Simulated Greeks for American Options*

Pascal Letourneau[†] and Lars Stentoft[‡]

March 9, 2022

Abstract

This paper develops a method to estimate price sensitivities, so-called Greeks, for American style options using flexible simulation methods combined with initially dispersed state variables. The asymptotic properties of the estimators are studied, and convergence of the method is established. A 2-stage method is proposed with an adaptive choice of optimal dispersion of state variables, which controls and balances off the bias of the estimates against their variance. Numerical results show that the method compares exceptionally well to existing alternatives, works well for very reasonable choices of dispersion sizes, regressors, and simulated paths, and it is robust to choices of these parameters. We apply the method to models with time varying volatility demonstrating that there are large differences between estimated Greeks with affine and non-affine models, that Greeks vary significantly through periods of crisis, and that the errors made when using Greeks implied from, e.g., misspecified models with constant volatility can be extremely large.

JEL Classification: C15, G12, G13

Keywords: Finance, Hedging, Least Squares Monte Carlo method, Price sensitivities.

*The authors thank participants at the 2014 SIAM Conference on Financial Mathematics & Engineering, the 2016 Financial Management Association Annual Meeting, and the 62nd Annual CORS Conference and at seminars at IIT Stuart School of Business, University of Western Ontario and Montreal-Quebec Actuarial and Financial Mathematics Laboratory (Quantact) for valuable comments. Lars Stentoft acknowledges support from CREATES - Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation. The paper was previously circulated with the title "Improved Greeks for American Options using Simulation".

[†]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.

1 Introduction

Option pricing, particularly for options that have early exercise features, remains a challenge. Moreover, while option pricing is interesting in itself, a much more important problem is calculating the various relevant hedging parameters or price sensitivities that market participants rely on for managing their positions. After all, you only price a derivative once, but once it is traded its risk exposures will generally need to be hedged repeatedly through time. Price sensitivities, or Greeks, 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 efficiently. Unlike prices which are observed in the market place, the Greeks are generally not observed and will instead have to be estimated. When the goal is to find a flexible method that is generally applicable and can be used in various settings, Monte Carlo simulation is essentially the only viable choice. This method has been used at least since Boyle (1977) to price European style derivatives, in general, and options, in particular. Simulation methods are flexible, easy to apply, and have nice properties since averages of independent random observations converge to expected values under mild assumptions. However, developing efficient and flexible methods that yield not only prices but also Greeks in realistic settings remains an open question and active area of research.

This paper studies the joint estimation of prices and Greeks for American style options using flexible simulation and regression based Monte Carlo methods combined with initially dispersed state variables. First, we contribute to the literature by deriving the asymptotic properties of the suggested estimators, obtained from an initial cross sectional regression, and we prove their convergence. Second, based on our theoretical developments, we propose an efficient 2-stage method, which combines the flexible simulation method with an adaptive choice of optimal initially dispersed state variables to automatically control and balance off the bias of the estimates against their variance. The proposed method is straightforward to implement in practice, provides precise and unbiased estimates of the price and Greeks, and compares exceptionally well to existing alternatives. Our numerical results show that it works well for very reasonable choices of dispersion sizes, regressors, and simulated paths, and it is robust to choices of these parameters. Finally, we apply our method to models with time-varying volatility, demonstrating that such features are important and that using misspecified models can lead to very large errors in the estimated Greeks.

Given an early exercise strategy calculating the Greeks for an American option is no more complicated than it is for an European style option.[1] For example, one could estimate the option Delta using the "finite difference approach" by simulating at two different values of the underlying stock price, a viable method when the two starting values are close enough by a smoothness argument. A drawback of the finite difference approach is that it is biased and potentially very inefficient, and since there is no guarantee that option payoffs are continuous in the underlying state variables, it may result in poorly behaved or non-existing estimates of the Greeks. Alternatives to this approach include the "pathwise derivative method" and the "likelihood ratio method". For applications of the pathwise derivative method see Piterbarg (2014) and for a recent approach that essentially uses the likelihood ratio method see Kaniel, Tompaidis, and Zemliano (2008). Though these two methods do produce unbiased estimates, possible drawbacks are that they require the payoff, respectively, the probability density be differentiable, which may not always be the case.

This paper considers instead simulation methods that can be used to jointly determine prices and Greeks. There are several reasons why this is preferred. First, by simultaneously estimating the stopping time strategy any issue of non-continuity in the state variables is mitigated. Second, a joint method is computationally more efficient as it does not require additional simulations to determine the Greeks. Finally, and most importantly, the numerical performance, in general, and the convergence, in particular, of any method that uses an exogenously given estimate of the optimal stopping time is clearly conditional on this particular estimate. Thus, it is difficult if not impossible to make any arguments about the actual performance of these methods that are generally and unconditionally applicable. Methods that jointly determine prices and Greeks, on the other hand, are easier to examine in terms of their convergence rates and numerical performance. In this paper we prove convergence of our proposed method and demonstrate numerically that optimal, in a Monte Carlo simulation context, convergence rates can be obtained.

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. Jain and Oosterlee (2015) instead suggest that Delta can be approximated using a finite difference

[1]For a general overview of methods available for estimating Greeks with simulation see, e.g., Glasserman (2004).

approach, in which the regression coefficients from the first early exercise points are used. This method, however, requires a "regress-later" type approach (see Glasserman and Yu (2002)), and therefore requires regressors that are martingales or have one step ahead conditional expectations known in closed form or with analytical approximations.[2] This restriction significantly limits the choice of potential regressors, and for more complicated models the proposed method would be infeasible. Finally, Chen and Liu (2014) considers jointly pricing options and estimating first order Greeks like the Delta using the pathwise derivative method mentioned above.

The method we use is, however, closest in spirit to Wang and Caflisch (2010) who suggests estimating Greeks by performing an additional regression using an initial dispersed sample of the state variables.[3] Though the idea behind using dispersed state variables for estimating Greeks is intuitive and simple, our theoretical and numerical results document that the quality of the estimates crucially depends on how and by how much state variables are initially dispersed and on the choice of the number of paths in the simulation and the order of the polynomial approximation used. In fact, our results show that blindly using the suggestions in Wang and Caflisch (2010) can result in statistically as well as economically significantly biased results. To address this shortcoming, we carefully develop a 2-stage method which combines the flexible simulation method with an adaptive choice of optimal initially dispersed state variables to control and balance off the bias of the estimates against their variance.

Our numerical results document that our method works extremely well for very reasonable choices of the number of regressors and simulated paths for a sample of options with empirically relevant characteristics. Moreover, the practical guidelines we develop are simple to implement and the resulting method is robust. We compare our proposed method to the pathwise derivative method, the likelihood ratio method, as well as the method suggested by Wang and Caflisch (2010). The results show that, among methods that provide statistically insignificant estimates of prices and Greeks, our proposed method is the most precise and provides estimates with the smallest bias. Since the computational complexity of these methods is roughly equivalent our results clearly demonstrate the value and importance of our suggested method.

The method we propose relies on nothing but simple regression based Monte Carlo simulation

[2]The 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.

[3]This is similar to using the extended Binomial Model to obtain Greeks (see Pelsser and Vorst (1994)).

methods and is extremely flexible easy to use. Thus we successfully proposes a method that can be used for simultaneous estimation of prices and Greeks for American style options in very general settings. For example, it is straightforward to obtain the Delta as well as the Gamma, and it is simple to apply our proposed method to, e.g., the multivariate case where sensitivities to each of the underlying assets along with cross-sensitivities should be estimated. More generally, our proposed methodology can be applied to obtain sensitivities with respect to any of the state variables that determining an option's price, which besides the value of the underlying asset include the volatility of this asset, the interest rate, and potentially the dividend yield. Our method should therefore be of immediate interest and have broad applications.

To showcase our proposed methodology, we consider an empirically relevant application illustrating how to obtain option Greeks in models with time varying volatility. In this setting, closed form benchmarks only exist for European options in a very restricted affine class of models. While our method straightforwardly replicates such benchmarks, it can be used to obtain results for American style options as well and more generally to compare different model specifications. Our numerical results first demonstrate that there are large differences between estimated Greeks from affine and simple alternative non-affine models that have been shown empirically to fit data better. Moreover, we show that estimated Greeks vary significantly through periods of crisis, and that the errors made when using Greeks implied from models with constant volatility can be extremely large when implied spot volatilities differ from their unconditional values. A direct implication is that a position in derivatives hedged using market implied Greeks would in fact be exposed to massive amounts of risk when volatility is time varying.

The rest of the paper is structured as follows: Section 2 provides the framework and discusses the idea behind obtaining Greeks using simulation and initial state dispersion, proves convergence of our proposed method theoretically, and provides numerical estimates of the convergence rates. Section 3 discusses several issues related to the implementation of the methodology, develops our proposed 2-stage method, and compares it to excisting methods from the literature. Section 4 demonstrates that the proposed method can be applied in a setting with time varying volatility, that it leads to estimated Greeks that differ across volatility regimes, and that hedging errors made when using wrong model dynamics can be extremely large. Finally, Section 5 concludes. Appendix A contains proofs and additional technical details and Appendix B contains some robustness checks.

2 Framework

The first step in implementing a numerical algorithm to price early exercise options is to assume that time can be discretized. We specify J exercise points as $t_0 = 0 < t_1 \leq t_2 \leq ... \leq t_J = T$, with t_0 and T denoting the current time and maturity of the option, respectively. Thus, we are approximating the American option by the so-called Bermudan option. The American option price is obtained in the limit by increasing the number of exercise points, J, see also Bouchard and Warin (2012) for a formal justification of this approach.[4] We assume a complete probability space $(\Omega, \mathcal{F}, \mathbb{Q})$ equipped with a discrete filtration $\left(\mathcal{F}\left(t_j\right)\right)_{j=0}^{J}$ and a unique pricing measure corresponding to the probability measure $\mathbb{Q}$. The derivative's value depends on one or more underlying assets modeled using a Markovian process, with state variables $\left(X\left(t_j\right)\right)_{j=0}^{J}$ adapted to the filtration. We denote by $\left(Z\left(t_j\right)\right)_{j=0}^{J}$ an adapted discounted payoff process for the derivative satisfying $Z\left(t_j\right) = \pi\left(X\left(t_j\right), t_j\right)$ for a suitable function $\pi\left(\cdot, \cdot\right)$ 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 π (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

$$P\left(X\left(t_0\right) = x_0\right) = \max_{\tau(t_1) \in \mathcal{T}(t_1)} \mathrm{E}\left[Z\left(\tau\right) | X\left(0\right) = x_0\right], \tag{1}$$

where $\mathcal{T}\left(t_j\right)$ denotes the set of all stopping times with values in $\{t_j, ..., t_J\}$, and x_0 is the vector of initial states.

The problem of calculating the option price in Equation (1) with $J > 1$ is referred to as a discrete time optimal stopping time problem and typically solved using the dynamic programming principle. Intuitively, this procedure can be motivated by considering the choice faced by the option holder at time t_j. The optimal choice is made by comparing the value from exercising the option immediately to the value from holding the option until the next period and behaving optimally onwards. Let $V\left(X\left(t_j\right)\right)$ denote the value of the option for state variables X at a time t_j prior to expiration and define $F\left(X\left(t_j\right)\right) \equiv \mathrm{E}[Z\left(\tau\left(t_{j+1}\right)\right) | X\left(t_j\right)]$ as the conditional expected payoff, where

[4] We do not stress any further this difference as the literature on pricing early exercise options using simulation generally refers to these as American style options, see, e.g., Longstaff and Schwartz (2001).

$\tau(t_{j+1})$ is the optimal stopping time. It then follows that

$$V(X(t_j)) = \max(Z(t_j), F(X(t_j))), \tag{2}$$

and the optimal stopping time can be derived iteratively as

$$\begin{cases} \tau(t_J) = T \\ \tau(t_j) = t_j \mathbf{1}_{\{Z(t_j) \geq F(X(t_j))\}} + \tau(t_{j+1}) \mathbf{1}_{\{Z(t_j) < F(X(t_j)=x)\}}, \quad 1 < j \leq J-1. \end{cases} \tag{3}$$

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

$$P(X(t_0) = x_0) = F(X(t_0)) = \mathrm{E}[Z(\tau(t_1)) | X(t_0) = x_0]. \tag{4}$$

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

2.1 Simulation and regression based prices and Greeks

It is generally not possible to implement the exact algorithm in Equation (3) because the conditional expectations are unknown and therefore the price estimate in Equation (4) is infeasible. Instead an approximate algorithm is needed, and among the different approaches available we study simulation based methods. In a standard Monte Carlo simulation, paths for the state variables are simulated from a single starting value x_0. In this paper, we consider option pricing with different deterministically determined starting values at $t = 0$. Note that, because of the Markovian assumption it does not matter where we start the process, and this does not change the determination of the pathwise optimal early exercise strategy at any of the potential early exercise points. Using initially dispersed state variables is key to our approach for determining option Greeks. We refer to the technique of simulating paths from different starting values as Initial State Dispersion (ISD).

Because conditional expectations can be represented as a countable linear combination of basis functions we may write $F(X(t_j)) = \sum_{m=0}^{\infty} \phi_m(X(t_j)) a_m(t_j)$, where $\{\phi_m(\cdot)\}_{m=0}^{\infty}$ form a basis.[5] To make this operational we further assume that the conditional expectation function can be well approximated with the first $M+1$ terms such that $F(X(t_j)) \approx F_M(X(t_j)) = \sum_{m=0}^{M} \phi_m(X(t_j))\, a_m(t_j)$, and that we can obtain an estimate of this function by

$$\hat{F}_M^N(X(t_j)) = \sum_{m=0}^{M} \phi_m(X(t_j))\, \hat{a}_m^N(t_j), \tag{5}$$

where the set of coefficients $\left\{\hat{a}_m^N(t_j) : m \in [0:M], j \in [1:J]\right\}$ is approximated or estimated at each time j using $N \geq M$ simulated paths. For example, in the Least Squares Monte Carlo (LSM) method of Longstaff and Schwartz (2001) these are determined from cross-sectional regressions of the discounted future pathwise payoffs on transformations of the state variables.

Based on the estimate in Equation (5) we can derive an estimate of the optimal stopping time as

$$\begin{cases} \hat{\tau}_M^N(t_J) = T \\ \hat{\tau}_M^N(t_j) = t_j \mathbf{1}_{\{Z(t_j) \geq \hat{F}_M^N(X(t_j))\}} + \hat{\tau}_M^N \mathbf{1}_{\{Z(t_j) < \hat{F}_M^N(X(t_j))\}}, \quad 1 < j \leq J-1. \end{cases} \tag{6}$$

From the algorithm in Equation (6) a natural estimate of the option value in Equation (4) is given by

$$\hat{P}_M^N(X(t_0) = x_0) = \hat{F}_M^N(X(t_0)) = \mathrm{E}[Z\left(\hat{\tau}_M^N(1)\right) | X(0) = x_0]. \tag{7}$$

In other words, the approximated option price equals the conditional expected payoff.

The algorithm in Equation (6) provides an optimal stopping time for each simulated path which can be used to determine the pathwise payoffs. In the special case when all the paths are started at the current values of the state variables, the conditional expectation in Equation (7) can be estimated by the sample average and an estimate of the option price is given by

$$\hat{P}_M^N(X(t_0) = x_0) = \frac{1}{N} \sum_{n=1}^{N} Z\left(n, \hat{\tau}_M^N(1,n)\right), \tag{8}$$

where $Z\left(n, \hat{\tau}_M^N(1,n)\right)$ is the payoff from exercising the option at the estimated optimal stopping time $\hat{\tau}_M^N(1,n)$ determined for path n according to Equation (6). This type of algorithm has nice

[5] This assumption is justified when approximating elements of the L^2 space of square-integrable functions relative to some measure. Since L^2 is a Hilbert space, it has a countable orthonormal basis (see, e.g., Royden (1988)).

properties and Stentoft (2014) documents that it is the most efficient method when compared to, e.g., the value function iteration method of Carriere (1996) or Tsitsiklis and Van Roy (2001).

In the general case with an ISD, $Z\left(n, \hat{\tau}_M^N(1,n)\right)$ corresponds to the payoff from exercising the option at the estimated optimal stopping time $\hat{\tau}_M^N(1,n)$ from Equation (6) for paths that are started at different initial values of the state variables and we can not simply average these to obtain an estimate of the price. Instead the simulated paths can be used to estimated the price function $\hat{P}_M^N(X(0)) = \mathrm{E}[Z\left(\hat{\tau}_M^N(1)\right)|X(0)]$. This is again a conditional expectation and we assume it can be well approximated using the first M_0+1 terms of a basis $\{\rho_m(\cdot)\}_{m=0}^{\infty}$, and that we can obtain an estimate of this function by

$$\hat{P}_{M,M_0}^N(X(0)) = \sum_{m=0}^{M_0} \rho_m(X(0))\, \hat{b}_m(0), \tag{9}$$

where the coefficients are estimated with a cross-sectional regression of $N \geq M_0$ discounted future pathwise payoffs on transformations of the initially dispersed state variables. Note that the number of terms and/or the basis used at $t=0$ can be different from what is used to estimate the optimal stopping time.

