Simulated Greeks for American Options *

Pascal Letourneau† and Lars Stentoft‡

December 14, 2019

Abstract

This paper considers estimation of price sensitivities, so-called Greeks, for American style options using flexible simulation methods combined with initial dispersed state variables. The asymptotic properties of the estimators are studied and convergence of the method is established under mild regularity conditions. A 2-step method is proposed with an adaptive choice of optimal initially dispersed state variables, that controls and balances off the bias of the estimates against their variance. Numerical results show that the method works extremely well for very reasonable choices of spread sizes, regressors, and simulated paths and demonstrate that the proposed method compares well to existing alternatives.

JEL Classification: C15, G12, G13

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

1 Introduction

Option pricing, in particular for options that have early exercise features, remains a challenge. However, while option pricing is interesting in itself, a much more important issue in finance is to calculate the various relevant hedging parameters or price sensitivities that market participants rely on for managing their positions. After all, you only price a derivative once, but when it is traded the risk exposures will generally need to be hedged through time. Price sensitivities, or Greeks for short, are used on a daily basis by financial institutions for risk management and having these readily available in real time is a necessity for these firms to conduct their business efficiently. Unlike prices which are observed in the market place, the Greeks are generally not observed and will instead always 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 and has been used at least since Boyle (1977) to price European style derivatives, in general, and options, in particular. Simulation methods are flexible, very easy to apply, and they have nice

*The authors thank participants at the 2014 SIAM Conference on Financial Mathematics & Engineering and the 2016 Financial Management Association Annual Meeting and at seminars at IIT Stuart School of Business and University of Western Ontario 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.

properties since averages of independent random observations converge to expected values under very mild assumptions. However, developing efficient methods that are flexible enough to price options in realistic setting that yield not only prices but also price sensitivities remains an active area of research.

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 initial 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. Based on our theoretical developments, we provide general guidelines for selecting the polynomial order for the estimation of the price and the Greeks, the number of paths to use in the simulation, and the size of the initial state dispersion. Second, a new 2-step 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, is proposed. The quality of the estimates crucially depends on how state variables are initially dispersed, and our proposed 2-step method automatically makes that selection. The proposed method is straightforward to implement in practice, provides precise and unbiased estimates of the price and price sensitivities, and our numerical results show that the method works extremely well for very reasonable choices of spread sizes, regressors, and simulated paths and that it compares exceptionally well to existing alternatives.

The use of Monte Carlo methods for calculating the Greeks for options with early exercise is much less explored than for the European versions.[1] However, given an early exercise strategy calculating the Greeks for an American option is no more complicated than it is for the European style option. For example, one could estimate the option Delta using the "finite difference approach" by simulating at two different values of the underlying stock price and estimating the option value using the same early exercise strategy, a viable method when the two starting values are close enough by a smoothness argument. A drawback of this method 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. For similar 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 if the optimal early exercise strategy is used, possible drawbacks are that they require that the payoff, respectively, the probability density can be differentiated, which may not always be the case.

Unlike the research mentioned above, this paper considers simulation methods that can be used to jointly determine prices and sensitivities. There are several reasons for this. First, by simultaneously estimating the stopping time strategy the issue of non-continuity in the state variables mentioned above 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 anal-

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

ysis of 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 argument about the actual performance of these methods that are generally and unconditionally applicable. Methods that jointly determine prices and price sensitivities, on the other hand, are easier to examine in terms of their numerical performance and convergence rates.

There are to our knowledge only a few papers that consider the problem of option pricing and estimation of the Greeks jointly for American options. In Feng, Liu, and Sun (2013) an algorithm for determining the Greeks iteratively along with the value function is proposed. The paper, though, offers little evidence on the usefulness of the proposed method and very limited numerical results. Jain and Oosterlee (2013) instead suggest that the Delta can be approximated using a finite difference approach, in which the regression coefficients from the first early exercise points are used. This method, however, requires that one uses a "regress-later" type approach (see Glasserman and Yu (2002)), and one therefore needs regressors that are martingales or for which the one step ahead conditional expectations are known in closed form or have analytical approximations.[2] This restriction significantly limits the choice of potential regressors, and for more complicated models it may be impossible to find regressors that satisfy the restriction and the proposed method would be infeasible. Finally, Chen and Liu (2014) considers jointly pricing options and estimating first order derivatives like the Delta using the pathwise derivative method mentioned above.

The method used in the current paper is, however, closest in spirit to Wang and Caflisch (2010) which suggests that the Greeks may be estimated by performing an additional regression using an initial dispersed sample of the state variables.[3] However, 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-step method which combines the flexible simulation method with an adaptive choice of optimal initial dispersed state variable to control and balance off the bias of the estimates against their variance.

Our numerical results document that our proposed method works extremely well for very reasonable choices of the number of regressors and simulated paths for a sample of options with diverse and empirically relevant characteristics. The practical guidelines we develop are simple to implement and the resulting method is robust to alternative choices of, for example, the method used to generate the initially dispersed state variables. We compare our proposed method to the pathwise derivative method and the likelihood ratio method as well as the method suggested by Wang and

[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 method is similar to starting the Binomial Model before the actual current time, the so-called extended tree, to obtain Greeks from this method (see Pelsser and Vorst (1994)).

