample Algorithm.

<table>
<thead>
<tr>
<th>k</th>
<th>σ</th>
<th>T</th>
<th>BM Estimate (std)</th>
<th>Delta BM Estimate (std)</th>
<th>Gamma BM Estimate (std)</th>
</tr>
</thead>
<tbody>
<tr>
<td>36</td>
<td>10%</td>
<td>0.5</td>
<td>0.0304 (0.0006)</td>
<td>-0.0281 (0.0013)</td>
<td>0.0236 (0.0088) *</td>
</tr>
<tr>
<td>36</td>
<td>20%</td>
<td>0.5</td>
<td>0.4978 (0.0037)</td>
<td>-0.1607 (0.0045)</td>
<td>0.0449 (0.0221) *</td>
</tr>
<tr>
<td>36</td>
<td>40%</td>
<td>0.5</td>
<td>2.1993 (0.0110)</td>
<td>-0.2759 (0.0089) *</td>
<td>0.0305 (0.0190) *</td>
</tr>
<tr>
<td>36</td>
<td>10%</td>
<td>1.0</td>
<td>0.0895 (0.0012)</td>
<td>-0.0545 (0.0022) *</td>
<td>0.0305 (0.0092) *</td>
</tr>
<tr>
<td>36</td>
<td>20%</td>
<td>1.0</td>
<td>0.9166 (0.0062)</td>
<td>-0.1979 (0.0061) *</td>
<td>0.0381 (0.0145) *</td>
</tr>
<tr>
<td>36</td>
<td>40%</td>
<td>1.0</td>
<td>3.4366 (0.0155)</td>
<td>-0.2863 (0.0082) *</td>
<td>0.0227 (0.0102) *</td>
</tr>
<tr>
<td>36</td>
<td>10%</td>
<td>2.0</td>
<td>0.1713 (0.0018)</td>
<td>-0.0751 (0.0022) *</td>
<td>0.0305 (0.0092) *</td>
</tr>
<tr>
<td>36</td>
<td>20%</td>
<td>2.0</td>
<td>1.4317 (0.0075)</td>
<td>-0.2165 (0.0070) *</td>
<td>0.0381 (0.0145) *</td>
</tr>
<tr>
<td>36</td>
<td>40%</td>
<td>2.0</td>
<td>4.9643 (0.0184)</td>
<td>-0.2786 (0.0087) *</td>
<td>0.0227 (0.0102) *</td>
</tr>
<tr>
<td>40</td>
<td>10%</td>
<td>0.5</td>
<td>0.7347 (0.0031)</td>
<td>-0.4088 (0.0093) *</td>
<td>0.1846 (0.1097) *</td>
</tr>
<tr>
<td>40</td>
<td>20%</td>
<td>0.5</td>
<td>1.7915 (0.0072)</td>
<td>-0.4256 (0.0099) *</td>
<td>0.0790 (0.0563) *</td>
</tr>
<tr>
<td>40</td>
<td>40%</td>
<td>0.5</td>
<td>3.9718 (0.0151)</td>
<td>-0.4186 (0.0127) *</td>
<td>0.0367 (0.0282) *</td>
</tr>
<tr>
<td>40</td>
<td>10%</td>
<td>1.0</td>
<td>0.8893 (0.0037)</td>
<td>-0.3901 (0.0082) *</td>
<td>0.1505 (0.0532) *</td>
</tr>
<tr>
<td>40</td>
<td>20%</td>
<td>1.0</td>
<td>2.3141 (0.0091)</td>
<td>-0.4040 (0.0106) *</td>
<td>0.0597 (0.0280) *</td>
</tr>
<tr>
<td>40</td>
<td>40%</td>
<td>1.0</td>
<td>5.3120 (0.0191)</td>
<td>-0.3930 (0.0104) *</td>
<td>0.0265 (0.0136) *</td>
</tr>
<tr>
<td>40</td>
<td>10%</td>
<td>2.0</td>
<td>1.0241 (0.0042)</td>
<td>-0.3729 (0.0078) *</td>
<td>0.1301 (0.0296) *</td>
</tr>
<tr>
<td>40</td>
<td>20%</td>
<td>2.0</td>
<td>2.8846 (0.0105)</td>
<td>-0.3796 (0.0102) *</td>
<td>0.0468 (0.0166) *</td>
</tr>
<tr>
<td>40</td>
<td>40%</td>
<td>2.0</td>
<td>6.9171 (0.0218)</td>
<td>-0.3552 (0.0102) *</td>
<td>0.0195 (0.0082) *</td>
</tr>
<tr>
<td>44</td>
<td>10%</td>
<td>0.5</td>
<td>3.9473 (0.0025)</td>
<td>-0.9998 (0.0194) *</td>
<td>0.0010 (0.2581) *</td>
</tr>
<tr>
<td>44</td>
<td>20%</td>
<td>0.5</td>
<td>4.3091 (0.0091)</td>
<td>-0.7563 (0.0159) *</td>
<td>0.0907 (0.0981) *</td>
</tr>
<tr>
<td>44</td>
<td>40%</td>
<td>0.5</td>
<td>6.3262 (0.0182)</td>
<td>-0.5637 (0.0167) *</td>
<td>0.0389 (0.0375) *</td>
</tr>
<tr>
<td>44</td>
<td>10%</td>
<td>1.0</td>
<td>3.9474 (0.0024)</td>
<td>-0.9989 (0.0147) *</td>
<td>0.0054 (0.1270) *</td>
</tr>
<tr>
<td>44</td>
<td>20%</td>
<td>1.0</td>
<td>4.6535 (0.0108)</td>
<td>-0.6648 (0.0153) *</td>
<td>0.0765 (0.0446) *</td>
</tr>
<tr>
<td>44</td>
<td>40%</td>
<td>1.0</td>
<td>7.6104 (0.0220)</td>
<td>-0.4966 (0.0127) *</td>
<td>0.0291 (0.0172) *</td>
</tr>
<tr>
<td>44</td>
<td>10%</td>
<td>2.0</td>
<td>3.9480 (0.0026)</td>
<td>-0.9963 (0.0102) *</td>
<td>0.0161 (0.0664) *</td>
</tr>
<tr>
<td>44</td>
<td>20%</td>
<td>2.0</td>
<td>5.0832 (0.0127)</td>
<td>-0.5897 (0.0127) *</td>
<td>0.0639 (0.0232) *</td>
</tr>
<tr>
<td>44</td>
<td>40%</td>
<td>2.0</td>
<td>9.1820 (0.0250)</td>
<td>-0.4342 (0.0119) *</td>
<td>0.0219 (0.0096) *</td>
</tr>
</tbody>
</table>

An ISD is created using $\alpha = 1$ and 100,000 paths are simulated. An optimal exercise strategy is approximated using the simulated paths and $\alpha = 100,000$ paths are simulated. The optimal exercise strategy obtained from the previous step is applied. The holding function at $t = 1$ is approximated using only paths which are held and $M = 5$. The value function from $t = 1$ is discounted over one period and the value function at $t = 0$ is approximated using $M = 5$. The approximation is then used to get the price, and derivated to get the Greeks. This is repeated 1000 times and we report the mean and standard deviation of the estimates. A * represent an estimate which is within a 90% confidence interval of the benchmark value based on 1000 repetitions.