Given this approximation, an estimate of the option price for a given value of the state variables $X(0)=x$ is obtained by evaluating the approximation $\hat{P}_{M,M_0}^N$ at this value and hence we have

$$\hat{P}_{M,M_0}^N(X(0)=x) = \sum_{m=0}^{M_0} \rho_m(X(0)=x)\, \hat{b}_m(0). \tag{10}$$

In a similar way, we can estimate the sensitivity of the option price at $X(0)=x$ with respect to state variable X^i as

$$\frac{\partial \hat{P}_{M,M_0}^N(X(0)=x)}{\partial X^i} = \sum_{m=0}^{M_0} \frac{\partial \rho_m(X(0)=x)}{\partial X^i} \hat{b}_m(0) = \sum_{m=0}^{M_0} \rho_m'(X(0)=x)\, \hat{b}_m(0). \tag{11}$$

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

In the special case where the only state variable is the stock price, $(S(t_j))_{j=0}^{J}$, the formulas for the price, the first derivative, the Delta or Δ, and the second derivative, the Gamma or Γ, at

$S(0) = s$ are given by

$$\hat{P}^N_{M,M_0}(S(0)=s) = \sum_{m=0}^{M_0} \rho_m(S(0)=s)\,\hat{b}_m(0), \tag{12}$$

$$\hat{\Delta}^N_{M,M_0}(S(0)=s) = \sum_{m=0}^{M_0} \rho'_m(S(0)=s)\,\hat{b}_m(0), \text{ and} \tag{13}$$

$$\hat{\Gamma}^N_{M,M_0}(S(0)=s) = \sum_{m=0}^{M_0} \rho''_m(S(0)=s)\,\hat{b}_m(0), \tag{14}$$

respectively.

2.2 Theoretical results

The first step needed to prove convergence of LSM type estimators is to establish convergence of the estimated approximate conditional expectation function. This is done in the following Lemma.

Lemma 1 (Adapted from Theorem 2 of Stentoft (2004)). *Under some regularity and integrability assumptions on the conditional expectation function, F, see Stentoft (2004) for details, if $M = M(N)$ is increasing in N such that $M \to \infty$ and $M^3/N \to 0$, then $\hat{F}^N_M(X(t_j))$ converges to $F(X(t_j))$ in probability for $j = 1, ..., J$.*

Proof. See Stentoft (2004). □

It is now straightforward to prove that the standard LSM estimator in Equation (8), in which all the paths are started at the same value, converges. To do this we need the following Lemma.

Lemma 2 (Adapted from Proposition 1 of Stentoft (2004)). *Assuming the simulated paths are independent and started at the same value x, if the sets of coefficients $\left\{\hat{a}^N_m(t_j) : m \in [0:M], j \in [1:J]\right\}$ converges to $\left\{a^N_m(t_j) : m \in [0:M], j \in [1:J]\right\}$, then $\hat{P}^N_M(X(0)=x)$ converges to $P(X(0)=x)$ in probability.*

Proof. See Stentoft (2004). □

Combining the results in Lemma 1 and 2 demonstrates that when all the simulated paths are started at the current values of the state variable, then the estimate in Equation (8) converges to the true price in Equation (4). This then establishes convergence of the LSM method in a general multi-period setting.

In the general case where an ISD is used, Lemma 1 continues to hold and demonstrates that the approximated optimal stopping time converges. However, in this case more care has to be taken when analysing the properties of the estimated price and Greeks, and we need to study the regression at t_0 carefully. Specifically, what we are interested in is essentially estimating an unknown regression function $P(x_0) = \mathrm{E}[Z|X = x_0]$ and its derivatives $P^{(M)}(x_0)$ given a sample (X_1, Z_1), ..., (X_N, Z_N) of initial states and pathwise payoffs from the population (X, Z). Suppose the $(M+1)$'th derivative exists at x_0, we can then approximate the regression function $P(x)$ locally using a Taylor series expansion for x in a neighborhood of x_0 as

$$P(x) \approx P(x_0) + P^{(1)}(x_0)(x - x_0) + \frac{1}{2!}P^{(2)}(x_0)(x - x_0)^2 + \cdots + \frac{1}{M!}P^{(M)}(x_0)(x - x_0)^M. \quad (15)$$

This is a Local Polynomial Regression (LPR) problem which is solved by fitting locally the weighted least squares regression problem

$$\min_{\beta_i} \sum_{n=1}^{N} \{Z_n - \sum_{i=0}^{M} \beta_i (X_n - x_0)^i\}^2 \mathcal{K}_h (X_n - x_0), \quad (16)$$

where $\mathcal{K}_h(t) = \mathcal{K}\left(\frac{t}{h}\right)/h$ is a continuous kernel function having bounded support assigning weights to each observation and h is a bandwidth parameter. Estimators of the derivatives in Equation (15) are then $\hat{P}^{(i)}(x_0) = i!\hat{\beta}_i$, where $\hat{\beta}_i$, $i = 0, \ldots, M$, denotes the solution to the problem in Equation (16).

LPR was introduced by Stone (1977) and has been extensively studied since (see, e.g., Fan and Gijbels (1996) and references therein). Asymptotic properties of the estimated parameters in Equation (16) as well as asymptotic normality were established by Fan and Gijbels (1996) and strong uniform consistency was shown by Delecroix and Rosa (1996). The theory can be generalized to multivariate problems (see, e.g., Masry (1996)). We will use LPR convergence results to establish the convergence of the price and Greeks estimators in Equations (12), (13), and (14). For this we need the following well known lemma from the literature.

Lemma 3 (Adapted from Theorem 3.1 of Fan and Gijbels (1996)). *Let f_X denote the marginal density of X and let $\sigma^2(x)$ denote the conditional variance of Z given $X = x$. Assume that $f_X(x_0) > 0$ and $f'_X(x_0) = 0$, and that $f_X(\cdot)$, $P^{(M+1)}(\cdot)$, and $\sigma^2(\cdot)$ are continuous in a neighborhood of x_0.*

Further, assume that $h \to 0$ and $N \to \infty$ such that $Nh \to \infty$. Then, for some constants ξ_1, ξ_2, and ξ_3 that depend on M and the kernel used, see Fan and Gijbels (1996) for details, the asymptotic conditional variance of $\hat{P}^{(i)}(x_0)$ is given by

$$Var\{\hat{P}^{(i)}(x_0)|X\} = \xi_1 \frac{i!^2\sigma^2(x_0)}{f(x_0)Nh^{1+2i}} + o_P\left(\frac{1}{Nh^{1+2i}}\right), \tag{17}$$

where $o_P(1)$ denotes a random quantity that tends to zero in probability.
The asymptotic conditional bias for $(M-i)$ odd is given by

$$Bias\{\hat{P}^{(i)}(x_0)|X\} = \xi_2 i! P^{(M+1)}(x_0) h^{M+1-i} + o_P\left(h^{M+1-i}\right), \tag{18}$$

and for $(M-i)$ even it is given by

$$Bias\{\hat{P}^{(i)}(x_0)|X\} = \xi_3 i! P^{(M+2)}(x_0) h^{M+2-i} + o_P\left(h^{M+2-i}\right), \tag{19}$$

provided that $f_X^{'}(\cdot)$ and $P^{(M+2)}(\cdot)$ are continuous in a neighborhood of x_0 and that $Nh^3 \to \infty$.

Proof. This follows from Fan and Gijbels (1996) assuming explicitly that $f_X^{'}(x_0) = 0$. It is possible to relax this assumption without affecting the general results, though this introduces additional terms in Equation (19). □

To provide a more formal setting for how to estimate the price and the Greeks that allows us to prove convergence, we now propose a slight modification to the t_0 approximation in Equation (9). The modification consists in using polynomials formulated in terms of the distance to x_0 as the basis. With a slight abuse of notation we propose to estimate the price and Greeks from the coefficients that solve the following ordinary least squares (OLS) regression problem

$$\min_{b_i} \sum_{n=1}^{N} \{Z_n - \sum_{i=0}^{M_0} b_i (X_n - x_0)^i\}^2. \tag{20}$$

We now state the relevant assumptions under which convergence of the price and Greeks estimators $\hat{P}^{(i)}(x_0) = i!\hat{b}_i$, where $\hat{b}_i$, $i = 0, \ldots, M$ is the solution to Equation (20), can be established and then continue to do so in Theorem 1.

Assumption 1. *The sample of N initially dispersed state variables, X, is generated from a continuous symmetric kernel density, $\mathcal{K}_{ISD}(u)$, with support on $|u| \leq 1$. In particular, we assume that the sample is generated from*

$$X_n = x_0 + \alpha \mathcal{K}_{ISD}(U_n), \tag{21}$$

where U_n is equidistributed on the support and $\alpha \geq 0$ controls the size of the ISD.

Assumption 2. *The pathwise payoffs, Z, are obtained from applying a stopping time $\tau(1)$ to randomly simulated paths starting at X. In particular, we assume that the corresponding values are generated from*

$$Z_n = Z_n(\tau(1, n)), \tag{22}$$

where Z_n corresponds to the discounted payoff along the path starting at X_n.

Assumption 3. *The option price function $P = \mathrm{E}[Z|X]$ is continuously differentiable of at least order $M_0 + 2$ on the support of X.*

Assumption 4. *The estimated stopping time introduces at most weak dependence between the discounted payoffs Z_n.*

Theorem 1. *Under Assumptions 1-4, let $N \to \infty$, let $\alpha \to 0$ such that $N\alpha^3 \to \infty$, and let $M \to \infty$ such that $M^3/N \to 0$. When $N\alpha^{1+2i} \to \infty$ then the estimate of the i'th derivative of the option value function, the $\hat{P}^{(i)}$ obtained with the estimates from the OLS regression in Equation* (20)*, converges to the true i'th derivative $P^{(i)}$.*

Proof. See Appendix A.1. □

Remark 1. *Theorem 1 demonstrates that it is essential for convergence of the estimators that the number of observations, N, tends to infinity and the size of the ISD, α, tends to zero, although α can not tend to zero "too" fast when estimating higher order derivatives since otherwise the variance of the estimates will be unbounded. This is quite intuitive since the estimates correspond to the (local) derivatives and there is therefore naturally no requirement that the polynomial order M_0 tend to infinity. This is very different from Lemma 1, which shows that $\hat{F}_M^N(X(t_j))$ converges to $F(X(t_j))$, for $t_j > 0$, when N and M both tend to infinity together.*

Theorem 1 holds for any estimated stopping time $\hat{\tau}(1)$ and as a special case therefore also proves convergence in the case of a European option. It should be noted that the conditions in Theorem 1, in particular the requirement that $N\alpha^{1+2i} \to \infty$, are sufficient to ensure convergence, though it may be possible to relax these conditions somewhat. For example, in the regular LPR setting a larger and larger fraction of the paths ends up not being used in the regression as the bandwidth h decreases. The setting we propose above, on the other hand, ensures that all N paths are used in the regression even when α tends to zero because we directly control the ISD size.

2.3 Estimated convergence rates

The theoretical results above provide (lower) bounds on the convergence rates that can be achieved with our proposed method. In this section we estimate the actual rates numerically. We know from Theorem 1 that convergence requires that $N \to \infty$, while $\alpha \to 0$. Moreover, when the early exercise strategy is estimated we know that as N increases we need that $M \to \infty$. To study the convergence rates we consider a setup in which the number of paths N is increased, and where α shrinks in proportion to N and M increases in proportion to N, simultaneously. First, to ensure that $M^3/N \to 0$ as required by Lemma 1 we set $N \propto M^4$ to ensure convergence of the estimated early exercise strategy when this is estimated. Next, to examine the convergence rates we set $\alpha \propto N^{-1/p}$ for various values of p positive. Note that Theorem 1 requires that $p > 1 + 2i$ for the estimate of the i'th sensitivity to converge. In our numerical results, we choose $M \in \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$, set $N = 160 \times M^4$, and set the initial value of the bandwidth to $\alpha = 10$. All results are based on 100 independent simulations.

In Table 1 we present estimated convergence rates for the bias, standard deviation and the root mean squared error (RMSE) for different values of p and M_0, respectively. Panel A reports the rates obtained for an American put option and confirms that when p is too low convergence may not be obtained. In particular, when $p = 3$ the estimated convergence rate for Gamma is very close to zero, and when $p = 1$ neither Delta nor Gamma converges. However, as long as p is not too small the price and Greeks converge and the actual rates increase with p. When $p = 21$ all the estimated RMSE convergence rates are quite close to -0.5, the optimal convergence rate for standard Monte Carlo estimates. Panel B reports the rates across various values of M_0 and shows that estimated convergence rates indeed tend to increase with M_0 and this is particularly so for

Table 1: Estimated convergence rates

Panel A: Estimated convergence rates across p

	Price			Delta			Gamma		
p	Bias	StDev	RMSE	Bias	StDev	RMSE	Bias	StDev	RMSE
1	-0.7219	-0.5708	-0.5722	0.2765	0.2065	0.2064	1.1189	1.1707	1.1707
3	-0.5748	-0.5473	-0.5478	-0.1048	-0.3371	-0.3355	-0.0475	-0.0197	-0.0194
5	-0.6317	-0.5481	-0.5486	-0.3231	-0.4096	-0.4097	-0.3089	-0.2821	-0.2819
11	-0.5577	-0.5253	-0.5262	-0.4870	-0.5139	-0.5146	-0.5078	-0.4134	-0.4139
21	-0.5617	-0.5217	-0.5225	-0.5940	-0.5379	-0.5384	-0.5655	-0.4986	-0.4990

Panel B: Estimated convergence rates across M_0

	Price			Delta			Gamma		
M_0	Bias	StDev	RMSE	Bias	StDev	RMSE	Bias	StDev	RMSE
3	-0.6600	-0.5189	-0.5201	-0.4698	-0.4172	-0.4178	-0.2695	-0.3408	-0.3387
5	-0.5751	-0.5124	-0.5127	-0.4809	-0.4290	-0.4293	-0.4459	-0.3334	-0.3340
9	-0.5577	-0.5253	-0.5262	-0.4870	-0.5139	-0.5146	-0.5078	-0.4134	-0.4139
15	-0.5851	-0.5234	-0.5239	-0.0552	-0.4898	-0.4898	-0.3406	-0.4430	-0.4431

This table shows the estimated convergence rates across p and M_0 with an initial bandwidth of $\alpha = 10$. Biases and standard errors are calculated from 100 independent simulations with $N = 160 \times M^4$ paths for $M \in \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$. Panel A reports results across p for $M_0 = 9$. Panel B reports results across M_0 for $p = 11$. The put option considered has $T = 1$ year to maturity, a volatility of $\sigma = 20\%$, an interest rate of $r = 6\%$, the initial stock price is fixed at $S(0) = 40$, and the strike price is $K = 40$.

higher order derivatives. However, the estimated rates may in fact decrease when the polynomial order is increased beyond $M_0 = 9$, i.e., the estimated rate for Delta is marginally lower for $M_0 = 15$. Thus, it seems that for our particular application setting $M_0 = 9$ is a good choice. Unreported results show that the convergence rates for the European option are very similar and estimating the optimal stopping time thus has very little effect on the rates.[6]

When comparing the estimated convergence rates in Table 1 for the bias, the standard deviation, and the RMSE, respectively, it becomes apparent that what drives RMSE convergence is the rate for the standard deviation. The reason for this is that in our particular setup the bias is very small and its contribution to the RMSE is negligible. The relative size of the bias versus the standard deviation is determined by the size of the bandwidth. To explore the impact of the initial bandwidth further, Table 2 reports estimated convergence rates across p and M_0 when a large initial bandwidth of $\alpha = 25$ is used. The table shows that across p estimated rates are quite close to those found in Table 1. However, when $M_0 = 3$ is small, Panel B does show that the estimated convergence rates are now driven by the bias and the rates are significantly lower than those in Table 1, i.e., the estimated rate for Gamma is -0.19 versus -0.34. Even with a large initial bandwidth of $\alpha = 25$ optimal convergence rate can be obtained though, i.e., with $p = 101$ and $M_0 = 15$ the rate for

[6]One reason for this is that the estimated stopping time converges at least as fast and likely faster because of the smoothness of the conditional expectation function (see Theorem 1 of Stentoft (2004) and the discussion therein).