Caflisch (2010). The results show that, among the methods that provide estimates of the prices and Greeks that are statistically insignificant from the benchmark values, 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 methods, is extremely flexible and easy to use, and the current paper thus successfully proposes a method that can be used for simultaneous estimation of prices and price sensitivities for American style options in very general setting. In particular, it is straightforward to obtain the Delta, the first derivative of the option price with respect to the underlying asset value, as well as the Gamma, the corresponding second derivative, and it is simple to apply our proposed method to the multivariate case where sensitivities to each of the underlying assets along with cross-sensitivities should be estimated. More generally, sensitivities towards all the stochastic or time varying factors determining an option's price, which besides the value of the underlying asset are at least the volatility of this asset, the interest rate, and potentially also the dividend yield, are needed. Our proposed methodology can be applied to obtain price sensitivities with respect to any of these state variables and should therefore be of immediate interest and have broad applications.

The rest of the paper is structured as follows: Section 2 provides the theoretical framework for how American options can be priced using simulation, discusses the idea behind obtaining Greeks using initial state dispersion, and proves convergence of our proposed method. Section 3 discusses several issues related to the implementation of the methodology, proposes a simple 2-step method which improves on the shortcomings of the naive estimator, and documents that this method provides precise and unbiased estimates of the price, Delta, and Gamma. Section 4 provides additional numerical results and robustness checks. Finally, Section 5 offers some concluding remarks. The Appendix contains proofs and additional technical details.

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 essentially 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 $(\Omega, \mathcal{F}, \mathbb{P})$ 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 $\mathbb{P}$. 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 $Z\left(t_j\right) = \pi\left(X\left(t_j\right), t_j\right)$

[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).

for a suitable function $\pi(\cdot,\cdot)$ 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

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

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

The problem of calculating the option price in (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 will be to exercise immediately if the value of this is positive and larger than the expected payoff from holding the option until the next period and behaving optimally onwards. Let $V(X(t_j))$ denote the value of the option for state variables X at a time t_j prior to expiration and define $F(X(t_j)) \equiv \mathrm{E}[Z(\tau(t_{j+1}))\,|X(t_j)]$ as the expected conditional payoff, where $\tau(t_{j+1})$ is the optimal stopping time. It then follows that

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

and the optimal stopping time can be derived iteratively as

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

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

$$P(X(0)=x) = \mathrm{E}\left[Z(\tau(t_1))\,|X(0)=x\right]. \tag{4}$$

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

2.1 Simulation and regression methods

The idea behind using simulation for option pricing is quite simple and involves estimating expected values and therefore option prices by an average of a number of random draws. However, when the option is American, one needs to simultaneously determine the optimal early exercise strategy, and this complicates matters. In particular, it is generally not possible to implement the exact algorithm in (3) because the conditional expectations are unknown and therefore the price estimate in (4) is infeasible. Instead an approximate algorithm is needed. Because conditional expectations can be represented as a countable linear combination of basis functions we may write $F(X(t_j)) =$

$\sum_{m=0}^{\infty} \phi_m (X(t_j)) c_m (t_j)$, 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 $F(X(t_j)) \approx F_M(X(t_j)) = \sum_{m=0}^{M} \phi_m(X(t_j)) c_m(t_j)$ and that we can obtain an estimate of this function by

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

where the coefficients $\hat{c}_m^N(t_j)$ are approximated or estimated 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 a cross-sectional regression of the discounted future pathwise payoff on transformations of the state variables.

Based on the estimate in (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 (6) a natural estimate of the option value in (4) is given by

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

In the special case when all the paths are started at the current values of the state variables, i.e. $X(0) = x$, the conditional expectation in (7) can be estimated by the sample average given by

$$\hat{P}_M^N(X(0) = x) = \frac{1}{N} \sum_{n=1}^{N} Z(n, \hat{\tau}_M^N(1, n)), \tag{8}$$

where $Z(n, \hat{\tau}_M^N(1, n))$ 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 (6). Convergence of this type of estimates has been analyzed in detail in the existing literature. The first step in doing so is to establish convergence of the estimated approximate conditional expectation function which is done in, e.g., 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 $\hat{F}_M^N(X(t_j))$ converges to $F(X(t_j))$ in probability for $j = 1, ..., J$.*

Proof. See Stentoft (2004). □

The result in Lemma 1 can now be combined with Proposition 1 of Stentoft (2004) to demon-

[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)).

strate that when all the simulated paths are started at the current values of the state variable, i.e. $X(0) = x$ then the estimate in (8) converges to the true price which establishes convergence of the LSM method in a general multiperiod setting.[6] Moreover, this type of algorithm has nice 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).

2.2 Prices and Greeks with Initial State Dispersion

While in practice the stock price is known our formulation allows for considering option pricing with different deterministically determined starting values at $t = 0$. Using initially dispersed state variables is key to our approach for determining option price sensitivities. Note that, because of the Markovian assumption it does not matter where we start the process, Lemma 1 continues to hold, and this does not change the determination of the pathwise optimal early exercise strategy at any of the potential early exercise points. In this case, however, each realization of $Z\left(n, \hat{\tau}_M^N(1,n)\right)$ corresponds to the payoff from exercising the option at the estimated optimal stopping time $\hat{\tau}_M^N(1,n)$ from (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 of using a simple average, a natural alternative approach is to use the simulated paths 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 that this can be well approximated using the first $M_0 + 1$ terms of the 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 can be determined from a cross-sectional regression of $N \geq M_0$ discounted future pathwise payoffs on transformations of the initially dispersed state variables. A natural 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

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

[6] One of the important assumptions in Stentoft (2004) is that the support is bounded. In Glasserman and Yu (2002) convergence is studied in the unbounded case. This complicates the analysis and therefore they limit their attention to the normal and lognormal cases. See also Gerhold (2011) for generalizations of the results to other processes and Belomestny (2011) for a proof using nonparametric local polynomial regressions.

