On the consistency of regression based Monte Carlo methods for pricing Bermudan options in case of estimated financial models

Andreas Fromkorth and Michael Kohler*

Department of Mathematics, Technische Universität Darmstadt, Schlossgartenstr. 7, D-64289 Darmstadt, Germany, email: fromkorth@mathematik.tu-darmstadt.de, kohler@mathematik.tu-darmstadt.de

February 22, 2012

Abstract

In many applications of regression based Monte Carlo methods for pricing American options in discrete time parameters of the underlying financial model have to be estimated from observed data. In this paper suitably defined nonparametric regression based Monte Carlo methods are applied to paths of financial models where the parameters converge towards true values of the parameters. For various Black-Scholes, Garch and Levy models it is shown that in this case the price estimated from the approximate model converge to the true price.

AMS classification: Primary 65C05, 93E24; secondary 60G40.

Key words and phrases: American options, consistency, least squares estimates, nonparametric regression, robustness, regression based Monte Carlo methods.

1 Introduction

In this paper we study the problem of numerical evaluation of an American option in discrete time (also called Bermudan option). The holder of such an option has the right to buy or sell the underlying asset for a given strike price at one of the time points $0, 1, \ldots, T$,

*Corresponding author. Tel: +49-6151-16-6846, Fax: +49-6151-16-6822

Running title: *Pricing Bermudan options in case of estimated financial models*

where T is the so-called majurity of the option. It is well-known that in complete and arbitrage free markets the price V_0 of such an option is given by a solution of the optimal stopping problem

$$V_0 = \sup_{\tau \in \mathcal{T}(0,\ldots,T)} \mathbf{E}\left\{ f_\tau\left(X_\tau\right)\right\} \tag{1}$$

(cf., e.g., Karatzas and Shreve (1998)). Here f_t is the discounted payoff function, the underlying stochastic process is given by $X_0, X_1, \ldots, X_T$, and $\mathcal{T}(0,\ldots,T)$ is the class of all $\{0,\ldots,T\}$-valued stopping times, i.e. $\tau \in \mathcal{T}(0,\ldots,T)$ is a measurable function of $X_0,\ldots,X_T$ satisfying

$$\{\tau = \alpha\} \in \mathcal{F}(X_0,\ldots,X_\alpha) \text{ for all } \alpha \in \{0,\ldots,T\}.$$

Throughout this paper we assume $X_0 = x_0$ *a.s.* for some $x_0 \in \mathbb{R}^d$, i.e., we start at time zero with some fixed value. Furthermore we assume that X_0, X_1, $\ldots$, X_T is a $\mathbb{R}^d$–valued Markov process recording all necessary information about financial variables including prices of the underlying assets as well as additional risk factors driving stochastic volatility or stochastic interest rates. Neither the Markov property nor the form of the payoff as a function of the state X_t are very restrictive and can often be achieved by including supplementary variables.

One way to compute (1) is to determine an optimal stopping rule $\tau^* \in \mathcal{T}(0,\ldots,T)$ satisfying

$$V_0 = \mathbf{E}\left\{ f_{\tau^*}\left(X_{\tau^*}\right)\right\}. \tag{2}$$

Let

$$q_t(x) = \sup_{\tau \in \mathcal{T}(t+1,\ldots,T)} \mathbf{E}\left\{ f_\tau\left(X_\tau\right) \middle| X_t = x\right\} \tag{3}$$

be the so-called continuation value describing the value of the option at time t given $X_t = x$ in case of holding the option rather than excersising it. It follows from the general theory of optimal stopping (cf., e.g., Chow, Robbins and Siegmund (1971) or Shiryayev (1978)) that an optimal stopping rule can be defined by

$$\tau^* = \inf\{s \geq 0 : q_s(X_s) \leq f_s(X_s)\} \tag{4}$$

(cf., e.g., Section 8.1 in Glasserman (2004) or Theorem 1 in Kohler (2010)).

The Markov property implies the dynamic programming equations

$$\begin{aligned} q_T(x) &= 0, \\ q_t(x) &= \mathbf{E}\left\{ \max\left\{ f_{t+1}\left(X_{t+1}\right), q_{t+1}\left(X_{t+1}\right)\right\} \middle| X_t = x \right\} \qquad (t = 0, 1, \ldots, T-1) \end{aligned} \tag{5}$$

(cf., e.g., Section 8.1 in Glasserman (2004) or Theorem 2 in Kohler (2010)). In general these conditional expectations cannot be computed in applications. The basic idea of regression-based Monte Carlo methods for pricing American options is to apply recursively regression estimates to artificially created samples of

$$\left(X_t, \max\left\{ f_{t+1}\left(X_{t+1}\right), \hat{q}_{n,t+1}\left(X_{t+1}\right)\right\}\right) \tag{6}$$

to construct estimates $\hat{q}_{n,t}$ of q_t. This kind of recursive estimation scheme was firstly proposed by Carrier (1996) for the estimation of so-called value functions. In Tsitsiklis and Van Roy (1999) and Longstaff and Schwartz (2001) it was used in connection with parametric regression to construct estimates of continuation values. Various nonparametric regression estimates have been applied for the estimation of continuation values in Egloff (2005), Egloff, Kohler and Todorovic (2007), Kohler (2008), Belomestny (2011), Kohler and Krzyżak (2009) and Kohler, Krzyżak and Todorovic (2010). There results concerning consistency and rate of convergence of the resulting estimates of the price of the option have been derived.