Table 2: Estimated convergence rates with $\alpha = 25$

Panel A: Estimated convergence rates across p									
		Price			Delta			Gamma	
p	Bias	StDev	RMSE	Bias	StDev	RMSE	Bias	StDev	RMSE
1	-0.6447	-0.5828	-0.5825	0.0732	0.2016	0.2003	1.0157	1.0566	1.0566
3	-0.5689	-0.5565	-0.5563	-0.4197	-0.3053	-0.3074	-0.2037	-0.1016	-0.1023
5	-0.4947	-0.5382	-0.5378	-0.4862	-0.3994	-0.4008	-0.4228	-0.3042	-0.3047
11	-0.5490	-0.5163	-0.5166	-0.5425	-0.4679	-0.4693	-0.5616	-0.4011	-0.4031
21	-0.5543	-0.5135	-0.5144	-0.5881	-0.4726	-0.4740	-0.6954	-0.4532	-0.4549
Panel B: Estimated convergence rates across M_0									
		Price			Delta			Gamma	
M_0	Bias	StDev	RMSE	Bias	StDev	RMSE	Bias	StDev	RMSE
3	-0.2485	-0.5339	-0.3020	-0.1381	-0.4872	-0.2443	-0.1914	-0.3276	-0.1916
5	-0.5577	-0.4949	-0.4974	-0.6912	-0.4430	-0.4584	-0.5051	-0.2490	-0.2591
9	-0.5490	-0.5163	-0.5166	-0.5425	-0.4679	-0.4693	-0.5616	-0.4011	-0.4031
15	-0.6038	-0.5318	-0.5332	-0.5682	-0.4853	-0.4852	-0.7065	-0.4559	-0.4609

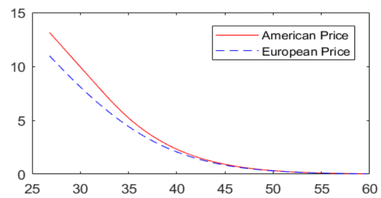
This table shows the estimated convergence rates across p and M_0 with an initial bandwidth of $\alpha = 25$. Biases and standard errors are calculated from 100 independent simulations with $N = 160 \times M^4$ paths for $M \in \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$. Panel A reports results across p for $M_0 = 9$. Panel B reports results across M_0 for $p = 11$. The put option considered has $T = 1$ year to maturity, a volatility of $\sigma = 20\%$, an interest rate of $r = 6\%$, the initial stock price is fixed at $S(0) = 40$, and the strike price is $K = 40$.

Gamma is estimated at -0.5.

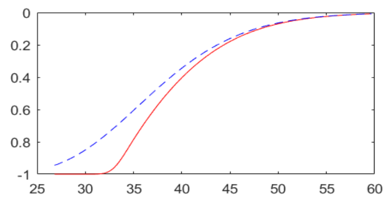
3 Finite sample results

The idea of using initially dispersed state variables to estimate the Greeks for options using simulation is intuitively very simple, but sometimes simple ideas are complicated to implement. To illustrate this, we plot the price, Delta, and Gamma against the stock price for both American, with $J = 50$ exercise possibilities, and European style put options in Figure 1. The figure shows that as the stock price decreases and the option becomes deep in the money, the price essentially becomes linear in the underlying asset, the Delta approaches a value of -1, and the Gamma approaches a value of 0 very quickly for the American option and much faster than for the European option. A simple polynomial in the underlying asset cannot easily approximate this type of price function and estimating Delta and Gamma by its derivatives may be difficult.

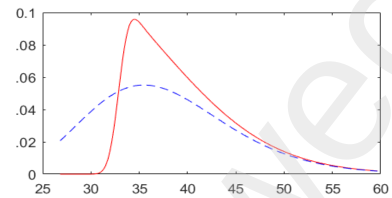
To demonstrate these issues, we first consider the results from a simple or "naive" method. For this method we observe a clear dependence of the variance and bias on the ISD size. Second, we proposed a 2-stage method which takes advantage of the control we have over the way the data is generated and addresses the issues observed with the "naive" method. Finally, our proposed



(a) Price



(b) Delta



(c) Gamma

Figure 1: Price, Delta, and Gamma across initial stock price

This figure plots the price, Delta, and Gamma across values of the stock price for an American, with $J = 50$ exercise possibilities, and a European style put option. The option strike prices are $K = 40$, the volatility is $\sigma = 20\%$, the interest rate is $r = 6\%$, and the maturity of the option is $T = 1$ year. The prices are obtained from a Binomial Model with 25 steps per trading day or 6,300 annual steps and the Delta and Gamma are obtained using simple one and two step ahead finite differences as described in, for example, Hull (2006).

2-stage method is benchmarked against some alternative simulation based methods demonstrating its superiority. Throughout this section we consider options with $T = 1$ year maturity, a volatility of $\sigma = 20\%$, and an interest rate of $r = 6\%$. The initial stock price is fixed at $S(0) = 40$ and the strike price is set equal to $K = 36$, 40, and 44, respectively, for the out of the money, OTM, at the money, ATM, and in the money, ITM, options.[7]

3.1 The naive method

The simplest possible and "naive" method is implemented using the following steps. First, create $N = 100,000$ initially dispersed stock prices X using Equation (21) and simulate a stock price path from each of these. Next, determine the pathwise payoffs Z from Equation (22) using the estimated optimal stopping time from the LSM algorithm obtained with a polynomial of order $M = 9$ in the cross-sectional regression using only ITM paths. Finally, regress Z on the distance of X to x_0 as in Equation (20) to approximate the price function. The estimated price and Greeks are found by scaling the coefficients from this regression appropriately. While Theorem 1 establishes convergence of the estimated prices and price sensitivities as the number of simulated paths, N, tends to infinity and the size of the ISD, α, shrinks to zero, Lemma 3 has important implications when it comes to implementing the methodology numerically. In particular, Lemma 3 provides insights for how to choose M_0, the order of the polynomial used in the regression at time $t = 0$, and on how the sample of N observations should be generated in terms of the chosen method for initially dispersing the

[7]In Appendix B.1 we consider a larger sample of options and show that the results presented here are robust.

state variables and the size of the ISD.

First, the theoretical results show that in order to estimate $P^{(i)}$, one needs to use a polynomial of at least order $M_0 \geq i$. More importantly though, the theoretical results show that increasing the order of the polynomial decreases the bias while it increases the variance. This means that the actual selection of M_0 essentially involves a compromise between reducing the bias and reducing the variance of the estimates. Note that Fan and Gijbels (1995a) show that increasing $M_0 - i$ from even to odd reduces the bias without affecting the variance, whereas increasing $M_0 - i$ from odd to even reduces the bias but increases the variance of the estimates. Thus, it is generally preferable to use $M_0 - i$ odd. In our application we are interested in $i = 0,\ 1,\ 2$, which are the price, Delta, and Gamma, respectively. As a compromise we select $M_0 \geq 3$ odd such that $M_0 - 0$ and $M_0 - 2$ are odd.[8] The specific order used in this section is $M_0 = 9$ but the robustness of this choice is examined further in Appendix B.2.

Second, note that the exact way in which the state variables are dispersed has potentially important implications on the properties of the estimators through the effect that this has on the implied density of the state variables at x_0, $f_X(x_0)$. In particular, the theoretical results show that the asymptotic variance is inversely proportional to $f_X(x_0)$. Thus, a concentrated distribution, as opposed to a uniform or bimodal distribution, should be preferred. Since the optimal kernel for the LPR regression is an Epanechnikov distribution we propose to generate the ISD using the following specification

$$\mathcal{K}_{ISD}(U) = 2 \times \sin\left(\operatorname{asin}(2 \times U - 1)/3\right), \tag{23}$$

where U is a vector of size N uniformly distributed on the unit interval. This results in an Epanechnikov distribution for the ISD, which is bounded, and indeed ensures that the distribution is symmetric around x_0 and peaked at this value.[9]

Finally, the theoretical results have important implications when choosing the size of the ISD, i.e., by how much to initially disperse the state variables around x_0. In particular, when we generate the initial stock prices as specified in Equation (21) and set $h = \max\left[X_{(N)} - x_0,\ x_0 - X_{(1)}\right]$, where $X_{(i)}$ denotes the ith order statistic such that all paths are used in the regression, then the choice

[8] Though $M_0 - i$ is in fact even for Delta and using $M_0 + 1$ would reduce the bias without increasing the variance, this avoids regressing multiple times at $t = 0$. Unreported results show that this is in fact inconsequential.

[9] We demonstrate that our method is robust towards several other ISD distribution types in Appendix B.3.

Table 3: Results with the naive method

		Price			Delta			Gamma		
K	α	BM	Estim	StDev	BM	Estim	StDev	BM	Estim	StDev
36	0.5	0.9166	0.9183	(0.0074)	-0.1979	-0.2039	(0.0540)	0.0381	0.0627	(0.5324)
40	0.5	2.3141	2.3156	(0.0147)	-0.4040	-0.4115	(0.0878)	0.0597	0.1693	(1.0262)
44	0.5	4.6535	4.6584	(0.0174)†	-0.6648	-0.6640	(0.1025)	0.0765	-0.0519	(1.1722)
36	5	0.9166	0.9173	(0.0097)	-0.1979	-0.1987	(0.0124)	0.0381	0.0390	(0.0133)
40	5	2.3141	2.3155	(0.0174)	-0.4040	-0.4031	(0.0182)	0.0597	0.0604	(0.0204)
44	5	4.6535	4.6576	(0.0203)	-0.6648	-0.6607	(0.0220)	0.0765	0.0742	(0.0235)
36	25	0.9166	0.8923	(0.0116)†	-0.1979	-0.1870	(0.0033)†	0.0381	0.0420	(0.0007)†
40	25	2.3141	2.2880	(0.0161)†	-0.4040	-0.4107	(0.0045)†	0.0597	0.0679	(0.0009)†
44	25	4.6535	4.7007	(0.0198)†	-0.6648	-0.6863	(0.0041)†	0.0765	0.0695	(0.0011)†

This table shows the estimated prices and Greeks for different values of α in Equation (21) with the naive method. We report averages of 100 independent simulations with $N = 100,000$ paths and standard errors in brackets. The optimal early exercise strategy is estimated using the LSM method using a polynomial of order $M = 9$ in the cross-sectional regressions. The options have $T = 1$ year to maturity, a volatility of $\sigma = 20\%$, an interest rate of $r = 6\%$, the initial stock price is fixed at $S(0) = 40$, and the strike prices are shown in the first column. The benchmark values, shown in the columns headed BM are estimated from the Binomial Model with 50,000 steps. The prices and Greeks are estimated using a polynomial of order $M_0 = 9$ at time $t = 0$.

† Indicates that the estimate is statistically different from the benchmark value at a 1% level.

of α is directly related to the bandwidth h. The theoretical results show that increasing the size of the ISD, i.e., increasing α or implicitly h, increases the bias while it decreases the variance. This means that picking the actual value of α involves a compromise between reducing the bias and reducing the variance of the estimates. To examine this further, we consider 3 different values of α in the following.

Table 3 shows the results for different choices of ISD determined by α. The table, first of all, shows that it is indeed possible to estimate prices and Greeks with this method. For example, when $\alpha = 5$ both prices and Greeks are generally very close to and insignificantly different from the benchmark values obtained with the Binomial Model. However, care has to be taken when choosing the initial α. For example, when $\alpha = 25$ prices and Greeks diverge and one clearly has to be careful choosing too large an ISD. Moreover, when $\alpha = 0.5$ higher order Greeks also diverge and the standard errors of all the estimated Greeks increase dramatically compared to when $\alpha = 5$. Thus, the table very nicely confirms the intuition from Theorem 1 that too small an initial spread leads to estimates with a large variance although they may be statistically unbiased, whereas too large an initial spread leads to estimates with a large bias but small variance and therefore results in statistically significant biases. This trade-off of bias against uncertainty makes it very difficult to provide general recommendations about the appropriate choice of the ISD.

3.2 A simple and consistent 2-stage method

The results above demonstrate that a first major shortcoming of the "naive" method is the large variance of estimates of higher order Greeks in particular obtained when using small ISDs. This is consistent with Equation (17) which shows that the variance of the estimates is directly proportional to the conditional variance of Z given $X = x$. Thus, finding a variance reduction technique that significantly reduces the variance of the Greek estimates is imperative when implementing this method. In Appendix A.2 we propose a simple variance reduction technique well adapted to the problem at hand. The method consist in discounting the cash flows from all simulated paths according to the approximated stopping time to t_1, estimate $\hat{F}_M^N\left(X\left(t_1\right)\right)$, exercise for paths for which it's optimal, and finally discount values to t_0. This "value function" method essentially amounts to using the Value Function as a smoother in the penultimate step which reduces $\sigma^2\left(x_0\right)$ significantly, and this turns out to be particularly effective for higher order Greeks.[10]

The "value function" method proposed in Appendix A.2 corrects one of the shortcomings of the "naive" method as it significantly decreases the standard deviation of the estimated values when using a small ISD. However, when the initial α is large, the estimates continue to be biased. To correct this a method that weighs the paths appropriately, and potentially truncates paths that are too far away from the stock price in question entirely, is needed. In other words, a way to determine the "optimal" α, α^*, used in Equation (21) is needed. In Appendix A.3 we propose a simple algorithm for obtaining the optimal bandwidth to be used. Once α^* is obtained we can simply truncate any paths that were initialized with values of the state variables outside this interval and the estimates are obtained by performing a new regression at time $t = 0$ using only the paths within this optimal ISD. The "truncated" method uses the original ISD but discards data points outside the optimal bandwidth in the final regression, and the major drawback of the method is that it leads to increased variability in the estimated quantities.

To fix the issue with the "truncated" method, we propose to, instead of simply truncating paths outside of the optimal α entirely, rescale or resimulate such that all the paths lie within the optimal ISD in a second stage. Table 4 shows the results for the different choices of α using our proposed "2-stage" method. The table, first of all, shows that this method indeed performs very

[10] Other variance reduction methods could be used with or instead of this method. In particular, using control variates and controlled continuation values as in Rasmussen (2005) is interesting. We leave this for future research.

Table 4: Results with the 2-stage method that rescales all paths to be within the optimal α

		Price			Delta			Gamma		
K	α	BM	Estim	StDev	BM	Estim	StDev	BM	Estim	StDev
36	0.5	0.9166	0.9181	(0.0061)	-0.1979	-0.1975	(0.0077)	0.0381	0.0303	(0.0297)†
40	0.5	2.3141	2.3164	(0.0098)	-0.4040	-0.4043	(0.0127)	0.0597	0.0469	(0.0598)
44	0.5	4.6535	4.6564	(0.0109)†	-0.6648	-0.6639	(0.0201)	0.0765	0.0576	(0.0977)
36	5	0.9166	0.9170	(0.0068)	-0.1979	-0.1978	(0.0046)	0.0381	0.0385	(0.0058)
40	5	2.3141	2.3149	(0.0102)	-0.4040	-0.4038	(0.0074)	0.0597	0.0602	(0.0097)
44	5	4.6535	4.6553	(0.0124)	-0.6648	-0.6626	(0.0102)	0.0765	0.0768	(0.0129)
36	25	0.9166	0.9156	(0.0071)	-0.1979	-0.1982	(0.0048)	0.0381	0.0390	(0.0043)
40	25	2.3141	2.3130	(0.0122)	-0.4040	-0.4041	(0.0077)	0.0597	0.0603	(0.0065)
44	25	4.6535	4.6541	(0.0134)	-0.6648	-0.6632	(0.0091)	0.0765	0.0769	(0.0088)

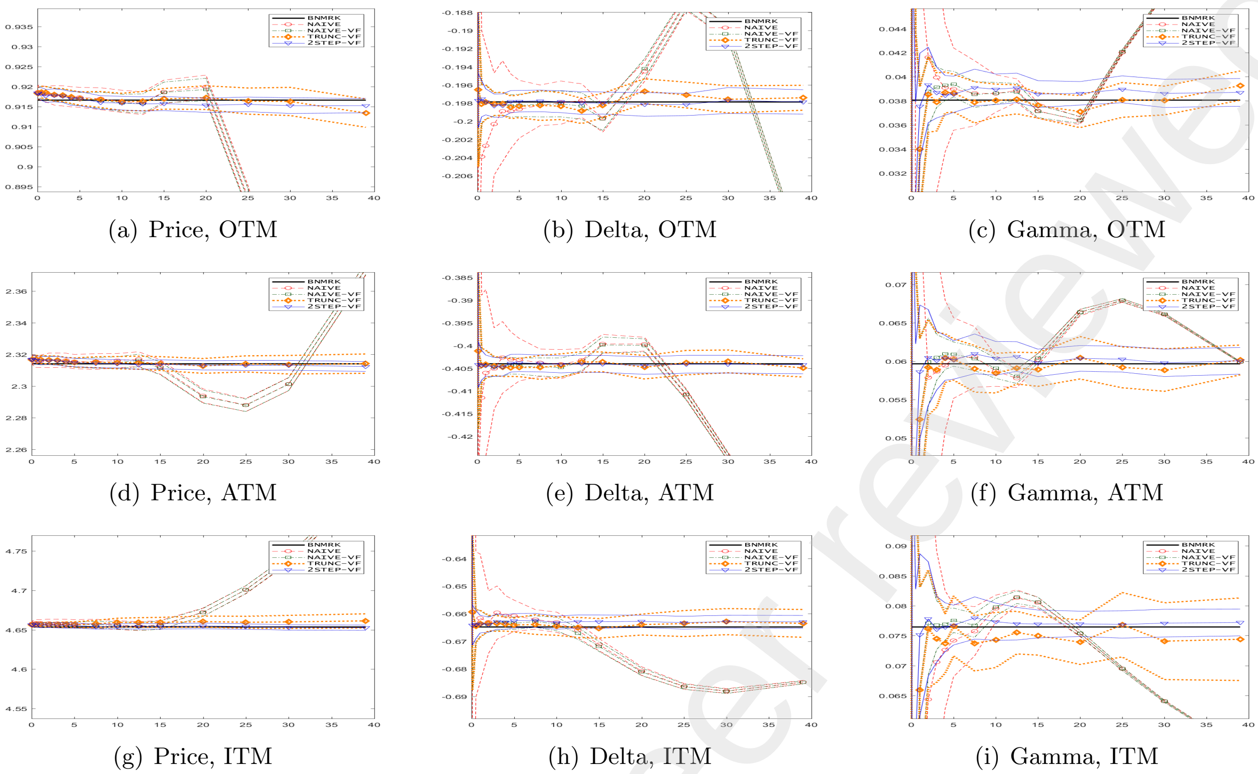
This table shows the estimated prices and Greeks for different values of α in Equation (21) with the 2-stage method that, in addition to using a value function iteration at the second to last step, rescales all paths to be within the optimal α and uses these to estimate prices and Greeks in a second step. See also the notes to Table 3.