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, i.e. the Delta or Δ, and the second derivative, i.e. the Gamma or Γ, at $S(0) = s$ are given by

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

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

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

respectively. Note that if the initial approximation $\hat{P}_{M,M_0}^{N}$ in (9) is a simple polynomial, the Greeks are particularly easy to calculate. However, while the use of polynomials is theoretically supported (see the next section), other types of approximation functions could be used and if analytical derivatives are difficult to obtain one can always use numerical differentiation to obtain the required sensitivities.

The method we outline above for estimating the price, Delta, and Gamma is very intuitive and it seems quite reasonable that it should work well and that the estimates should converge. However, while the convergence of the LSM method is well established the proof of this generally relies on assuming that the price is estimated as an average of pathwise payoffs according to (8). The fact that $\hat{F}_M^N(X(t_j))$ converges to $F(X(t_j))$, for $t_j > 0$, does by itself not necessarily imply that the estimated quantities in (12), (13), and (14) converge to the true price, Delta, and Gamma. In particular, what is used to calculate the price and Greeks at time $t_j = 0$ is the actual estimated approximation when using initial state dispersion (ISD). In the rest of the LSM algorithm, the estimated approximation is only used to decide whether to exercise or not and the actual cash flows used for valuation are those realized along a particular path. Because of this, more care has to be taken when analysing the properties of the estimated price and Greeks.

2.3 Convergence of the price and Greeks estimator

We now provide a formal setting for how to estimate the price and the Greeks and we prove that these estimates converge. To fix notation, suppose we want to estimate an unknown regression function $m(x) = \mathrm{E}[Y|X=x]$ using an i.i.d. random sample of data (X_1, Y_1), ..., (X_N, Y_N) from the population (X, Y). Using a Taylor series expansion we can approximate $m(x)$ around a point x_0 by

$$m(x) \simeq m(x_0) + m^{(1)}(x_0)(x-x_0) + \frac{1}{2} m^{(2)}(x_0)(x-x_0)^2 + \cdots + \frac{1}{M!} m^{(M)}(x_0)(x-x_0)^M \tag{15}$$

$$= m(x_0) + \beta_1(x-x_0) + \beta_2(x-x_0)^2 + \cdots + \beta_M(x-x_0)^M, \tag{16}$$

where $m^{(i)}$, the ith derivative of $m(\cdot)$, is assumed to exist and M is the order of the polynomial used to approximate the true function. This is a Local Polynomial Regression (LPR) problem in which the β_i coefficients solve the following weighted least squares regression

$$\min_{\beta_i} \sum_{n=1}^{N} \{Y_n - \sum_{i=0}^{M} \beta_i (X_n - x_0)^i\}^2 \mathcal{K}_h (X_n - x_0), \tag{17}$$

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.

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 (17) 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)) and thus is applicable to approximating (9) in general. We can use LPR convergence results to establish the convergence of the price and Greeks estimators in (12), (13), and (14). For this we need the following well known lemma from the literature on LPR.

Lemma 2 (Adapted from Theorem 3.1 of Fan and Gijbels (1996))**.** *In the context of the LPR problem in* (17)*, let f_X denote the density of X and let $\sigma^2(x)$ denote the conditional variance of Y given $X = x$. Assume that $f_X(x_0) > 0$, $f_X^{'}(x_0) = 0$, and that $f(\cdot)$, $\beta_{(M+1)}(\cdot)$, 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 is given by*

$$Var\{\hat{\beta}_i|X\} \quad = \quad \xi_1 \frac{i!^2 \sigma^2(x_0)}{f(x_0) Nh^{1+2i}} + o_P\left(\frac{1}{Nh^{1+2i}}\right), \tag{18}$$

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

$$Bias\{\hat{\beta}_i|X\} \quad = \quad \xi_2 i! \beta_{(M+1)}(x_0) h^{M+1-i} + o_P\left(h^{M+1-i}\right). \tag{19}$$

Further, for $(M-i)$ even the asymptotic conditional bias is given by

$$Bias\{\hat{\beta}_i|X\} \quad = \quad \xi_3 i! \beta_{(M+2)}(x_0) h^{M+2-i} + o_P\left(h^{M+2-i}\right), \tag{20}$$

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

Proof. This follows from Fan and Gijbels (1996) assuming explicitly that the derivative of the density $f_X^{'}(x_0) = 0$ and by expressing everything in terms of the β_i's.[7] $\square$

[7] It is possible to relax the assumption that $f_X^{'}(x_0) = 0$ without affecting the general results though this introduces additional terms in (20).

To cast our problem of estimating the price and the price sensitivities in the framework of LPR we first note that at $t = 0$ given a stopping time $\tau^*(1)$ the price function P^* is indeed a conditional expectation or regression function similar to m above. Moreover, using simulation we can obtain a sample of N realizations of discounted payoffs $Z(n, \tau^*(1, n))$, which we may simply refer to as the Y_n, for paths started at different initial values of the state variable X_n. Next, what we are after are the actual derivatives of this function and not the coefficients in the LPR regression. However, when the basis used corresponds to polynomials centered at x_0 these two are identical. Finally, note that using ordinary least squares (OLS) regression, which gives equal weight to all observations, corresponds to using a uniform kernel, given by

$$\mathcal{K}_h(X_n - x_0) = \frac{1}{2h}\mathbf{1}\left(\left|\frac{X_n - x_0}{h}\right| \leq 1\right), \tag{21}$$

in the LPR setting.

With a slight abuse of notation we now propose to estimate directly the price and price sensitivities as the coefficients that solve the following OLS problem

$$\min_{P^{*(i)}} \sum_{n=1}^{N} \{Y_n - \sum_{i=0}^{M_0} P^{*(i)} (X_n - x_0)^i\}^2. \tag{22}$$

We now state the relevant assumptions under which convergence of the price and price sensitivities from (22) 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*

$$X_n = x_0 + \alpha \mathcal{K}_{ISD}(U_n), \tag{23}$$

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

Assumption 2. *The pathwise payoffs, Y, 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*

$$Y_n = Z_n(\tau^*(1, n)), \tag{24}$$

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

Assumption 3. *The price function given $\tau^*(1)$, denoted by $P^* = \mathrm{E}[Y|X]$, is continuously differentiable of at least order $M_0 + 2$ on the support of X.*

Theorem 1. *Under Assumptions 1, 2 and 3, let $\alpha \to 0$ and $N \to \infty$ such that $N\alpha^3 \to \infty$ then the estimated i'th derivative of the option value function, the price sensitivity $\hat{P}^{*(i)}$ from (22), is asymptotically unbiased for $i = 0, 1, ..., M_0$. Moreover, if $N\alpha^{1+2i} \to \infty$ then the estimated price sensitivity, $\hat{P}^{*(i)}$, converges to the price sensitivity, $P^{*(i)}$, in mean square sense for $i = 0, 1, ..., M_0$.*

Proof. See Appendix A. □

Remark 1. *Theorem 1 demonstrates that it is essential for convergence of the estimators that the size of the ISD, i.e. the α, tends to zero and the number of observations, the N, tends to infinity. 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. Finally, note that α can not tend to zero "too" fast when estimating higher order derivatives since otherwise the variance of the estimates will be unbounded.*

Theorem 1 holds for any estimated stopping time $\hat{\tau}(1)$ and we can now combine it with the results of, e.g., Lemma 1, which shows that $\hat{F}_M^N(X(t_j))$ converges to $F(X(t_j))$, for $t_j > 0$, and that therefore the implied stopping time converges to the "true" stopping time, to conclude that our estimates converges to the true price and price sensitivities. A similar argument is used in Chen and Liu (2014). We state this result, without proof, in the following corollary to Theorem 1.

Corollary 1. *Under the setup given above, assume that $N \to \infty$, that $M \to \infty$ such that $M^3/N \to 0$, and let $\alpha \to 0$ such that $N\alpha^{1+2i} \to \infty$. Then the estimated i'th derivative of the option value function, the price sensitivity $\hat{P}^{(i)}$ from (22), converges to the true price sensitivity $P^{(i)}$.*

Remark 2. *It should be noted that the conditions in Corrolary 1, in particular the requirement that $N\alpha^{1+2i} \to \infty$, are sufficient to ensure convergence. However, it may be possible to relax these conditions as they may not be necessary. In particular, the context outlined above ensures that all N paths are used in the regression even when α tends to zero. In the regular LPR setting, on the other hand, a larger and larger fraction of the paths ends up not being used in the regression as the bandwith h decreases.*[8]

2.4 Discussion

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 2 has important implications when it comes to implementing the methodology numerically and, in particular, on 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 state variables. From this we can draw several conclusions which will also help us develop our proposed 2-step methodology in the following section.

First of all, 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

[8]Unreported simulation results document that the convergence rates may in fact be significantly faster then those established in Lemma 2 for the problem at hand.

and reducing the variance of the estimates. Finally, note that Fan and Gijbels (1995a) show that increasing $M_0 - i$ from even to odd reduces the bias while this does not affect 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.[9]

Next, the theoretical results have important implications when choosing the size of the ISD, i.e. by how much to initial disperse the state variables around x_0. In particular, when we generate the initial stock prices as specified in (23) 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 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 value of 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.

Finally, 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, $f(x_0)$. In particular, the theoretical results show that the asymptotic variance is inversely proportional to $f(x_0)$. 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

$$\mathcal{K}_{ISD}(U) = 2 \times \sin\left(\operatorname{asin}(2 \times U - 1)/3\right), \tag{25}$$

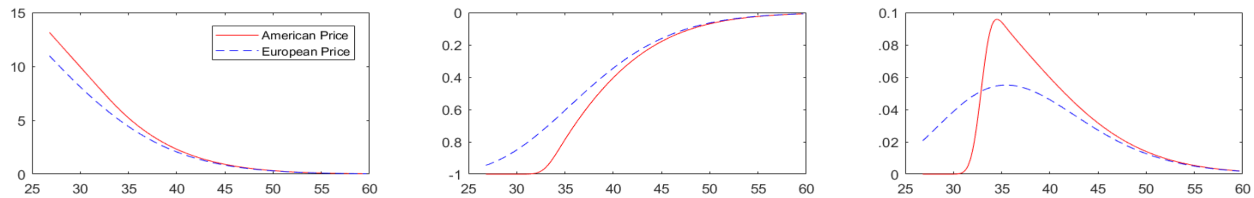
where U is a vector of size N uniformly distributed on the unit interval. This results in an Epanechnikov distribution for the ISD, which is bounded, and indeed ensures that the distribution is symmetric around x_0 and peaked at this value.

3 Implementation

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 start by plotting 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.

In the rest of this section, we consider the case in which the optimal stopping time is estimated

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



(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).