† Indicates that the estimate is statistically different from the benchmark value at a 1% level.

well. In particular, the 2-stage method produces results for higher order Greeks when α is small that have smaller standard deviation than with any of the other methods, and the estimates are in fact also generally less biased. Moreover, when α is large, the bias is significantly improved for all estimates and though the standard deviation is slightly higher than with, e.g., the value function method it remains very low even for higher order Greeks. Across the choices of initial values of α the table shows that the 2-stage method nicely balances off the bias from using a large ISD against the uncertainty stemming from using a small ISD. If anything, the recommendation from the table in terms of choosing the initial value of α is that this should be chosen large since the 2-stage procedure will correct for choosing it too extreme and still produce unbiased estimates with reasonable standard deviation.

Based on this, our proposed 2-stage method for estimating option prices and Greeks is:

1. Create a sample of size N of initially dispersed stock prices X with a given α in Equation (21) and simulate a stock price path from each of these. Calculate the pathwise payoffs Z using the estimated optimal stopping time determined with the LSM method in Equation (6).
2. Use the value function estimated at $t = 1$ to "smooth" the payoffs and discount these back to time $t = 0$ as in Appendix A.2.
3. Use the optimal bandwidth selector from Appendix A.3 to determine an α^* based on the above smoothed Z and X.
4. Rescale the initially dispersed state variables to the optimal bandwith, apply the estimated



(a) Price, OTM (b) Delta, OTM (c) Gamma, OTM

(d) Price, ATM (e) Delta, ATM (f) Gamma, ATM

(g) Price, ITM (h) Delta, ITM (i) Gamma, ITM

Figure 2: Results with estimated early exercise strategy

This figure plots estimated prices and Greeks along with 99% confidence intervals across different values of α in Equation (21) for the three options considered in this section. Results are based on 100 independent simulations with $N = 100,000$ paths. Prices and Greeks are estimated using a polynomial of order $M_0 = 9$ at time $t = 0$. The optimal exercise strategy is estimated using a polynomial of order $M = 9$.

optimal stopping time to the paths, use the value function method at $t = 1$ to smooth the pathwise discounted payoffs, and estimate the Greeks using the new paths.

In the spirit that a picture, or a figure, is worth a thousand words, or numbers, Figure 2 shows the results for the three options considered across various choices of initial ISD size. The figure clearly shows that as the size of the initial spread increases the naive method, labeled "NAIVE" in the plots, and even the method that uses the value function method, labeled "NAIVE-VF" in the plots, breaks down, and only methods that rely on results with optimal bandwidth selection, labeled "TRUNC-VF" and "2STEP-VF", can consistently provide unbiased estimates of the price, Delta, and Gamma. As such, the plots are in line with the theoretical results from Section 2 and confirm that the variance of the estimates decreases with increasing ISD whereas the size of the bias increases with increasing ISD. However, and this was not evident from the tables presented previously, the biases of the naive method do not increase monotonically making it difficult to provide any heuristic arguments for how to pick the size of initial α for these methods. Contrary to this, the plots show almost straight horizontal lines for the 2-stage method indicating that the choice of initial α is essentially inconsequential for the performance of the method, demonstrating that our proposed method works well irrespective of the initial α used.

3.3 Comparison with some alternative methods

In the Introduction we mention other possible methods that could be used to obtain prices and Greeks for options using Monte Carlo simulation and we now compare our results to some of these. First, we consider the pathwise derivative method (PDM), in which the option payoff is differentiated directly. This method produces unbiased and very precise estimators when they exist but it cannot be used to estimate the Gamma in our setup. Second, we consider the likelihood ratio method (LRM), in which the probability density of the underlying price is instead differentiated. This method also produces unbiased estimators though they may have large variances, in particular when compared to the PDM. For completeness, we also report the results from the Modified Least Squares Method (MLSM) of Wang and Caflisch (2010). This method essentially corresponds to our naive method though with an alternative random and non-symmetric ISD kernel and a choice of α that depends on the volatility and the maturity of the option but is in no way optimized, see their Equation (13). The results for these three methods are compared to our 2-stage method in Table 5.[11]

Table 5 shows that the PDM and LRM methods also produce estimated prices and Greeks that are very close to the benchmark values when they exist, but the estimates from the MLSM method of Wang and Caflisch (2010) have larger biases, except for the out of the money option. In terms of the estimated Greeks, our proposed 2-stage method always has the smallest bias. The biggest differences, however, are found in terms of the standard deviation of the estimates. As expected, the PDM offers the most precise estimates of the Delta, though these estimates are actually statistically different from the true value, and compared to this the LRM has a much larger standard deviation. The standard deviations of the estimated Deltas with our proposed 2-stage method are smaller than what the LRM offers though still somewhat larger than those obtained with the PDM. When considering the Gamma, which can only be estimated well with the

[11] Results for other combinations of volatility and maturity are available upon request.

Table 5: Comparison with alternative models

K	Model	Price BM	Price Estim	Price StDev	Delta BM	Delta Estim	Delta StDev	Gamma BM	Gamma Estim	Gamma StDev
36	PDM	0.9166	0.9180	(0.0056)	-0.1979	-0.1973	(0.0014)†	0.0381		
40	PDM	2.3141	2.3149	(0.0082)	-0.4040	-0.4017	(0.0022)†	0.0597		
44	PDM	4.6535	4.6545	(0.0096)	-0.6648	-0.6607	(0.0038)†	0.0765		
36	LRM	0.9166	0.9180	(0.0056)	-0.1979	-0.1990	(0.0055)	0.0381	0.0393	(0.0079)
40	LRM	2.3141	2.3149	(0.0082)	-0.4040	-0.4049	(0.0094)	0.0597	0.0610	(0.0147)
44	LRM	4.6535	4.6545	(0.0096)	-0.6648	-0.6676	(0.0159)	0.0765	0.0780	(0.0212)
36	MLSM	0.9166	0.9150	(0.0083)	-0.1979	-0.1991	(0.0026)†	0.0381	0.0389	(0.0014)†
40	MLSM	2.3141	2.3058	(0.0123)†	-0.4040	-0.4084	(0.0031)†	0.0597	0.0632	(0.0014)†
44	MLSM	4.6535	4.6621	(0.0123)†	-0.6648	-0.6755	(0.0038)†	0.0765	0.0746	(0.0014)†
36	2-stage	0.9166	0.9159	(0.0067)	-0.1979	-0.1978	(0.0047)	0.0381	0.0390	(0.0049)
40	2-stage	2.3141	2.3141	(0.0102)	-0.4040	-0.4041	(0.0073)	0.0597	0.0604	(0.0085)
44	2-stage	4.6535	4.6544	(0.0123)	-0.6648	-0.6632	(0.0093)	0.0765	0.0773	(0.0120)

This table shows the estimated prices and Greeks with different methods. We report averages and standard deviations in brackets from 100 independent simulations with $N = 100,000$ paths in each simulation using a polynomial of order $M = 5$ to determine the optimal early exercise strategy, except in the 2-stage method where $M = 9$. Results in the rows headed PDM correspond to the Pathwise Derivative Method in which the option payoff is differentiated and for which only estimates of Delta are available. Results in the rows headed LRM correspond to the Likelihood Ratio Method in which the density is differentiated. Results in the rows headed MLSM correspond to the Modified Least Squares Method with the particular ISD from Equation (13) in Wang and Caflisch (2010) with a polynomial of order $M_0 = 5$ to estimate the price and Greeks. Results in the rows headed 2-stage correspond to our proposed method implemented with an initial $\alpha = 10$ with a polynomial of order $M_0 = 9$ to estimate the price and Greeks.
† Indicates that the estimate is statistically different from the benchmark value at a 1% level.

LRM and our 2-stage method, the table shows that the relative precision of our proposed method increases and for these higher order sensitivities the standard deviation is roughly half that of the LRM. Thus, among the methods that provide estimates of the prices and Greeks which are statistically insignificant from the benchmark values our proposed method is the most precise. Since the computational complexity of all these methods is roughly equivalent and our proposed method is more flexible and generally offers more precise estimates of the Greeks, this clearly demonstrate the value of our suggested method.

4 Time-varying volatility application

The constant volatility BSM model considered in Section 3 is, though widely used, unrealistic and suffers from a number of shortcomings when it comes to fitting financial asset dynamics empirically. An alternative class of models that allow more flexibility are those within the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) framework. The class of GARCH models was developed by Engle (1993) and Bollerslev (1986) and allows for time varying volatility which is filtered out from historical returns alone and therefore estimation is possible using standard Maximum Likelihood techniques. Duan (1995) was among the first to show how to price options in a (Gaussian) GARCH model and this framework has since been widely used empirically. See Christoffersen, Jacobs, and Ornthanalai (2013) for a detailed survey of the use of GARCH option pricing models.

In this section we apply our method for simulating the Greeks in this more realistic setting. First we demonstrate that our method indeed generalizes by benchmarking it against a specific GARCH type model that allows for closed form prices and Greeks for European options. Next we compare the Greeks obtained for American options with alternative models that fit empirical data better. Finally, we analyse the estimated Greeks in periods of financial crisis. GARCH models are typically specified at a daily frequency. In an effort to stay consistent with the implementation above and to decrease the computational burden when pricing American options, we assume that one year has $T = 250$ trading days and consider exercise only every 5'th day leading to a total of $J = 50$ possible early exercise points per year as was used with the constant volatility BSM model.

4.1 Generalization to models with time-varying volatility

To demonstrate that our method indeed generalizes, we first consider the case of European options in the Affine GARCH specification proposed in Heston and Nandi (2000) (HN-GARCH), for which closed form option valuation and calculation of the Greeks is possible. In the HN-GARCH model the dynamics used for pricing are given by

$$R_t = r - \tfrac{1}{2}h_t + \sqrt{h_t}z_t, \tag{24}$$

where $z_t \mid \mathcal{F}_{t-1} \sim N(0,1)$, with $\mathcal{F}_t$ denoting the information set at time t, and where the conditional variance, h_t, follows the HN-GARCH process given by

$$h_t = \omega + \beta h_{t-1} + \alpha\left(z_{t-1} - \gamma\sqrt{h_{t-1}}\right)^2, \tag{25}$$

where ω, β, α, and γ are parameters governing the variance dynamics yielding a volatility persistency of $\Psi = \beta + \alpha\gamma^2$ and a long run unconditional variance of $\tilde{h} = (\omega + \alpha)/(1 - \Psi)$.

Table 6: Estimated Greeks for European options under the Heston-Nandi model

Panel A: Benchmark values												
K	Price		Delta		Gamma		Vega		Vomma		Vanna	
36	0.9835		-0.1595		0.0262		2.0155		9.1524		0.1566	
40	2.0901		-0.3031		0.0421		2.8337		11.6570		0.0887	
44	3.8611		-0.4885		0.0540		3.1562		12.8214		-0.0892	
Panel B: Using $N = 317^2 = 100,489$ paths with a double ISD												
	Price		Delta		Gamma		Vega		Vomma		Vanna	
K	Estim	StDev	Estim	StDev	Estim	StDev	Estim	StDev	Estim	StDev	Estim	StDev
36	0.9866	(0.0367)	-0.1613	(0.0157)	0.0267	(0.0107)	2.0502	(0.2128)	9.0004	(2.4126)	0.1713	(0.1033)
40	2.1009	(0.0513)	-0.3036	(0.0202)	0.0420	(0.0131)	2.8658	(0.2913)	10.4755†	(3.1597)	0.0988	(0.1232)
44	3.8770	(0.0642)	-0.4871	(0.0261)	0.0534	(0.0161)	3.1933	(0.3870)	11.2441†	(3.8788)	-0.0606	(0.1535)
Panel C: Using $N = 500^2 = 250,000$ paths with a double ISD												
	Price		Delta		Gamma		Vega		Vomma		Vanna	
K	Estim	StDev	Estim	StDev	Estim	StDev	Estim	StDev	Estim	StDev	Estim	StDev
36	0.9803	(0.0231)	-0.1610	(0.0089)	0.0272	(0.0065)	2.0399	(0.1007)	9.3254	(1.8341)	0.1617	(0.0594)
40	2.0897	(0.0311)	-0.3056	(0.0117)	0.0432	(0.0086)	2.8382	(0.1341)	11.3151	(2.3420)	0.0934	(0.0777)
44	3.8611	(0.0378)	-0.4897	(0.0164)	0.0545	(0.0109)	3.1491	(0.1799)	12.3592	(2.8201)	-0.0772	(0.0957)
Panel D: Using $N = 1000^2 = 1,000,000$ paths with a double ISD												
	Price		Delta		Gamma		Vega		Vomma		Vanna	
K	Estim	StDev	Estim	StDev	Estim	StDev	Estim	StDev	Estim	StDev	Estim	StDev
36	0.9829	(0.0101)	-0.1597	(0.0050)	0.0264	(0.0032)	2.0336	(0.0600)†	9.1701	(0.8086)	0.1595	(0.0248)
40	2.0903	(0.0141)	-0.3042	(0.0070)	0.0427	(0.0043)	2.8420	(0.0872)	11.4277	(1.0748)	0.0910	(0.0324)
44	3.8612	(0.0180)	-0.4897	(0.0085)	0.0548	(0.0055)	3.1596	(0.1087)	12.5257	(1.4066)	-0.0817	(0.0409)

This table shows the estimated pries and Greeks for European options under the HN-GARCH model in Equations (24) - (25) with $\beta = 0.86$, $\alpha = 0.000003$, $\gamma = 200$, yielding a persistency of $\Psi = 0.98$ and where ω is set to ensure an unconditional volatility of 20%. The Greeks are obtained with our 2-stage method implemented with a bivariate ISD with different numbers of paths and using a regression of order $M_0^S = 9$ and $M_0^V = 5$ at time $t = 0$. All results are based on $I = 100$ independent repetitions.

† Indicates that the estimate is statistically different from the benchmark value at a 1% level.

Table 6 shows the results for European option prices and several Greeks and compares to the closed-form benchmark results. In Panel A of this table we report the benchmark values obtained with the available closed form solutions in the literature. In Panels B through D we report the estimates obtained with increasing sample sizes when using our proposed 2-stage method generalized to the bivariate case. Specifically, here we use a bivariate ISD and a regression on the complete set of polynomials of order $M_0^S = 9$ and $M_0^V = 5$ in the stock and volatility dimension, respectively, at time $t = 0$.[12] First, Panel B of the table uses a sample size close to that used with univariate ISDs and clearly demonstrates that our method generalizes to the bivariate case. In particular, all the estimated Greeks, in the stock dimension, in the volatility dimension as well as in the cross section, are close to the true values. In fact, with the exception of the ATM and ITM Vommas the estimates are statistically insignificantly different from the benchmark values.

Compared to the estimated Greeks for the constant volatility model shown in the bottom rows in Table 4, we observe that the standard deviations reported are larger for the estimates from the HN-GARCH model in Panel B of Table 6. This is somewhat expected since we now consider a bivariate model. To examine this further, Panel C reports results with $N = 500^2 = 250,000$ paths. In this case all the standard deviations are lowered significantly as are most of the biases. In particular, when using $N = 250,000$ paths the biases of the ATM and ITM Vommas are reduced to about a third of their sizes in Panel B. As a result, all the Greeks are now very precisely estimated and insignificantly different from the benchmark values. Panel D of the table, which reports results with $N = 1,000^2 = 1,000,000$ paths, shows that further improvements can be made. However, in this case the ATM Vega is now significantly different from the benchmark, indicating that, perhaps, higher order polynomials are needed in the cross sectional regressions when using this many paths. In the following we therefore use $N = 250,000$ paths in our applications.

4.2 Estimated Greeks with alternative models

While the affine HN-GARCH model benefits from allowing closed form solutions for European option prices, this is not the case for the corresponding American options. Since the majority of the options traded empirically are American style, this is a major concern. Moreover, it is well

[12] We use a "square" ISD with the same number of grid points in both dimensions. We ensure that initial volatilities are positive by truncating the vector U used in Equation (23) from below. The optimal bandwidth in each dimension is selected using the procedure in Appendix A.3 and is chosen with respect to second order derivatives.

known in the empirical option pricing literature that the analytical convenience of affine option valuation models comes at a price since the errors for simple non-affine models may be much lower (see, e.g, Christoffersen, Dorion, Jacobs, and Wang (2010) for results in the univariate case and Rombouts, Stentoft, and Violante (2020) for a generalization of these to the multivariate setting). Because of these drawbacks with the HN-GARCH model, we now examine the Greeks obtained with alternative models with other and empirically more realistic specifications for the time-varying volatility.