Table 1: 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 (23) 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 $M_\tau = 9$ 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.

using the LSM methodology with $N = 100,000$ simulated paths and a polynomial of order $M_\tau = 9$ in the cross-sectional regression using only paths that are in the money. 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. We first consider the results from a naive method using a polynomial of order $M_0 = 9$ at time $t = 0$ and, building on this method's performance and inspired by the theoretical results in the previous section, we develop a 2-step procedure that corrects its shortcomings.

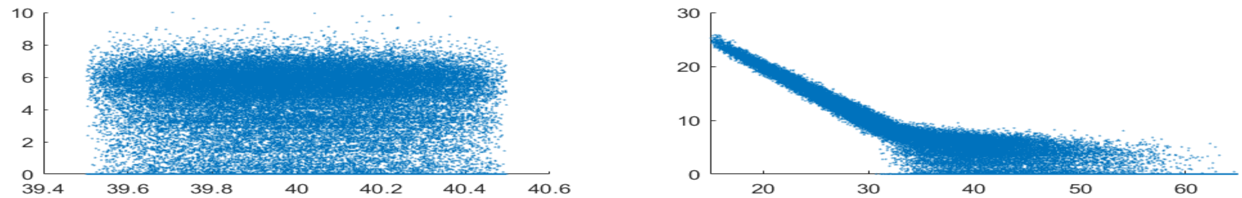
3.1 Results with the naive method

The simplest possible and "naive" method is implemented as follows: First, create N initially dispersed stock prices X using (23) and simulate a stock price path from each of these. Next, determine the pathwise payoffs Y from (24) using the estimated optimal stopping time from the LSM algorithm. Finally, regress Y on X as in (22) to approximate the price function and estimate the price and Greeks by evaluating this function and its derivatives at the current level of the stock price. Table 1 shows the results for different choices of 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 the estimates of all 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 though they may be statistically unbiased, whereas too large an initial spread leads to estimates with a large bias but with small variance and therefore statistically biased. This trade-off of bias against uncertainty makes it very difficult to provide general recommendations about the appropriate choice of the ISD.

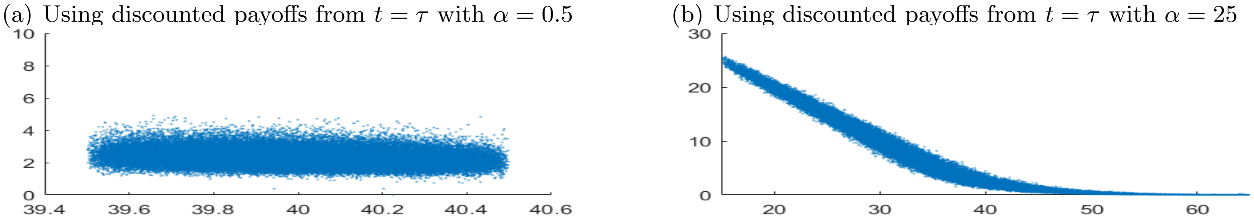
Given the above findings, the only possible general recommendation may appear to be to use a small ISD to get unbiased estimates, and to find a variance reduction technique that significantly reduces the variance of the Greek estimates. To motivate our method, we first illustrate in more detail the problem 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 2 show examples of data available at $t = 0$ for the regression. When a small ISD is used as in Sub-figure 2(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 2(b) the data, on the other hand, displays some structure and when a 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.

3.2 Variance reduction by imposing structure

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 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 paths 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 $t = 0$. Figures 2(c) and 2(d) show the sample data when



(a) Using discounted payoffs from $t = \tau$ with $\alpha = 0.5$ (b) Using discounted payoffs from $t = \tau$ with $\alpha = 25$



(c) Using the value function from $t = 1$ with $\alpha = 0.5$ (d) Using the value function from $t = 1$ with $\alpha = 25$

Figure 2: 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 (23). 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.

Table 2: 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 (23) with a method that uses a value function iteration in the second to last step. See also the notes to Table 1.

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

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.[10] Our proposed method essentially reduces the conditional variance $\sigma^2(x)$ and turns out to be particularly effective for higher order Greeks

Table 2 shows the result for the different choices of α using this "value function" method. The table, first of all, shows that the method indeed produces estimated prices and Greeks with lower standard deviation in particular 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

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

Table 3: 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 (23) 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 1.

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

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 2 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 $\alpha = 0.5$. 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.

3.3 Bias reduction with optimal bandwidth selection

The method proposed above corrects one of the shortcomings of the naive simulation-based method for obtaining Greeks 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 (23) is needed. Using LPR theory we can obtain such an α^* and in Appendix A we propose a simple algorithm for obtaining the optimal bandwidth to be used. Once α^* is obtained we 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.

Table 3 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

Table 4: Results with the 2-step 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 (23) with our proposed 2-step 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 1.

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

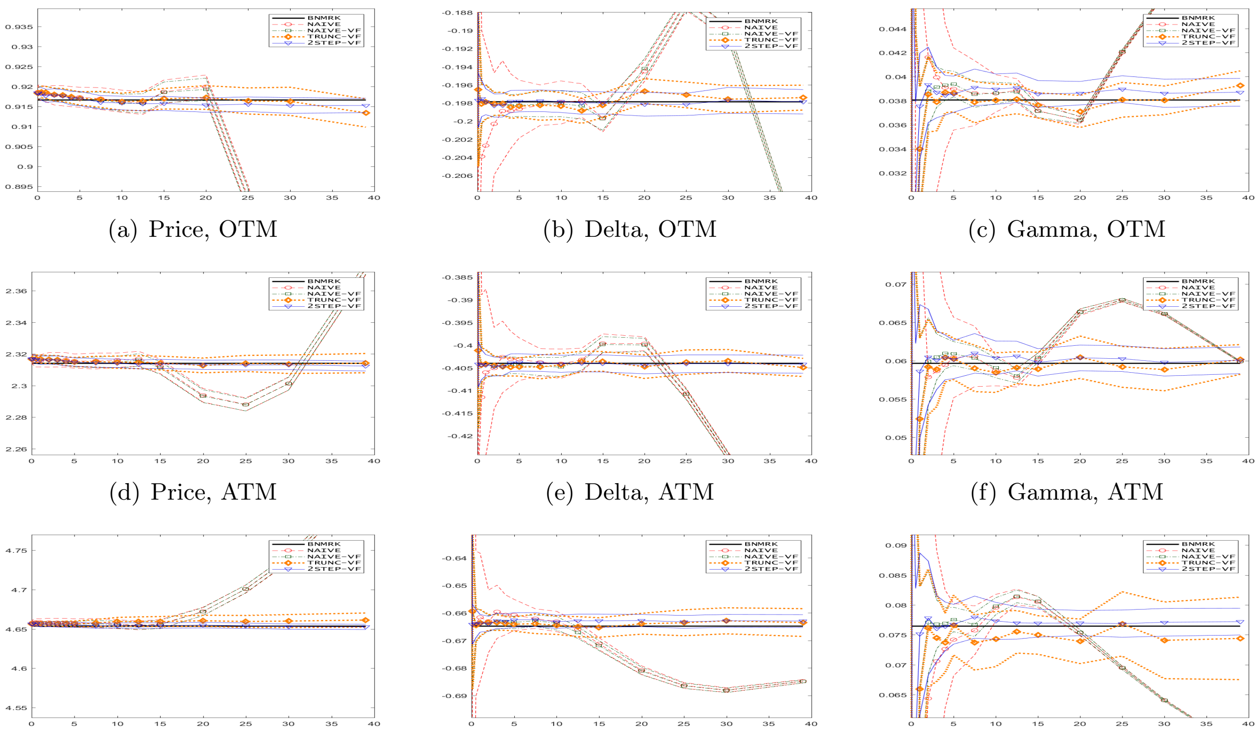
in the standard deviation of the estimate as a large fraction of the paths are potentially truncated. Note that using this method to determine the optimal α also decreases the bias for higher order Greeks when using a small ISD though not to the same extent as when using a large ISD.

3.4 An improved 2-step rescaled method

The major drawback of the method that truncates paths is that this leads to increased variability in the estimated quantities. To fix this, 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 step. In our setup, rescaling is possible and can be achieved by simply dividing the simulated stock prices along a given path by an appropriately chosen constant. Table 4 shows the results for the different choices of α using our proposed "2-step" method. The table, first of all, shows that this method indeed performs very well. In particular, the 2-step 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 for even higher order Greeks. Across the choices of initial values of α the table shows that the 2-step 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-step procedure will correct for choosing it too extreme and still produce unbiased estimates with reasonable standard deviation.

Based on these findings, we propose that option prices and price sensitivities should be estimated with the following algorithm:

1. Create a sample of size N of initially dispersed stock prices X with a given α in (23) and simulate a stock price path from each of these. Determine the pathwise payoffs Y using the estimated optimal stopping time determined with the LSM method in (6) using cross sectional



(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 3: Results with estimated early exercise strategy

This figure plots estimated prices and Greeks along with 99% confidence intervals across different values of α in (23) for the three options and four methods considered in this section. Results are based on 100 independent simulations with $N = 100,000$ paths. The 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 $M_\tau = 9$.

regressions with $M_\tau + 1$ regressors. The naive method corresponds to using a simple OLS regression at time $t = 0$ of Y on X according to (22) with $M_0 + 1$ terms to determine the prices and Greeks.

2. Use the value function estimated at $t = 1$ to "smooth" the payoffs at time $t = 1$ and discount these back to time $t = 0$. The value function method corresponds to using these discounted payoffs as the Y in the regression at time $t = 0$.
3. Use the optimal bandwidth selector from Appendix A to determine α^* based on the above smoothed Y and X. The truncated method corresponds to using only the paths that are within the optimal bandwith in the regression at time $t = 0$.
4. Rescale the initially dispersed state variables to all lie within the optimal bandwith, apply the estimated optimal stopping time to these paths, and use the value function method at $t = 1$ to smooth the pathwise discounted payoffs from these new paths. Our proposed 2-step method corresponds to using all these paths in the regression at time $t = 0$.

In the spirit that a picture, or in our case a figure, is worth a thousand words, or here estimates, Figure 3 shows the results for the three options considered in this section across various choices of initial ISD size. The figure clearly shows that as the size of the initial spread increases the naive method, labelled "NAIVE" in the plots, and even the method that uses the value function method, labelled “NAIVE-VF” in the plots, breaks down, and only methods that rely on results with optimal bandwidth selection, labelled “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 methods 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-step method indicating that the choice of initial α is essentially inconsequential for the performance of the method and demonstrating that this method works well irrespective of the initial α used.

4 Results

The previous section develops our 2-step methodology and provides strong support for it. In this section, we test the robustness of this finding along several dimensions. First, we consider a much large range of options then in Section 3 with characteristics of empirical relevance. Next, we consider the robustness of the algorithm to different choices of the number of simulated paths and the number of regressors used at time $t = 0$, as well as to using alternative polynomial orders to determine the optimal stopping time. Then, we consider robustness to the choice of the ISD and the method for determining the optimal α. Finally, we compare our suggested methodology to some other existing methods for estimating the Greeks like the pathwise derivative method and the likelihood ratio method.