As the most obvious and simplest possible alternative we consider the asymmetric (non-affine) NGARCH model of Engle and Ng (1982) which can be used for option pricing with the approach of Duan (1995) for obtaining the risk-neutral dynamics. In this option pricing model, returns under the pricing measure $\mathbb{Q}$ are given by

$$R_t = r - \tfrac{1}{2}h_t + \sqrt{h_t}z_t, \tag{26}$$

where $z_t \mid \mathcal{F}_{t-1} \sim N(0,1)$, with $\mathcal{F}_t$ denoting the information set at time t, and where the conditional variance, h_t, follows the asymmetric NGARCH process given by

$$h_t = \omega + \beta h_{t-1} + \alpha h_{t-1}\left(z_{t-1} - \gamma\right)^2, \tag{27}$$

where ω, β, α, and γ are parameters governing the variance dynamics yielding a volatility persistency of $\Psi = \beta + \alpha\left(1+\gamma^2\right)$ and a long run unconditional variance of $\tilde{h} = \omega/\left(1-\Psi\right)$. We will refer to this model simply as the NGARCH model, and we note that a symmetric GARCH model is obtained when $\gamma = 0$ and that the Constant Volatility (CV) model from above amounts to setting all parameters except ω equal to zero. The LSM method was first used to price American style options with GARCH and NGARCH models by Stentoft (2005).[13]

Table 7 shows estimated prices and Greeks for American options in the HN-GARCH, NGARCH and GARCH models in Panels B, C, and D, respectively. For comparison we also report the corresponding estimates for European options in the HN-GARCH model in Panel A. When American

[13]We note that since the conditional variance or volatility becomes a state variable in these models, the estimation of $\hat{F}_M^N\left(X\left(t_j\right)\right)$ for the approximation of the optimal stopping time requires a bivariate polynomial. In this paper, we chose the complete set of polynomials of order $M^S = 9$ in the stock and $M^V = 5$ in the volatility, respectively.

Table 7: Prices and Greeks with alternative models

	Price		Delta		Gamma		Vega		Vomma		Vanna	
Panel A: European options with HN-GARCH												
K	Estim	StDev	Estim	StDev	Estim	StDev	Estim	StDev	Estim	StDev	Estim	StDev
36	0.9803	(0.0231)	-0.1610	(0.0089)	0.0272	(0.0065)	2.0399	(0.1007)	9.3254	(1.8341)	0.1617	(0.0594)
40	2.0897	(0.0311)	-0.3056	(0.0117)	0.0432	(0.0086)	2.8382	(0.1341)	11.3151	(2.3420)	0.0934	(0.0777)
44	3.8611	(0.0378)	-0.4897	(0.0164)	0.0545	(0.0109)	3.1491	(0.1799)	12.3592	(2.8201)	-0.0772	(0.0957)
Panel B: American options with HN-GARCH												
K	Estim	StDev	Estim	StDev	Estim	StDev	Estim	StDev	Estim	StDev	Estim	StDev
36	1.0785	(0.0216)	-0.1835	(0.0086)	0.0315	(0.0057)	2.3964	(0.1015)	10.4937	(1.6148)	0.2040	(0.0521)
40	2.3622	(0.0255)	-0.3604	(0.0107)	0.0541	(0.0075)	3.4627	(0.1323)	11.4758	(1.9879)	0.1591	(0.0618)
44	4.5308	(0.0346)	-0.6442	(0.0136)	0.0951	(0.0104)	3.5969	(0.1790)	12.1549	(2.5269)	-0.4406	(0.0731)
Panel C: American options with NGARCH												
K	Estim	StDev	Estim	StDev	Estim	StDev	Estim	StDev	Estim	StDev	Estim	StDev
36	1.0297	(0.0248)	-0.1796	(0.0114)	0.0328	(0.0061)	2.4226	(0.0995)	10.0442	(1.8887)	0.1730	(0.0550)
40	2.3142	(0.0293)	-0.3669	(0.0137)	0.0561	(0.0071)	3.3145	(0.1548)	10.1267	(2.0101)	0.0928	(0.0596)
44	4.5209	(0.0388)	-0.6561	(0.0163)	0.0943	(0.0099)	3.2522	(0.1901)	9.8574	(2.5657)	-0.4550	(0.0765)
Panel D: American options with GARCH												
K	Estim	StDev	Estim	StDev	Estim	StDev	Estim	StDev	Estim	StDev	Estim	StDev
36	0.8809	(0.0467)	-0.1877	(0.0137)	0.0373	(0.0054)	2.2859	(0.1013)	10.2664	(1.4215)	0.2380	(0.0546)
40	2.2186	(0.1001)	-0.3915	(0.0143)	0.0605	(0.0089)	3.4540	(0.2126)	10.4683	(1.8093)	0.1639	(0.0827)
44	4.5344	(0.0832)	-0.6835	(0.0253)	0.0889	(0.0110)	3.6529	(0.2075)	10.0787	(2.2078)	-0.3944	(0.1070)

This table shows the estimated prices and Greeks for various alternative models. Panel A and B are based on the HN-GARCH model in Equations (24) - (25) with $\beta = 0.86$, $\alpha = 0.000003$, $\gamma = 200$, yielding a persistency of $\Psi = 0.98$ and where ω is set to ensure an unconditional annualized volatility of 20%, and report results for European and American prices, respectively. Panel C is based on the NGARCH model in Equations (26) - (27) with $\beta = 0.90$, $\alpha = 0.04$, $\gamma = 1.00$, yielding a persistency of $\Psi = 0.98$ and where ω is set to ensure an unconditional annualized volatility of 20%. Panel D uses the same model but with $\alpha = 0.08$, $\gamma = 0.00$ yielding again a persistency of $\Psi = 0.98$ and in both these panels results are reported for American options. In all cases, the Price and all the Greeks are obtained with our 2-stage method implemented with $N = 250,000$ paths and using a regression of order $M_0^S = 9$ in the stock dimension and $M_0^V = 5$ at time $t = 0$. The optimal stopping time, when needed, is estimated with $M_\tau^S = 9$ in the stock dimension and $M_\tau^V = 5$.

options are considered we no longer have a closed form benchmark with which to compare even in the HN-GARCH model. However, when comparing the estimated prices and Greeks for American options in Panel B with the corresponding European case in Panel A the results are indeed very meaningful. For example, the price is higher, which it should be by definition, and the estimated Delta is also larger in absolute value. Also, and this is consistent with, e.g., Figure 1 for the CV case, the estimated Gamma is significantly larger for ITM American options. More generally, the panels show that almost all the Greeks are larger for American options than for European options.

Panel C of Table 7 shows estimated prices and Greeks for American options in the, arguably, empirically more realistic non-affine NGARCH model. Compared to Panel B, these results show that estimated Greeks may be very different with the non-affine NGARCH model than with the affine HN-GARCH model. Moreover, while the NGARCH prices are all slightly lower, up to 4.73% for the OTM option, with the exception of Vomma the differences in the Greeks change non-monotonically across moneyness. For example, the Delta is around 2% larger in absolute value with the HN-GARCH model for OTM options but around 2% smaller for ITM options. The relative difference in the Greeks may be as large as 71.56%, which occurs for the estimated Vanna for the ATM option.

Finally, compared with the results for the symmetric GARCH model in Panel D in Table 7, our results show that asymmetries are important and neglecting them leads to significantly different estimates of option Greeks. The largest differences are found for Vanna, which for the ATM option differ by as much as 43.42%, but even a first order sensitivity like Delta is estimated to be at least 4% smaller when asymmetries are present. It should also be noted that while Delta and Vomma are lower with an NGARCH model for all moneyness categories, the other estimated Greeks again change non-monotonically across moneyness and whereas the largest differences in Gamma are found for OTM options it is the ITM Vega that differs the most among the tree moneyness categories.

4.3 Estimated Greeks in periods of crisis

One of the main benefits of the NGARCH framework is that it can accommodate time-varying conditional volatility, i.e., the level of the (spot) volatility changes with current market conditions. For example, given a set of parameter values observed returns can be used to filter out a time series of conditional volatility and Equation (27) illustrates that a large shock to returns today will increase the conditional volatility tomorrow significantly. An alternative way to obtain a time series of conditional volatilities is to use the corresponding option implied volatilities, e.g, the VIX backed out from the market prices of options written on the S&P 500 Index. To understand how this can be done, we note that in the absence of jumps it is possible to derive a model implied VIX in this class of discrete time models. In the NGARCH model we consider, this is given by

$$\frac{1}{\tau}\left(\frac{VIX_t}{100}\right)^2 = h_{t+1}\frac{1-\Psi^T}{(1-\Psi)\,T} + \tilde{h}\left(1-\frac{1-\Psi^T}{(1-\Psi)\,T}\right), \tag{28}$$

where h_{t+1} is the conditional variance, $\tilde{h}$ is the unconditional (long run) variance, Ψ is the implied persistency of the model, and τ and T are respectively the number of days per year and month (see, e.g, Kanniainen, Lin, and Yang (2014)).[14] Put differently, this means that we can back out the conditional volatilities from the VIX from

$$h_t = \frac{(1-\Psi)\,T}{1-\Psi^T}\left[\frac{1}{\tau}\left(\frac{VIX_{t-1}}{100}\right)^2 - \tilde{h}\left(1-\frac{1-\Psi^T}{(1-\Psi)\,T}\right)\right], \tag{29}$$

which yields a time series of conditional variances, which is consistent with observed option market prices through time as summarized by the VIX time series.

To illustrate this, we plot the time series of conditional annualized volatilities for the NGARCH model filtered out from the VIX for the period from 2000 to 2021 in Figure 3. The plot shows that the implied conditional volatility follows the VIX very closely and varies significantly through time with large spikes at the time of the Global Financial Crisis in 2008 and with the onset of the COVID pandemic in late 2019 and early 2020. The average level of the annualized conditional volatility for this sample is 19.51%, very close to the unconditional level assumed for the NGARCH model above, with a minimum value of 3.95% and a maximum value of 90.67%. The distribution of conditional volatilities is highly skewed with a median of only 17.25% and when rounded to integer values the mode is a mere 11%. Given these characteristics for the time series of volatility, it is obvious that assuming volatility to be constant may lead to very poor and unrealistic estimates of option prices in general and option Greeks in particular.

[14] Consistent with market standard, we take a year to be $\tau = 252$ (trading) days and a month to be $T = 21$ days.

Table 8: Prices and Greeks in period of crisis

	Price		Delta		Gamma		Vega		Vomma		Vanna	
Panel A: NGARCH model started at 10% annualized volatility												
K	Estim	StDev	Estim	StDev	Estim	StDev	Estim	StDev	Estim	StDev	Estim	StDev
36	0.8433	(0.0174)	-0.1639	(0.0114)	0.0328	(0.0069)	1.2519	(0.2937)	13.2777	(3.5179)	0.1524	(0.1040)
40	2.0392	(0.0332)	-0.3546	(0.0147)	0.0595	(0.0103)	2.0940	(0.3687)	14.1057	(4.1673)	0.1623	(0.1225)
44	4.2480	(0.0420)	-0.7037	(0.0183)	0.1181	(0.0140)	2.1083	(0.4468)	13.2122	(4.8855)	-0.5200	(0.1513)
Panel B: NGARCH model started at 40% annualized volatility												
K	Estim	StDev	Estim	StDev	Estim	StDev	Estim	StDev	Estim	StDev	Estim	StDev
36	1.6828	(0.0184)	-0.2135	(0.0095)	0.0300	(0.0057)	3.9202	(0.1022)	5.4764	(0.3986)	0.1463	(0.0385)
40	3.1423	(0.0242)	-0.3748	(0.0109)	0.0458	(0.0067)	4.7521	(0.1368)	4.9374	(0.5218)	0.0016	(0.0451)
44	5.3346	(0.0308)	-0.5781	(0.0119)	0.0623	(0.0076)	4.6933	(0.1607)	5.2085	(0.6751)	-0.3365	(0.0493)
Panel C: Constant Volatility model with 10% annualized volatility												
K	Estim	StDev	Estim	StDev	Estim	StDev	Estim	StDev	Estim	StDev	Estim	StDev
36	0.0949	(0.0386)	-0.0616	(0.0174)	0.0182	(0.0132)	5.5859	(0.4641)	63.5696	(8.2411)	1.5656	(0.2038)
40	0.8872	(0.0323)	-0.4305	(0.0148)	0.1678	(0.0193)	11.7855	(0.4364)	40.8173	(9.6314)	-0.4459	(0.2051)
44	3.8077	(0.0479)	-0.8845	(0.0196)	0.0450	(0.0189)	4.0366	(0.6819)	100.7200	(13.3851)	-2.4406	(0.2501)
Panel D: Constant Volatility model with 40% annualized volatility												
K	Estim	StDev	Estim	StDev	Estim	StDev	Estim	StDev	Estim	StDev	Estim	StDev
36	3.3994	(0.0277)	-0.2880	(0.0116)	0.0241	(0.0079)	13.6204	(0.1816)	6.6368	(1.3726)	0.1972	(0.0743)
40	5.2728	(0.0309)	-0.3812	(0.0131)	0.0213	(0.0093)	15.4881	(0.2156)	-0.2359	(1.5792)	-0.0631	(0.0732)
44	7.5069	(0.0383)	-0.4899	(0.0158)	0.0342	(0.0104)	15.9302	(0.2676)	7.7648	(1.8381)	-0.5428	(0.0921)

This table shows the estimated prices and Greeks in different volatility regimes. Panel A and B are based on the NGARCH model in Equations (26) - (27) with $\beta = 0.90$, $\alpha = 0.04$, $\gamma = 1.00$, yielding a persistency of $\Psi = 0.98$ and where ω is set to ensure an unconditional annualized volatility of 20% but with different initial values for h_t. Panels C and D are based on the Constant Volatility model in which the long run unconditional volatility is scaled directly. In all cases, the Price and all the Greeks are obtained with our 2-stage method implemented with $N = 250,000$ paths and using a regression of order $M_0^S = 9$ in the stock dimension and $M_0^V = 5$ at time $t = 0$. The optimal stopping time is estimated with $M_\tau^S = 9$ in the stock dimension and $M_\tau^V = 5$ in the NGARCH model an with $M_\tau^S = 9$ in the stock dimension only in the Constant Volatility model.

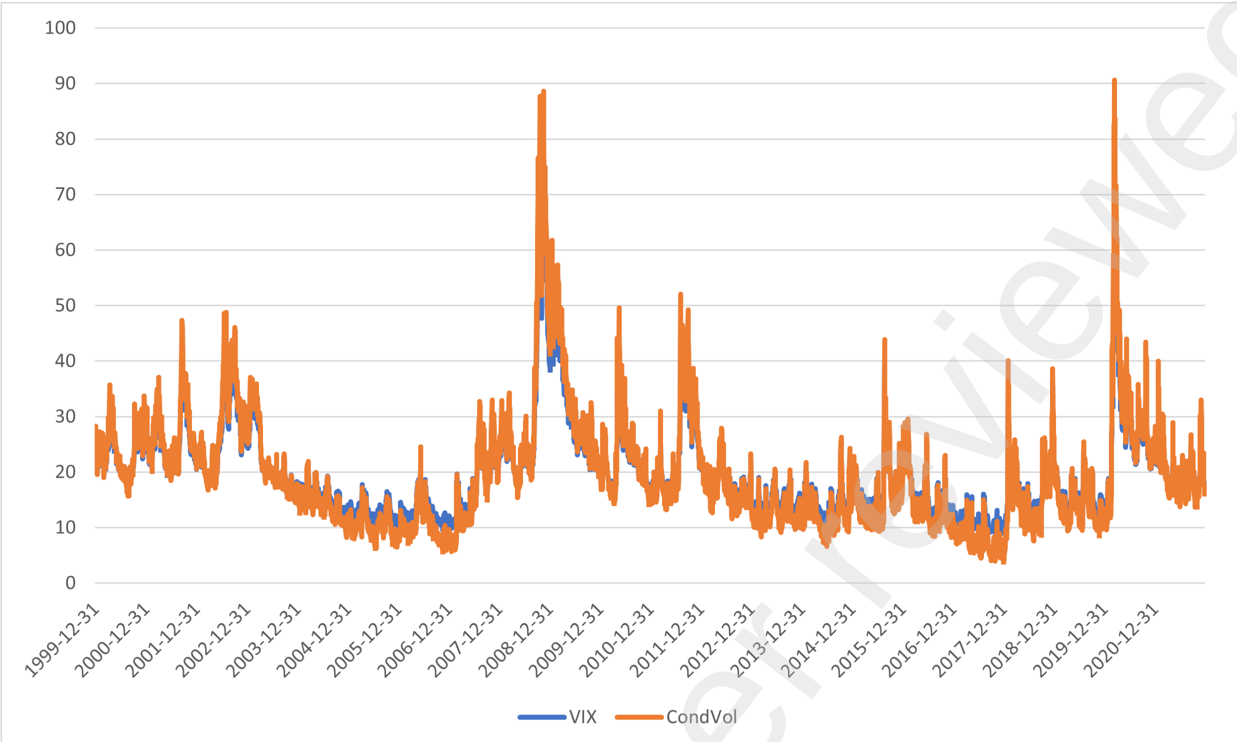


Figure 3: VIX and implied conditional volatility

This figure plots the time series of the VIX and the conditional volatility implied by the NGARCH model from Equation (28). We use the same parameter values as above and set $\beta = 0.90$, $\alpha = 0.04$, $\gamma = 1.00$, which yields a persistency of $\Psi = 0.98$, and where ω is set to ensure an unconditional annualized volatility of 20%.