4.1 Robustness across option characteristics

We first report results for a wider range of options with strike prices of $K \in \{36, 40, 44\}$, volatilities of $\sigma \in \{10\%, 20\%, 40\%\}$, and maturities of $T \in \{0.5, 1, 2\}$ years, for a total of 27 different configurations. 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 $\alpha = 10$. We maintain the assumption of Black-Scholes-Merton style dynamics for ease of calculating benchmark values. However, there is nothing in our methodology that is restrictive to this setting and, since we are using simple simulations, our approach can be generalized to other dynamics and could even be used to get other types of Greeks as long as the underlying risk factors can be simulated.

For the large sample of options considered here, Table 5 shows that our 2-step 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. The supplementary document accompanying this paper demonstrates that this finding is not particular to the choice of initial α and our proposed method works irrespective of the choice of initial α.

Table 5: 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 $M_\tau = 9$ 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.

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 6 which allows us to conclude that our method works well across various levels of interest rates and dividend yields. The only case in which the method produces marginally significantly different estimates occurs when both the interest rate and dividend yield are zero. For these options the t-statistics for the prices are 1.99, 1.83, and 2.13, respectively. In this case, however, it is known that the option should never be exercised. However, the simulation algorithm at times will choose to exercise, essentially because there is a slight degree of over fitting, and this generates the significant differences.[11] Note

[11]The slight over fitting 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 6: Results across interest rates and dividend yields

K	r	d	Price BM	Price Estim	Price StDev	Delta BM	Delta Estim	Delta StDev	Gamma BM	Gamma Estim	Gamma 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 $M_\tau = 9$ 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.

that this only affects the prices whereas the Greeks continue to be estimated insignificantly different from their true values.

4.2 Robustness across choices of regressors and number of paths

The results presented above were based on using $M_0 = 9$ regressors in the initial regression, using $N = 100,000$ simulated paths, and using $M_\tau = 9$ 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 7.

Table 7 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 $t = 0$.

The results in Table 7 also confirm the intuition from Theorem 1. In particular, the table clearly

Table 7: 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 $M_\tau = 9$.

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

documents that in almost all cases the bias of the estimated prices and sensitivities decrease when the polynomial order M_0 increases. Moreover, the table shows that the choice of N clearly affects the standard deviation of the estimates and as expected these decrease with increased number of paths. Combined the results demonstrate that as the number of simulated paths and for a given order of the polynomial used in the regression at time $t = 0$ the estimates converge as expected.

We next consider the effect of using different orders of the polynomial used to determine the optimal early exercise strategy, M_τ. 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_τ, $M_\tau = 5$ and $M_\tau = 15$. The results are shown in Table 8 which documents that the choice of the order of the polynomial

Table 8: Results across stopping time polynomial order M_τ

			Price			Delta			Gamma	
K	M_τ	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 used to estimate the optimal stopping time, M_τ. 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.

used to determine the optimal early exercise strategy in fact has surprisingly little influence on the estimates obtained with our proposed 2-step method.

4.3 Robustness to the choice of ISD and optimal α

The results presented above were based on using a particular 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. This section shows that our method is also robust to these fundamental choices of how to implement the proposed methodology. 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 (25) is instead the Uniform distribution. We also consider the case where the ISD is generated from a random uniformly distributed vector U in (25).

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 9 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 9 does indicate that a deterministic and peaked kernel is preferred.

Table 9: Results across ISD 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 $M_\tau = 9$ 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.

Table 9 also documents a clear effect of choosing the optimal α differently. For example, 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 $\alpha = 0$, i.e. 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 of the value function which is essentially a straight line. 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.

4.4 Comparison with some competing methods

In the Introduction we mention other possible methods that could be used to obtain price sensitivities for options using Monte Carlo simulation and we now compare our results to some of these.

Table 10: Comparison with alternative models

		Price			Delta			Gamma		
K	Model	BM	Estim	StDev	BM	Estim	StDev	BM	Estim	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-step	0.9166	0.9159	(0.0067)	-0.1979	-0.1978	(0.0047)	0.0381	0.0390	(0.0049)
40	2-step	2.3141	2.3141	(0.0102)	-0.4040	-0.4041	(0.0073)	0.0597	0.0604	(0.0085)
44	2-step	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 for the three options from Section 3 when using 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 $M_\tau = 5$ to determine the optimal early exercise strategy, except in the 2-step method where $M_\tau = 9$. 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) in which a polynomial of order $M_0 = 5$ is used to estimate the price and Greeks. Results in the rows headed 2-step correspond to our proposed method implemented with an initial $\alpha = 10$ in which a polynomial of order $M_0 = 9$ is used to estimate the price and Greeks.

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

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 e.g. 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 sensitivity estimators though the estimates 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-step method in Table 10.[12]

Table 10 shows that the PDM and LRM methods also produce estimated prices and Greeks that are very close to the benchmark values when they exist. 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, however, our proposed 2-step method always has the smallest bias. The biggest differences though 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.

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

The standard deviations of the estimated Deltas with our proposed 2-step 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 LRM and our 2-step 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 these results clearly demonstrate the value of our suggested method.

5 Conclusion

Simulation methods are important for option pricing because of their flexibility and efficient algorithms now exist for pricing 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 options with American style features because of the need to simultaneously determine the optimal early exercise strategy which significantly complicates matters.

This paper proposes a new method which combines flexible simulation methods with initially dispersed state variables to jointly estimate prices and price sensitivities of American style options. 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. Based on our theoretical developments, we provide general guidelines for selecting the polynomial order used in the estimation of the price and the Greeks, the number of paths to use in the simulation, and the size of the initial state dispersion. Second, a new numerical method, which combines flexible simulation methods with an adaptive choice of optimal initially dispersed state variables to control and balance off the bias of the estimates against their variance, is proposed. The quality of the estimates crucially depends on how state variables are initially dispersed, and our proposed 2-step method automatically makes that selection.

A large numerical exercise shows that the method works extremely well for very reasonable choices of the number of regressors and simulated paths for a sample of options with diverse and empirically relevant characteristics. The practical guidelines we develop are simple to implement and the resulting method is robust to alternative choices of, for example, the method used to generate the initially dispersed state variables. The results also show that our proposed method is less biased, more precise, and more flexible than alternative methods that can be used to obtain relevant price sensitivities. Since the method we propose relies on nothing but simple polynomial approximations using simulated paths 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 price sensitivities for American style options in very general settings using simple regression-based simulation methods.

References

Belomestny, D. (2011): "Pricing Bermudan Options by Nonparametric Regression: Optimal Rates of Convergence for Lower Estimates," *Finance and Stochastics*, 15(4), 655–683.

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.

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.

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

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.

Gerhold, S. (2011): "The Longstaff–Schwartz Algorithm for Lévy Models: Results on Fast and Slow Convergence," *Annals of Applied Probability*, 21(2), 589–608.

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.

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

JAIN, S., AND C. W. OOSTERLEE (2013): "The Stochastic Grid Bundling Method: Efficient Pricing of Bermudan Options and their Greeks," *Working paper.*

KANIEL, R., S. TOMPAIDIS, AND A. ZEMLIANO (2008): "Efficient Computation of Hedging Parameters for Discretely Exercisable Options," *Operations Research*, 56(4), 811–826.

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.

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.

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.

——— (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 Proofs and technical details

This Appendix contains the proof of Theorem 1 and technical details on how to implement the proposed algorithm for optimally selecting α.

A.1 Proof of convergence

Proof of Theorem 1. First, note that though Lemma 2 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). Next, 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 $f(x_0) > 0$, $f'(x_0) = 0$, and $f(x_0)$ is continuous in a neighborhood of x_0.
2. By Assumption 2 and the properties of the payoffs $\sigma^2(\cdot)$ exists, is bounded and continuous in a neighborhood of x_0.
3. By Assumption 3 β_{M+2}, the $M+2$'th derivative of the regression function P, exists and it is continuous in a neighborhood of x_0.

Finally, letting $\alpha \to 0$ and assuming a uniform kernel with $h = \max\left[X_{(N)} - x_0,\ x_0 - X_{(1)}\right]$, where $X_{(i)}$ denotes the ith order statistic, it follows that $h \to 0$, and thus all the assumptions of Lemma 2 are satisfied. Since the i'th derivative of the option value function corresponds to the coefficient β_i Lemma 2 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 concludes the proof. $\square$

A.2 Optimal bandwidth selection

As it is the case with most non-parametric methods the choice of bandwidth is crucial for the LPR methodology, and while there is a vast literature on bandwidth selection for methods such as the LPR, many of these methods are not suitable in our setting where the data results from the LSM method and could involve hundreds of thousands of data points. In particular, cross-validation methods which require N regressions (one regression per path) are very computer intensive, and instead, we need to resort to more direct and simpler "rule of thumb" based methods. We now propose a simple solution that works well in our setting.

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, instead of a global, bandwidth selector. Fan and Gijbels (1995b) provide a theoretically optimal local bandwidth formulation which is easy to approximate. It is given by

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

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 variable 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. Fan and Gijbels (1995b) provide means to estimate these unknown values and hence a way to implement the bandwidth selector in practice.

In our setting, we have an important advantage because the distribution of X is known since we specify 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. To do so, 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.[13] Here we propose to estimate these quantities using a simple Rule of Thumb (ROT) method for global bandwidth selection.

The ROT estimator we consider originally comes from Fan and Gijbels (1992) and is given by

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

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

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

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

This ROT is a very crude approximation that is often used as a pilot bandwidth for more sophisticated bandwidth selectors. In fact, we could use the global bandwidth directly in the simulation method, but results are not as robust as when a local bandwidth is used. Thus, our suggested method proceeds as follows: First, obtain an approximation of the optimal global bandwidth using the ROT in (A.2). Second, use the ROT to locally fit a polynomial and estimate $\sigma^2(x_0)$ and β_{M+1}. Finally, plug those values in (A.1) to obtain an approximation of the optimal local bandwidth. This procedure is quick, simple, and in our application provides good results.

The pricing method we propose uses the optimal bandwidth selector from above in a first step to estimate the optimal ISD size and then applies the simulation methodology to estimate the price and price sensitivities using the estimated optimal ISD. However, the bandwidth selection method used to find the optimal ISD size does not guarantee the optimal ISD size is found after just one iteration and one might consider iterating multiple times. Our results nevertheless show that when using a reasonable large number of regressors, say $M_0 = 9$, and initial ISD, say $\alpha = 10$, our bandwidth selector converges very fast in as little as one or two iterations.

[13] If we knew the optimal local bandwidth this could have been used, but then we would not need to determine it!