To analyze this further, we first report the estimated prices and Greeks for the NGARCH model with different values for the (spot) conditional volatility in Table 8. Panel A shows results for the case with a low initial conditional volatility set at 10%, close to the mode, and Panel B shows results for the case with a high initial conditional volatility set at 40% corresponding to the 95.6'th percentile. Comparing these two panels to the results in Panel C of Table 7 shows that the level of the (spot) volatility significantly impacts not only the estimated price but much more so the estimated Greeks. In particular, comparing the two panels it is seen that estimated Vegas are particularly sensitive to the volatility regime and can be more than twice as large in a high volatility regime than in a low volatility regime. Interestingly, although differences tend to be of the same sign across moneyness, i.e., all the prices increase whereas all the Gammas decrease, Delta increases with conditional volatility for OTM options whereas it decreases for ITM options.

For comparison, we next report the estimated prices and Greeks that would be obtained if a CV model was used with the two different values of implied volatilities as the (long run unconditional) level of volatility in Panels C and D, respectively. These are essentially the Greeks that would be obtained if one was to use the implied volatility from the observed market prices of options directly in a CV model and then calculate the Greeks based on this model, a standard approach used in the industry when reporting the Greeks for empirical option data. The first thing to notice from these two panels is that the differences in estimates are much larger for the CV model than for the NGARCH model. This is particularly so for the prices which differ by a factor of 35 for the OTM option. While differences in Greeks are less extreme, for the OTM option the Delta is still almost 4 times larger (in absolute value) in a high volatility regime.

Finally, by comparing the results for the NGARCH and CV models Table 8 demonstrates the importance of allowing for time varying volatility and showcases that large errors are made in estimating prices and Greeks if this is ignored. For example, when volatility is high the estimated price with the CV model is more than twice the NGARCH price whereas the estimated Delta is 35% large and the estimated Gamma is 20% smaller. When volatility is low the corresponding errors on the estimated Greeks are even bigger and in this case both of them are underestimated with a CV model. Thus, our results clearly show that if the true model has time varying volatility, the errors that would be made when hedging options using Greeks from a wrong model could be significant, even if the correct level of spot volatility is used and a position in derivatives hedged using market implied Greeks would in fact be exposed to massive amounts of risk when volatility is time varying. We leave a thorough analysis of this important empirical question for future research.

5 Conclusion

Simulation methods are important for option pricing because of their flexibility, and efficient algorithms now exist for pricing both European and American style derivatives. However, a much more important issue is to calculate the various hedging parameters, price sensitivities, or Greeks that market participants rely on for managing their positions. The Greeks are used by financial institutions for hedging and risk assessment purposes throughout the life of the option. While several methods have been developed to calculate the Greeks of European style options, much less research has dealt with American style options because of the need to simultaneously determine the optimal early exercise strategy which significantly complicates matters.

This paper provides a comprehensive study of the joint estimation of prices and price sensitivities for American style options using flexible simulation and regression based Monte Carlo methods combined with initially dispersed state variables. First, we contribute to the literature by studying the asymptotic properties of the suggested estimators, obtained from an initial cross sectional regression, and we prove convergence under mild regularity conditions. Second, based on our theoretical developments, we propose an efficient 2-stage method, which combines the flexible simulation method with an adaptive choice of optimal initially dispersed state variables to control and balance off the bias of the estimates against their variance.

The quality of the estimates crucially depends on how state variables are initially dispersed, and our proposed 2-stage method automatically makes that selection and is straightforward to implement in practice. Our numerical results show that the method compares exceptionally well to existing alternatives, works for very reasonable choices of dispersion sizes, regressors, and simulated paths and is robust to choices of these parameters. We apply the method to models with time varying volatility demonstrating that there are large differences between estimated Greeks with affine and non-affine models, that Greeks vary significantly through periods of crisis, and that the errors made by using Greeks implied from misspecified models with constant volatility can be extremely large.

Since the method we propose relies on nothing but simple polynomial approximations in regression based Monte Carlo simulation methods it is extremely flexible and easy to use. Thus, our paper successfully proposes a method that is generally applicable and can be used for simultaneous estimation of prices and Greeks for American style options in very general settings using simple regression-based simulation methods. In particular, our proposed methodology can be applied to obtain price sensitivities with respect to any of the stochastic or time varying state variable that determine an option's price and should therefore be of immediate interest and have broad applications.

References

BOLLERSLEV, T. (1986): "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31(3), 307–327.

BOUCHARD, B., AND X. WARIN (2012): "Monte-Carlo Valuation of American Options: Facts and new Algorithms to Improve Existing Methods," in *Numerical Methods in Finance*, pp. 215–255. Springer.

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

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

CHEN, N., AND Y. LIU (2014): "American Option Sensitivities Estimation via a Generalized Infinitesimal Perturbation Analysis Approach," *Operations Research*, 62(3), 616–632.

CHOW, Y. S., H. ROBBINS, AND D. SIEGMUND (1971): *Great Expectations: The Theory of Optimal Stopping*. Houghton Mifflin, New York.

CHRISTOFFERSEN, P., C. DORION, K. JACOBS, AND Y. WANG (2010): "Volatility Components, Affine Restrictions, and Nonnormal Innovations," *Journal of Business and Economic Statistics*, 28(4), 483–502.

CHRISTOFFERSEN, P., K. JACOBS, AND C. ORNTHANALAI (2013): "GARCH Option Valuation: Theory and Evidence," *The Journal of Derivatives*, 21(2), 8–41.

DELECROIX, M., AND A. C. ROSA (1996): "Nonparametric Estimation of a Regression Function and its Derivatives under an Ergodic Hypothesis," *Journal of Nonparametric Statistics*, 6(4), 367–382.

DUAN, J.-C. (1995): "The GARCH Option Pricing Model," *Mathematical Finance*, 5(1), 13–32.

DUFFIE, D. (1996): "Dynamic asset pricing theory," *Princeton University*.

Engle, R. F. (1993): "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, pp. 987–1007.

Engle, R. F., and V. K. Ng (1982): "Measuring and Testing the Impact of News on Volatility," *Journal of Finance*, 48(5), 1749–1778.

Fan, J., and I. Gijbels (1992): "Variable Bandwidth and Local Linear Regression Smoothers," *The Annals of Statistics*, pp. 2008–2036.

Fan, J., and I. Gijbels (1995a): "Adaptive Order Polynomial Fitting: Bandwidth Robustification and Bias Reduction," *Journal of Computational and Graphical Statistics*, 4(3), 213–227.

——— (1995b): "Data-driven Bandwidth Selection in Local Polynomial Fitting-Variable Bandwidth and Spatial Adaptation," *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(2), 371–394.

——— (1996): *Local Polynomial Modelling and its Applications: Monographs on Statistics and Applied Probability 66*, vol. 66. CRC Press.

Feng, G., G. Liu, and L. Sun (2013): "A Nonparametric Method for Pricing and Hedging American Options," in *Proceedings of the 2013 Winter Simulation Conference*, ed. by R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl, pp. 114–147. Edward Elgar Publishing.

Glasserman, P. (2004): *Monte Carlo Methods in Financial Engineering*. Springer-Verlag, New York, Inc., New York, USA.

Glasserman, P., and B. Yu (2002): "Simulation for American Options: Regression Now or Regression Later?," in *Monte Carlo and Quasi-Monte Carlo Methods 2002*, ed. by H. Niederreiter, Berlin. Springer.

Heston, S. L., and S. Nandi (2000): "A Closed-Form GARCH Option Valuation Model," *The Review of Financial Studies*, 13(3), 585–625.

Hull, J. C. (2006): *Options, Futures, and Other Derivatives*. Prentice-Hall, Inc., Upper Saddle River, New Jersey.

Jain, S., and C. W. Oosterlee (2015): "The Stochastic Grid Bundling Method: Efficient Pricing of Bermudan Options and their Greeks," *Applied Mathematics and Computation*, 269, 412–431.

Kaniel, R., S. Tompaidis, and A. Zemliano (2008): "Efficient Computation of Hedging Parameters for Discretely Exercisable Options," *Operations Research*, 56(4), 811–826.

Kanniainen, J., B. Lin, and H. Yang (2014): "Estimating and using GARCH models with VIX data for Option Valuation," *Journal of Banking and Finance*, (43), 200–211.

Karatzas, I. (1988): "On the Pricing of American Options," *Applied Mathematics and Optimization*, 17(1), 37–60.

Longstaff, F. A., and E. S. Schwartz (2001): "Valuing American Options by Simulation: A Simple Least-Squares Approach," *Review of Financial Studies*, 14, 113–147.

Masry, E. (1996): "Multivariate Local Polynomial Regression for Time Series: Uniform Strong Consistency and Rates," *Journal of Time Series Analysis*, 17(6), 571–599.

Masry, E., and J. Fan (1997): "Local Polynomial Estimation of Regression Functions for Mixing Processes," *Scandinavian Journal of Statistics*, 24(2), 165–179.

Pelsser, A., and T. Vorst (1994): "The Binomial Model and the Greeks," *Journal of Derivatives*, 1, 45–49.

Piterbarg, V. V. (2014): "Pricing and Hedging Callable Libor Exotics in Forward Libor Models," *Journal of Computational Finance*, 8(2), 65–117.

Rasmussen, N. (2005): "Control Variates for Monte Carlo Valuation of American Options," *Journal of Computational Finance*, 9(1), 2–12.

Rombouts, J., L. Stentoft, and F. Violante (2020): "Pricing individual stock options using both stock and market index information," *Journal of Banking and Finance*, 111, 105727.

Royden, H. L. (1988): *Real Analysis*. Prentice Hall, Inc., Upper Saddle River, New Jersey.

Stentoft, L. (2004): "Convergence of the Least Squares Monte Carlo Approach to American Option Valuation," *Management Science*, 9(50), 1193–1203.

——— (2005): "Pricing American Options when the Underlying Asset Follows GARCH Processes," *Journal of Empirical Finance*, 12(4), 576–611.

——— (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), 1–56.

Stone, C. J. (1977): "Consistent Nonparametric Regression," *The Annals of Statistics*, 5(4), 595–620.

Tsitsiklis, J. N., and B. Van Roy (2001): "Regression Methods for Pricing Complex American-Style Options," *IEEE Transactions on Neural Networks*, 12(4), 694–703.

Wang, Y., and R. Caflisch (2010): "Pricing and Hedging American-Style Options: A simple Simulation-Based Approach," *Journal of Computational Finance*, 13(4), 95–125.

A Technical details

This appendix contains the proof of our main theorem along with some technical details on how to reduce the variance and control the bias in the proposed 2-stage method.

A.1 Proof of Theorem 1

Proof of Theorem 1. We prove Theorem 1 in three steps as follows:

Step 1. Random ISD with known stopping time: Note that under Assumptions 1, 2 and 3 the following holds:

1. By Assumption 1 the density of X is continuous, symmetric and bounded away from zero and thus $f(x_0) > 0$ and $f'(x_0) = 0$, and $f(\cdot)$ is continuous in a neighborhood of x_0.
2. By Assumption 2 and the properties of the payoffs, the conditional variance $\sigma^2(\cdot)$ exists and it is bounded and continuous in a neighborhood of x_0.
3. By Assumption 3 P^{M+2}, the $M+2$'th derivative of the regression function P, exists and it is continuous in a neighborhood of x_0.

Finally, note that using an OLS regression, which gives equal weight to all observations, corresponds to using a uniform kernel, given by

$$\mathcal{K}_h\left(X_n - x_0\right) = \frac{1}{2h}\mathbf{1}\left(\left|\frac{X_n - x_0}{h}\right| \leq 1\right), \tag{A.1}$$

in the LPR setting. Letting $\alpha \to 0$ and setting $h = \max\left[X_{(N)} - x_0,\ x_0 - X_{(1)}\right]$, where $X_{(i)}$ denotes the ith order statistic, it follows that $h \to 0$, and thus all the assumptions of Lemma 3 are satisfied. This implies that the derivatives are asymptotically unbiased when $N\alpha^3 \to \infty$ and that the asymptotic variance tends to zero when $N\alpha^{1+2i} \to \infty$, which proves convergence.

Step 2. Estimated stopping time: When the optimal stopping time is estimated using the same paths that are used for pricing, dependence is introduced between the pathwise payoffs. However, under Assumption 4 this dependence is weak. Weak dependence functions as a substitute for strong mixing and under this assumption Masry and Fan (1997) shows that Lemma 3 continues to hold.

Step 3. Deterministic ISD: Though Lemma 3 is stated in terms of an i.i.d. sample of data it remains valid for "fixed designs" and therefore holds in our setting with deterministically generated initially dispersed state variables, see Section 3.2.4 of Fan and Gijbels (1996).

We now combine the above results with Lemma 1, which shows that when $M \to \infty$ such that $M^3/N \to 0$ then $\hat{F}_M^N\left(X\left(t_j\right)\right)$ converges to $F\left(X\left(t_j\right)\right)$, for $t_j > 0$. As a result, the estimated stopping time converges to the "true" stopping time and since the derivatives are asymptotically unbiased and the asymptotic variance tends to zero under the given assumptions, they converge to the true price and Greeks. □

A.2 Variance reduction using the Value Function at time $t = 1$

To motivate our method, we first illustrate in more detail the problem occurring with the regression at $t = 0$. At this time, the naive method uses the payoffs along each path discounted from the optimal exercise time. The top sub-figures in Figure A.1 show examples of data available at $t = 0$ for the regression. When a small ISD is used as in Sub-figure A.1(a) the data displays little or no structure at all, and when a polynomial is fitted to the data it will approximate the true function relatively well locally, though there will be a lot of variability for the Greeks. When a large ISD is used as in Sub-figure A.1(b) the data, on the other hand, displays some structure and when a

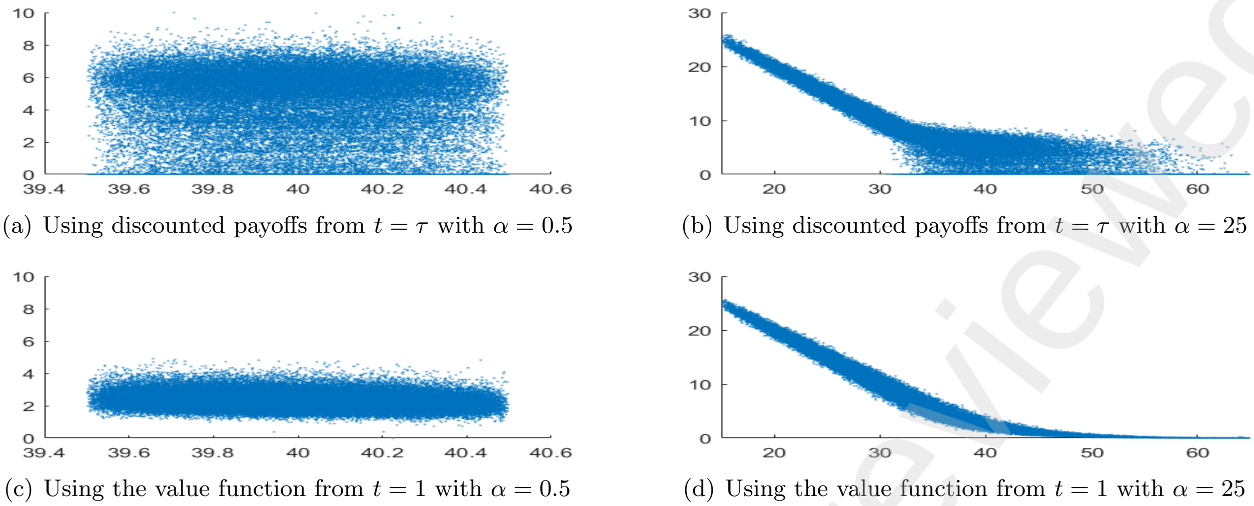


Figure A.1: Example of data used in the regression at $t = 0$ for different values of α
This figure shows an example of $N = 100,000$ paths of simulated data used in the regression at $t = 0$ when using small ISD with $\alpha = 0.5$ and a large ISD with $\alpha = 25$ in Equation (21). Top plots show the results when using the discounted payoffs whereas the bottom plots show the results when using a value function iteration in the second to last step. The initial stock price is fixed at $S(0) = 40$ and the option strike price is $K = 40$, the volatility is $\sigma = 20\%$, the interest rate is $r = 6\%$, and the maturity of the option is $T = 1$ year.

polynomial is fitted to the data it will approximate the true function relatively well overall, though there might be some small local biases on the price and large biases on the Greeks.

Our proposed solution is to add more structure to the data at $t = 0$ even when using a small ISD and to do so we use information available at $t = 1$. In the LSM algorithm the result of the regression at $t = 1$, $\hat{F}_M^N(X(t_1))$, is used only to determine which paths should be exercised. In particular, $\hat{F}_M^N(X(t_1))$, an approximation of the value of holding the option, can be compared to $Z(t_1)$, the value of exercising the option, to determine which paths should be exercised. Since the payoffs from paths that are not exercised at $t = 1$ may come from periods far in the future this creates a lot of variability. As an alternative to discounting the payoffs for each path from when it is optimal to exercise, we propose instead to simply discount directly the value function estimated at $t = 1$ given by $\hat{V}_M^N(X(t_1)) = \max\left(Z(t_1), \hat{F}_M^N(X(t_1))\right)$ for one period. This simple step removes a lot of variability in the sample data at $t = 0$. Sub-figures A.1(c) and A.1(d) show the sample data when discounting the value function from $t = 1$ for $\alpha = 0.5$ and $\alpha = 25$, respectively. Note how compact the data is compared to not using the value function.[15]

[15] The use of value function iteration should generally be avoided in this type of algorithms as shown by Stentoft (2014) since it introduces a high bias on the estimated pathwise payoffs which accumulates fast. In our proposed

Table A.1: Results with the value function method

		Price			Delta			Gamma		
K	α	BM	Estim	StDev	BM	Estim	StDev	BM	Estim	StDev
36	0.5	0.9166	0.9186	(0.0057)†	-0.1979	-0.1981	(0.0083)	0.0381	0.0232	(0.0793)
40	0.5	2.3141	2.3169	(0.0094)†	-0.4040	-0.4044	(0.0148)	0.0597	0.0305	(0.1619)
44	0.5	4.6535	4.6567	(0.0107)†	-0.6648	-0.6645	(0.0232)	0.0765	0.0318	(0.2652)
36	5	0.9166	0.9169	(0.0065)	-0.1979	-0.1983	(0.0048)	0.0381	0.0394	(0.0042)†
40	5	2.3141	2.3147	(0.0110)	-0.4040	-0.4045	(0.0070)	0.0597	0.0609	(0.0061)
44	5	4.6535	4.6559	(0.0138)	-0.6648	-0.6635	(0.0087)	0.0765	0.0776	(0.0079)
36	25	0.9166	0.8917	(0.0110)†	-0.1979	-0.1868	(0.0033)†	0.0381	0.0421	(0.0006)†
40	25	2.3141	2.2879	(0.0153)†	-0.4040	-0.4109	(0.0043)†	0.0597	0.0679	(0.0008)†
44	25	4.6535	4.7009	(0.0192)†	-0.6648	-0.6864	(0.0038)†	0.0765	0.0695	(0.0011)†

This table shows the estimated prices and Greeks for different values of α in Equation (21) with a method that uses a value function iteration in the second to last step. See also the notes to Table 3.

† Indicates that the estimate is statistically different from the benchmark value at a 1% level.

Table A.1 shows the result for the different choices of α using the "value function" method. The table, first of all, shows that the method indeed produces estimated prices and Greeks with lower standard deviation, and this particularly so when α is small. The improvement in precision is most spectacular when it comes to the Greeks, and the standard deviations of the estimated Delta and Gamma for the OTM option are more than 6 times larger with the naive method. When $\alpha = 5$, introducing the value function step at time $t = 1$ more than halves the standard deviation of the estimated Greeks. When $\alpha = 25$, the standard deviations are only around 10% larger with the naive method. Secondly, and this is an additional benefit of the value function approach, Table A.1 shows that our proposed improvement also reduces the bias of the estimated Greeks. This is particularly evident for the estimated Gamma of the ITM option when $\alpha = 0.5$, and in this case, the estimated Gamma with the naive method is -0.0519, i.e., estimated with the wrong sign. The corresponding estimate from the value function method is 0.0318 which, though still biased, has the right sign and is much closer to the benchmark value of 0.0765.

A.3 Bias control with optimal bandwidth selection

Our proposed method for controlling the bias is to use an "optimal" bandwidth to estimate the prices and Greeks. In our setting, where we are estimating the value function and the sensitivities evaluated at a particular point x_0, the objective is to find an optimal local bandwidth selector. Fan and Gijbels (1995b) provide a theoretically optimal local bandwidth formulation which is easy to

method, we only do value function iteration in one step and the bias introduced turns out to be negligible.

approximate given by

$$h_i(x_0) = \left(\frac{(2i+1)\, a_i \sigma^2(x_0)}{2(M+1-i)\, b_i^2 \beta_{M+1}^2 N f(x_0)} \right)^{\frac{1}{2M+3}}, \tag{A.2}$$

where i is the derivative to estimate, a_i is the $(i+1)^{th}$ diagonal element of the matrix $Q^{-1}Q^{*}Q^{-1}$, b_i is the $(i+1)^{th}$ diagonal element of the Q matrix, which are in terms defined as $Q = (\mu_{j+l})_{0 \le j,l \le M}$ and $Q^{*} = (\nu_{j+l})_{0 \le j,l \le M}$ where $\mu_j = \int u^j K(u)\, \mathrm{d}u$ and $\nu_j = \int u^j K^2(u)\, \mathrm{d}u$. This optimal local bandwidth contains three easy to interpret quantities: $\sigma^2(x_0)$, β_{M+1}, and $f(x_0)$. The first quantity measures the noise in the data and the noisier the data the larger the optimal bandwidth will be. The second quantity measures the roughness of the function we are trying to approximate and the rougher this function is the smaller the optimal bandwidth will be in order to avoid biased estimates. The third quantity measures how X is distributed around x_0 and the more centered this distribution is, the smaller the optimal bandwidth will be.

In our setting, the distribution of X is determined by the ISD used to generate the data. Thus, we only need to approximate $\sigma^2(x_0)$ and β_{M+1} and we need to do so around x_0. We could run a simple OLS regression using all the data, however, a better and more stable method is, naturally, to use a weighted regression with an appropriately chosen bandwidth. Here we propose to estimate these quantities using a simple Rule of Thumb (ROT) method for global bandwidth selection from Fan and Gijbels (1992) given by

$$\hat{h}_{ROT}(x_0) = C_{i,M}(K) \left[\frac{\hat{\sigma}^2 \int w(x_0)/f(x)\, dx}{\sum_{j=1}^{N} \{\hat{F}^{(M+1)}(X_i)\}^2 w_0(X_j)\}} \right]^{1/(2M+3)}, \tag{A.3}$$

where $\hat{F}$ is first obtained using a polynomial regression of order $M+3$ using all data, $\hat{\sigma}^2$ is the standardized residual sum of squares from the later polynomial regression, and where w_0 can be taken as the indicator function. In this bandwith estimator

$$C_{i,M}(K) = \left[\frac{(M+1)!^2 (2i+1) \int K_i^{*2}(t)\, dt}{2(M+1-i) \{\int t^{M+1} K_i^{*}(t)\, dt\}^2} \right]^{1/(2M+3)}, \tag{A.4}$$

where K_i^{*} denoting the equivalent kernel, see Fan and Gijbels (1996) Chapter 3 for details.

The optimal bandwidth selector from above can now be used to estimate the locally optimal

Table A.2: Results with the method that truncates paths outside the optimal α

		Price			Delta			Gamma		
K	α	BM	Estim	StDev	BM	Estim	StDev	BM	Estim	StDev
36	0.5	0.9166	0.9186	(0.0057)†	-0.1979	-0.1981	(0.0083)	0.0381	0.0232	(0.0793)
40	0.5	2.3141	2.3169	(0.0094)†	-0.4040	-0.4044	(0.0148)	0.0597	0.0305	(0.1619)
44	0.5	4.6535	4.6567	(0.0107)†	-0.6648	-0.6645	(0.0232)	0.0765	0.0318	(0.2652)
36	5	0.9166	0.9170	(0.0064)	-0.1979	-0.1984	(0.0049)	0.0381	0.0386	(0.0061)
40	5	2.3141	2.3149	(0.0112)	-0.4040	-0.4048	(0.0077)	0.0597	0.0602	(0.0105)
44	5	4.6535	4.6563	(0.0147)	-0.6648	-0.6641	(0.0099)	0.0765	0.0766	(0.0194)
36	25	0.9166	0.9164	(0.0127)	-0.1979	-0.1971	(0.0055)	0.0381	0.0381	(0.0056)
40	25	2.3141	2.3140	(0.0197)	-0.4040	-0.4038	(0.0104)	0.0597	0.0592	(0.0103)
44	25	4.6535	4.6595	(0.0296)	-0.6648	-0.6636	(0.0183)	0.0765	0.0768	(0.0209)

This table shows the estimated prices and Greeks for different values of α in Equation (21) with a method that, in addition to using a value function iteration at the second to last step, truncates paths outside the optimal α and uses only the paths that are inside to estimate the price and the Greeks. See also the notes to Table 3.

† Indicates that the estimate is statistically different from the benchmark value at a 1% level.

ISD size. Specifically, our suggested method proceeds as follows: First, obtain an approximation of the optimal global bandwidth using the ROT in Equation (A.3). Second, use the ROT to locally fit a polynomial and estimate $\sigma^2(x_0)$ and β_{M+1}. Finally, plug those values into Equation (A.2) to obtain an approximation of the optimal local bandwidth. This procedure is quick, simple, and in our application provides good results in terms of bandwidth selection.[16] The prices and Greeks can then be estimated using only the paths that fall within this bandwidth by simply truncating all the paths that fall outside this bandwidth.

Table A.2 shows the results for the different choices of initial α using this "truncated" method. The table, first of all, shows that when the initial spread is generated with an $\alpha = 25$ the method corrects very efficiently the bias from the large spread. This is particularly clear for the estimated Gammas, which are biased by as much as $13,4\%$ when using the value function method alone. When selecting the optimal α and using this to estimate the Greeks this bias is almost eliminated and estimated Gammas are less than 1% biased. The price to pay for this, though, is an increase in the standard deviation of the estimate as a large fraction of the paths are potentially truncated. Note that using this method to determine the optimal α also decreases the bias for higher order Greeks when using a small ISD though not to the same extent as when using a large ISD.

[16] Appendix B.3 shows that the best approach is to select the bandwidth with respect to the highest order derivative.

Table B.1: Results for a large sample of options

			Price			Delta			Gamma		
K	σ	T	BM	Estim	StDev	BM	Estim	StDev	BM	Estim	StDev
36	10%	0.5	0.0304	0.0304	(0.0008)	-0.0281	-0.0281	(0.0012)	0.0236	0.0240	(0.0027)
40	10%	0.5	0.7347	0.7339	(0.0048)	-0.4088	-0.4079	(0.0063)	0.1846	0.1882	(0.0088)†
44	10%	0.5	3.9473	3.9478	(0.0035)	-0.9998	-1.0002	(0.0057)	0.0010	-0.0020	(0.0189)
36	20%	0.5	0.4978	0.4984	(0.0042)	-0.1607	-0.1604	(0.0038)	0.0449	0.0451	(0.0043)
40	20%	0.5	1.7915	1.7925	(0.0090)	-0.4256	-0.4244	(0.0066)	0.0790	0.0791	(0.0085)
44	20%	0.5	4.3091	4.3099	(0.0118)	-0.7563	-0.7542	(0.0101)	0.0907	0.0918	(0.0135)
36	40%	0.5	2.1993	2.2011	(0.0119)	-0.2759	-0.2747	(0.0076)	0.0305	0.0303	(0.0097)
40	40%	0.5	3.9718	3.9740	(0.0155)	-0.4186	-0.4167	(0.0102)	0.0367	0.0360	(0.0142)
44	40%	0.5	6.3262	6.3287	(0.0200)	-0.5637	-0.5613	(0.0131)	0.0389	0.0377	(0.0187)
36	10%	1.0	0.0895	0.0893	(0.0017)	-0.0545	-0.0545	(0.0020)	0.0305	0.0308	(0.0040)
40	10%	1.0	0.8893	0.8888	(0.0050)	-0.3901	-0.3901	(0.0058)	0.1505	0.1517	(0.0110)
44	10%	1.0	3.9474	3.9477	(0.0030)	-0.9989	-0.9998	(0.0059)	0.0054	0.0039	(0.0175)
36	20%	1.0	0.9166	0.9159	(0.0067)	-0.1979	-0.1978	(0.0047)	0.0381	0.0390	(0.0049)
40	20%	1.0	2.3141	2.3141	(0.0102)	-0.4040	-0.4041	(0.0073)	0.0597	0.0604	(0.0085)
44	20%	1.0	4.6535	4.6544	(0.0123)	-0.6648	-0.6632	(0.0093)	0.0765	0.0773	(0.0120)
36	40%	1.0	3.4366	3.4368	(0.0165)	-0.2863	-0.2858	(0.0078)	0.0227	0.0232	(0.0088)
40	40%	1.0	5.3120	5.3137	(0.0216)	-0.3903	-0.3894	(0.0104)	0.0265	0.0271	(0.0115)
44	40%	1.0	7.6104	7.6133	(0.0241)	-0.4966	-0.4947	(0.0128)	0.0291	0.0292	(0.0149)
36	10%	2.0	0.1713	0.1714	(0.0026)	-0.0751	-0.0754	(0.0028)	0.0313	0.0310	(0.0049)
40	10%	2.0	1.0241	1.0232	(0.0057)	-0.3729	-0.3720	(0.0064)	0.1301	0.1302	(0.0116)
44	10%	2.0	3.9480	3.9483	(0.0031)	-0.9963	-0.9975	(0.0056)	0.0161	0.0171	(0.0168)
36	20%	2.0	1.4317	1.4319	(0.0102)	-0.2165	-0.2163	(0.0056)	0.0311	0.0308	(0.0066)
40	20%	2.0	2.8846	2.8840	(0.0144)	-0.3796	-0.3791	(0.0083)	0.0468	0.0472	(0.0092)
44	20%	2.0	5.0832	5.0826	(0.0157)	-0.5897	-0.5881	(0.0100)	0.0639	0.0641	(0.0116)
36	40%	2.0	4.9643	4.9649	(0.0226)	-0.2786	-0.2779	(0.0072)	0.0168	0.0157	(0.0089)
40	40%	2.0	6.9171	6.9187	(0.0269)	-0.3552	-0.3538	(0.0089)	0.0195	0.0186	(0.0111)
44	40%	2.0	9.1820	9.1833	(0.0317)	-0.4342	-0.4326	(0.0111)	0.0219	0.0209	(0.0134)

This table shows the estimated prices and Greeks for 27 different options. We report averages of 100 independent simulations with $N = 100,000$ paths. The strike price, volatility, and time to maturity are shown in the first three columns. The initial stock price is fixed at $S(0) = 40$ and the interest rate is fixed at $r = 6\%$. The benchmark values are from the Binomial Model with $50,000$ steps and $J = 50$ early exercise points a year. The initial alpha is set to $\alpha = 10$. The optimal early exercise strategy is estimated with $M = 9$ and the prices and Greeks are estimated using a polynomial of order $M_0 = 9$.

† Indicates that the estimate is statistically different from the benchmark value at a 1% level.

B Robustness

In this Appendix, we demonstrate the robustness of our proposed method along several dimensions. First, we consider a much large sample of options than in Section 3. Next, we consider different choices of the number of simulated paths and the number of regressors used at time $t = 0$, as well as using alternative polynomial orders to determine the optimal stopping time. Finally, we consider different choices of the ISD and methods for determining the optimal α.

Table B.2: Results across interest rates and dividend yields

			Price			Delta			Gamma		
K	r	d	BM	Estim	StDev	BM	Estim	StDev	BM	Estim	Stdev
36	0%	0%	1.4356	1.4376	(0.0099)	-0.2654	-0.2660	(0.0065)	0.0410	0.0407	(0.0075)
40	0%	0%	3.1862	3.1891	(0.0155)	-0.4602	-0.4608	(0.0091)	0.0496	0.0498	(0.0110)
44	0%	0%	5.7168	5.7211	(0.0200)	-0.6467	-0.6472	(0.0120)	0.0465	0.0466	(0.0148)
36	6%	0%	0.9166	0.9166	(0.0061)	-0.1979	-0.1976	(0.0046)	0.0381	0.0381	(0.0054)
40	6%	0%	2.3141	2.3144	(0.0109)	-0.4040	-0.4045	(0.0066)	0.0597	0.0591	(0.0093)
44	6%	0%	4.6535	4.6531	(0.0119)	-0.6648	-0.6660	(0.0092)	0.0765	0.0752	(0.0135)
36	6%	6%	1.3646	1.3652	(0.0089)	-0.2534	-0.2534	(0.0057)	0.0395	0.0394	(0.0070)
40	6%	6%	3.0420	3.0416	(0.0127)	-0.4433	-0.4431	(0.0082)	0.0490	0.0497	(0.0101)
44	6%	6%	5.4907	5.4910	(0.0163)	-0.6315	-0.6305	(0.0105)	0.0482	0.0471	(0.0133)

This table shows the estimated prices and Greeks for the three options from Section 3 for different values of the interest rate and dividend yield. We report averages of 100 independent simulations with $N = 100,000$ paths. The strike price, interest rate, and dividend yield are shown in the first three columns. The initial stock price is fixed at $S(0) = 40$, the volatility is fixed at $\sigma = 20\%$, and the time to maturity is $T = 1$ year. The benchmark values are from the Binomial Model with $50,000$ steps and $J = 50$ early exercise points a year. The initial alpha is set to $\alpha = 10$. The optimal early exercise strategy is estimated with $M = 9$ and the prices and Greeks are estimated using a polynomial of order $M_0 = 9$.

† Indicates that the estimate is statistically different from the benchmark value at a 1% level.

B.1 Robustness across option characteristics

We first report results for a wider range of options with different moneyness and volatilities of $\sigma \in \{10\%, 20\%, 40\%\}$ and maturities of $T \in \{0.5, 1, 2\}$ years, for a total of 27 different options. For all options, we take the initial stock price to be $S(0) = 40$, the interest rate to be $r = 6\%$, and we consider $J = 50$ early exercise points per year. The initial alpha is set to $\alpha = 10$. For this sample of options, Table B.1 shows that our 2-stage method works well across the board. In particular, of the 27 prices, Deltas, and Gammas only one of the estimates is statistically different from the benchmark value at a 1% level, roughly the number you would expect at this significance level for a sample of this size. Indeed, irrespective of the maturity and volatility, when sensitivities exist our proposed method works very well.

We next consider a subset of options with various different values of the risk free interest rate and of the dividend yield. In particular, we consider the limiting cases when the risk neutral drift of the underlying approaches the zero level, i.e. when the dividend yield is equal to the interest rate, and the case when the interest rate is very low. The results are shown in Table B.2 which allows us to conclude that our method works well across various levels of interest rates and dividend yields. The only case for which the t-statistics are somewhat large occurs when both the interest rate and dividend yield are zero. In this case, however, it is known that the option should never be exercised, but the simulation algorithm at times will choose so, essentially because there is a slight degree of over fitting, and this generates the significant differences.[17] Note that this only affects the prices whereas the Greeks continue to be estimated insignificantly different from their true values.

B.2 Robustness across choices of regressors and number of paths

The results presented so far are based on using $M_0 = 9$ regressors in the initial regression, using $N = 100,000$ simulated paths, and using $M = 9$ regressors in the cross-sectional regressions to determine the optimal stopping time. The properties of the estimated price, Delta, and Gamma are expected to depend on the parameter choices for the number of regressors at time $t = 0$ and the number of simulated paths in particular and to analyse this we now consider alternative choices for each of these parameters. We consider two different values of M_0, $M_0 = 5$ and $M_0 = 15$, and of N, $N = 50,000$ and $N = 200,000$, the results for which are reported in Table B.3.

Table B.3 shows that the choice of M_0 and N indeed affects the quality of the estimated prices and the estimated Greeks in particular. For example, with respect to the polynomial order the table shows that when this is low and $M_0 = 5$ several of the estimates are statistically different from the benchmark and this is particularly so when N is low also. When $M_0 = 15$, on the other hand, only one of the estimates is statistically different from the benchmark and this is, again, for the case when N is low. For a reasonable choice of polynomial order, e.g. $M_0 = 9$, the table shows that the statistical significance does not change with the number of simulated paths N. With respect to the number of simulated paths the table shows that most of the cases for which the estimates are statistically significant happen when this is low and $N = 50,000$. When $N = 200,000$ only two of the estimated sensitivities are significant and again this happens when a low order polynomial is used at $t = 0$.

We next consider the effect of using different orders of the polynomial used to determine the optimal early exercise strategy, M. The choice of the polynomial order used in the cross-sectional regressions could potentially also affect the performance of the algorithm for determining price, Delta, and Gamma. To examine the effect of using an alternative number of regressors to estimated the optimal early exercise strategy we now consider two different values of $M = 5$ and $M = 15$. The

[17]The slight overfitting is due to using "only" $100,000$ simulated paths. When increasing the number of simulated paths to $500,000$ prices, as well as the Greeks, are insignificantly different from the benchmark values.

Table B.3: Results across number of paths, N, and polynomial order for initial regression, M_0

				Price			Delta			Gamma	
K	N'	M_0	BM	Estim	StDev	BM	Estim	StDev	BM	Estim	StDev
36	50	5	0.9166	0.9073	(0.0239)†	-0.1979	-0.1980	(0.0065)	0.0381	0.0397	(0.0037)†
40	50	5	2.3141	2.3314	(0.0894)	-0.4040	-0.4103	(0.0174)†	0.0597	0.0612	(0.0047)†
44	50	5	4.6535	4.6961	(0.1119)†	-0.6648	-0.6731	(0.0100)†	0.0765	0.0738	(0.0098)†
36	100	5	0.9166	0.9108	(0.0192)†	-0.1979	-0.1985	(0.0047)	0.0381	0.0393	(0.0029)†
40	100	5	2.3141	2.3105	(0.0239)	-0.4040	-0.4047	(0.0090)	0.0597	0.0618	(0.0033)†
44	100	5	4.6535	4.6700	(0.0512)†	-0.6648	-0.6705	(0.0088)†	0.0765	0.0763	(0.0072)
36	200	5	0.9166	0.9135	(0.0164)	-0.1979	-0.1981	(0.0078)	0.0381	0.0386	(0.0024)
40	200	5	2.3141	2.3116	(0.0123)	-0.4040	-0.4046	(0.0062)	0.0597	0.0607	(0.0029)†
44	200	5	4.6535	4.6619	(0.0380)	-0.6648	-0.6689	(0.0075)†	0.0765	0.0775	(0.0060)
36	50	9	0.9166	0.9183	(0.0105)	-0.1979	-0.1977	(0.0074)	0.0381	0.0379	(0.0088)
40	50	9	2.3141	2.3166	(0.0166)	-0.4040	-0.4027	(0.0117)	0.0597	0.0599	(0.0128)
44	50	9	4.6535	4.6562	(0.0167)	-0.6648	-0.6622	(0.0156)	0.0765	0.0756	(0.0166)
36	100	9	0.9166	0.9159	(0.0067)	-0.1979	-0.1978	(0.0047)	0.0381	0.0390	(0.0049)
40	100	9	2.3141	2.3141	(0.0102)	-0.4040	-0.4041	(0.0073)	0.0597	0.0604	(0.0085)
44	100	9	4.6535	4.6544	(0.0123)	-0.6648	-0.6632	(0.0093)	0.0765	0.0773	(0.0120)
36	200	9	0.9166	0.9172	(0.0052)	-0.1979	-0.1982	(0.0035)	0.0381	0.0378	(0.0036)
40	200	9	2.3141	2.3142	(0.0076)	-0.4040	-0.4053	(0.0055)	0.0597	0.0595	(0.0070)
44	200	9	4.6535	4.6533	(0.0080)	-0.6648	-0.6646	(0.0082)	0.0765	0.0765	(0.0103)
36	50	15	0.9166	0.9162	(0.0087)	-0.1979	-0.1958	(0.0071)†	0.0381	0.0371	(0.0117)
40	50	15	2.3141	2.3157	(0.0138)	-0.4040	-0.4015	(0.0118)	0.0597	0.0588	(0.0213)
44	50	15	4.6535	4.6539	(0.0153)	-0.6648	-0.6611	(0.0183)	0.0765	0.0758	(0.0288)
36	100	15	0.9166	0.9164	(0.0062)	-0.1979	-0.1974	(0.0046)	0.0381	0.0382	(0.0082)
40	100	15	2.3141	2.3136	(0.0099)	-0.4040	-0.4045	(0.0075)	0.0597	0.0616	(0.0142)
44	100	15	4.6535	4.6528	(0.0128)	-0.6648	-0.6642	(0.0101)	0.0765	0.0765	(0.0217)
36	200	15	0.9166	0.9160	(0.0044)	-0.1979	-0.1979	(0.0037)	0.0381	0.0379	(0.0049)
40	200	15	2.3141	2.3136	(0.0073)	-0.4040	-0.4044	(0.0061)	0.0597	0.0590	(0.0084)
44	200	15	4.6535	4.6527	(0.0087)	-0.6648	-0.6654	(0.0085)	0.0765	0.0763	(0.0124)

This table shows the estimated prices and Greeks for the three options from Section 3 for different values of the number of simulated paths and polynomial order used at time $t = 0$. We report averages of 100 independent simulations. The strike price, number of paths (in thousands), and polynomial order are shown in the first three columns. The initial stock price is fixed at $S(0) = 40$, the volatility is fixed at $\sigma = 20\%$, the interest rate is fixed at $r = 6\%$, and the time to maturity is $T = 1$ year. The benchmark values are from the Binomial Model with $50,000$ steps and $J = 50$ early exercise points a year. The initial alpha is set to $\alpha = 10$. The optimal early exercise strategy is estimated with $M = 9$.

† Indicates that the estimate is statistically different from the benchmark value at a 1% level.

results are shown in Table B.4 which documents that the choice of the order of the polynomial used to determine the optimal early exercise strategy has surprisingly little influence on the estimates obtained with our proposed 2-stage method.

B.3 Robustness to the choice of ISD and optimal α

The results presented so far are based on using a deterministic method for generating the ISD and an optimal α determined based on minimizing the root mean squared error (RMSE) of the estimated second order derivative, the Gamma. We now show that our method is also robust to

Table B.4: Results across stopping time polynomial order M

			Price			Delta			Gamma	
K	M	BM	Estim	StDev	BM	Estim	StDev	BM	Estim	StDev
36	5	0.9166	0.9163	(0.0066)	-0.1979	-0.1972	(0.0045)	0.0381	0.0386	(0.0057)
40	5	2.3141	2.3131	(0.0107)	-0.4040	-0.4042	(0.0072)	0.0597	0.0611	(0.0087)
44	5	4.6535	4.6518	(0.0132)	-0.6648	-0.6645	(0.0094)	0.0765	0.0774	(0.0121)
36	9	0.9166	0.9159	(0.0067)	-0.1979	-0.1978	(0.0047)	0.0381	0.0390	(0.0049)
40	9	2.3141	2.3141	(0.0102)	-0.4040	-0.4041	(0.0073)	0.0597	0.0604	(0.0085)
44	9	4.6535	4.6544	(0.0123)	-0.6648	-0.6632	(0.0093)	0.0765	0.0773	(0.0120)
36	15	0.9166	0.9163	(0.0066)	-0.1979	-0.1971	(0.0045)	0.0381	0.0385	(0.0056)
40	15	2.3141	2.3131	(0.0105)	-0.4040	-0.4040	(0.0070)	0.0597	0.0607	(0.0086)
44	15	4.6535	4.6524	(0.0132)	-0.6648	-0.6642	(0.0091)	0.0765	0.0766	(0.0113)

This table shows the estimated prices and Greeks for the three options from Section 3 for different values of the polynomial order M used to estimate the optimal stopping time. We report averages of 100 independent simulations. The strike price and the polynomial order are shown in the first two columns. The initial stock price is fixed at $S(0) = 40$, the volatility is fixed at $\sigma = 20\%$, the interest rate is fixed at $r = 6\%$, and the time to maturity is $T = 1$ year. The benchmark values are from the Binomial Model with $50,000$ steps and $J = 50$ early exercise points a year. The initial alpha is set to $\alpha = 10$. The prices and Greeks are estimated using a polynomial of order $M_0 = 9$.

† Indicates that the estimate is statistically different from the benchmark value at a 1% level.

these fundamental choices of how to implement the proposed 2-stage method. First, it is well known from the literature on nonparametrics, in general, and on LPR, in particular, that the optimal kernel distribution for a weighted least square regression is the Epanechnikov distribution. However, other choices are possible and here we consider the case when the ISD kernel in Equation (23) is instead the Uniform distribution. We also consider the case where the ISD is generated from a random uniformly distributed vector U in Equation (23).

Theorem 1 shows that the performance of the method is improved when the sample density is peaked at the current value of the state variables and this is the case when generating the ISD from an Epanechnikov distribution. In this respect using a uniform ISD is among the worst possible choices one could consider. Table B.5 indeed shows a clear negative effect on the estimates, particularly of higher order derivatives, of using the Uniform kernel. For example, the standard deviation of the Gamma for the ITM options is more then 4 times larger when using a Uniform kernel than when using the Epanechnikov kernel. The bias is also often higher with the Uniform kernel. With respect to using a random ISD the table shows that this in general increases the standard deviation of the estimates although the effect on the bias is minor. Thus, Table B.5 does indicate that a deterministic and peaked kernel is preferred.

Table B.5 also documents a clear effect of choosing the optimal α differently. For example,

Table B.5: Results across ISD type and optimal α

			Price			Delta			Gamma		
K	ISD	i	BM	Estim	StDev	BM	Estim	StDev	BM	Estim	StDev
36	Uni.D.	2	0.9166	0.9167	(0.0060)	-0.1979	-0.1974	(0.0054)	0.0381	0.0391	(0.0169)
40	Uni.D.	2	2.3141	2.3142	(0.0096)	-0.4040	-0.4036	(0.0088)	0.0597	0.0613	(0.0315)
44	Uni.D.	2	4.6535	4.6522	(0.0115)	-0.6648	-0.6638	(0.0133)	0.0765	0.0780	(0.0472)
36	Epa.D.	2	0.9166	0.9163	(0.0068)	-0.1979	-0.1970	(0.0045)	0.0381	0.0386	(0.0056)
40	Epa.D.	2	2.3141	2.3132	(0.0108)	-0.4040	-0.4040	(0.0070)	0.0597	0.0611	(0.0083)
44	Epa.D.	2	4.6535	4.6521	(0.0133)	-0.6648	-0.6643	(0.0089)	0.0765	0.0769	(0.0115)
36	Epa.R.	2	0.9166	0.9164	(0.0087)	-0.1979	-0.1973	(0.0045)	0.0381	0.0377	(0.0061)
40	Epa.R.	2	2.3141	2.3134	(0.0132)	-0.4040	-0.4041	(0.0070)	0.0597	0.0593	(0.0087)
44	Epa.R.	2	4.6535	4.6527	(0.0161)	-0.6648	-0.6648	(0.0095)	0.0765	0.0759	(0.0128)
36	Epa.D.	0	0.9166	0.9168	(0.0054)	-0.1979	-0.1984	(0.0086)	0.0381	0.0385	(0.0718)
40	Epa.D.	0	2.3141	2.3139	(0.0084)	-0.4040	-0.4050	(0.0150)	0.0597	0.0609	(0.1301)
44	Epa.D.	0	4.6535	4.6523	(0.0106)	-0.6648	-0.6672	(0.0227)	0.0765	0.0759	(0.1928)
36	Epa.D.	1	0.9166	0.9167	(0.0056)	-0.1979	-0.1976	(0.0051)	0.0381	0.0391	(0.0143)
40	Epa.D.	1	2.3141	2.3140	(0.0091)	-0.4040	-0.4045	(0.0080)	0.0597	0.0616	(0.0220)
44	Epa.D.	1	4.6535	4.6522	(0.0114)	-0.6648	-0.6652	(0.0105)	0.0765	0.0788	(0.0290)
36	Epa.D.	2	0.9166	0.9163	(0.0068)	-0.1979	-0.1970	(0.0045)	0.0381	0.0386	(0.0056)
40	Epa.D.	2	2.3141	2.3132	(0.0108)	-0.4040	-0.4040	(0.0070)	0.0597	0.0611	(0.0083)
44	Epa.D.	2	4.6535	4.6521	(0.0133)	-0.6648	-0.6643	(0.0089)	0.0765	0.0769	(0.0115)

This table shows the estimated prices and Greeks for the three options from Section 3 for different methods to generate the ISD and to obtain the optimal α. We report averages of 100 independent simulations. The strike price, method to generate the ISD, where "Epa." denotes the Epanechnikov, "Uni." denotes the Uniform kernel, "D." is denotes a deterministic, and "R." denotes a random ISD, and the derivative, i, alpha is optimized to are shown in the first three columns. The initial stock price is fixed at $S(0) = 40$, the volatility is fixed at $\sigma = 20\%$, the interest rate is fixed at $r = 6\%$, and the time to maturity is $T = 1$ year. The benchmark values are from the Binomial Model with $50,000$ steps and $J = 50$ early exercise points a year. The initial alpha is set to $\alpha = 10$. The optimal early exercise strategy is estimated with $M = 9$ and the prices and Greeks are estimated using a polynomial of order $M_0 = 9$.

† Indicates that the estimate is statistically different from the benchmark value at a 1% level.

when selecting the optimal α to minimize the RMSE of the Delta the first thing to notice is that our proposed method continues to provide very good estimates of the price as well as the Delta. The results for the estimated Gamma though change and the standard deviation of the estimates increase significantly and in some cases almost triples. When choosing the optimal bandwidth for the price the results differ quite a bit and now the method only provides reasonable results for the actual price estimates. Note that when the objective is to estimate only the price one should pick $\alpha = 0$ and use the standard LSM method. Our optimal bandwidth selector adapts to this and picks α^* so small that it is impossible to estimate higher order derivatives. Thus, the results show that though the method is relatively robust to picking an α that is optimal for estimating slightly lower order derivatives, the best results overall are obtained when α is chosen to be optimal for the highest order derivative to be estimated.