From the theoretical point of view there is still one important problem. In applications the distribution of the stock values is unknown. Usually in practice one considers a stochastic model for the stock values (e.g., a Black-Scholes-model), estimates the model parameters (in this case the volatility of the underlying asset) and generates sample paths with this estimated distribution. Here it is assumed that the real model is known, but the real model parameters are unknown. So instead of (6) artificially generated samples of

$$\left(\bar{X}_t, \bar{Y}_t\right) = \left(\bar{X}_t, \max\left\{ f_{t+1}\left(\bar{X}_{t+1}\right), \hat{q}_{n,t+1}\left(\bar{X}_{t+1}\right)\right\}\right)$$

are given, where it is assumed that the distribution of $\bar{X}_t$ is close to the distribution of X_t in the sense that it is generated with the same model but slightly different values of the parameters.

In the sequel we investigate how the estimated price of the option behaves in case that the parameters of a given model converge to the true parameter values. Our main result will be that for suitably defined least squares estimates of the continuation values the estimated prices in case of a Black-Scholes model, of a GARCH-model or of a Levy model converges to the true price in this situation.

The outline of this article is as follows: The definition of the nonparametric regression based Monte Carlo methods which we analyze is given in Section 2. The main result concerning consistency of the estimate in case of an application with estimated parameters of a Black-Scholes, of a GARCH and of a Levy model is described in Section 3. In Section 4 the results are illustrated by simulated data. The proofs are given in Section 5.

2 Definition of the estimates

In the sequel we consider a $\mathbb{R}^d$-valued stochastic process

$$(X_t)_{t=0,\ldots,T}$$

containing the log prices of the underlyings and at least all informations needed to compute the payoff for arbitrary t. We denote the payoff function with respect to $(X_t)_{t=0,\ldots,T}$ at time t by f_t and the corresponding continuation value by q_t.

The consideration of log prices instead of ordinary prices simplifies the integrability condition of the sample paths, which we will need in our theoretical results.

Instead of $(X_t)_{t=0,\ldots,T}$ we have only given artificially generated samples of an estimate of $(X_t)_{t=0,\ldots,T}$ denoted by

$$\left(\bar{X}_{t,i}^{(n)}\right)_{t=0,\ldots,T}, \qquad i \in \mathbb{N}.$$

We will use these sample paths to estimate the continuation values q_t $(t = 0, \ldots, T)$. We start with

$$\hat{q}_{n,T} = 0. \tag{7}$$

Given the estimate $\hat{q}_{n,t+1}$ of q_{t+1} for some $t \in \{0, 1, \ldots, T-1\}$, we estimate the conditional expectation in (5) by applying the principle of least squares to the data

$$\left\{ \left(\bar{X}_{t,i}^{(n)}, \max\{f_{t+1}(\bar{X}_{t+1,i}^{(n)}), \hat{q}_{n,t+1}(\bar{X}_{t+1,i}^{(n)})\} \right) \quad : \quad i = 1, \ldots, n \right\}.$$

To do this, we choose a set $\mathcal{F}_n$ of functions $f : \mathbb{R}^d \to \mathbb{R}$ and define

$$\hat{q}_{n,t}(\cdot) = \arg \min_{g \in \mathcal{F}_n} \frac{1}{n} \sum_{i=1}^{n} \left| g\left(\bar{X}_{t,i}^{(n)}\right) - \max\{f_{t+1}(\bar{X}_{t+1,i}^{(n)}), \hat{q}_{n,t+1}(\bar{X}_{t+1,i}^{(n)})\} \right|^2,$$

where $x_0 = \arg\min_{x\in\mathcal{D}} h(x)$ means $x_0 \in \mathcal{D}$ and $h(x_0) = \min_{x\in\mathcal{D}} h(x)$ for a function $h : \mathcal{D} \to \mathbb{R}$. Here we assume for simplicity, that the minimum exists, but we do not require its uniqueness.

To compute the least squares estimate above, we have to specify suitable function sets $\mathcal{F}_n$. In the sequel we will use sets of polynomial spline functions.

Choose $M \in \mathbb{N}_0$, $K_n \in \mathbb{N}$, $A, B \in \mathbb{R}$ with $A < B$ and set $u_k = A + k \cdot (B - A)/K_n$ for $k \in \mathbb{Z}$. Let B_{j,M,K_n}, $j = 1, ..., K_n + M$ be the B–spline with support $[u_j, u_{j+M+1}]$ with respect to the knot sequence $(u_k)_{k\in\mathbb{Z}}$ (see, e.g., de Boor (1978), Chapter IX or Györfi et al. (2002), Section 14.1). The spline spaces which we will use for our estimates in case $d = 1$ will be defined as subspaces of

$$S_{K_n,M}([A,B]) = \left\{ \sum_{j\in\mathbb{Z}: supp(B_{j,M,K_n})\cap[A,B]\neq\emptyset} a_j \cdot B_{j,M,K_n} : j \in \mathbb{Z}, a_j \in \mathbb{R} \right\}.$$

Restricted on $[A, B]$ the space $S_{K_n,M}([A, B])$ consists of all functions f that are $(M-1)$-times continuously differentiable on $[A, B]$ and that are on each interval $[u_j, u_{j+1})$ equal to a polynomial of degree M (or less). For our function space we restrict the coefficients in $S_{K_n,M}([A, B])$ such that the functions are bounded and Lipschitz continuous. More precisely, we set

$$\begin{aligned} S_{K_n,M,\beta_n,\gamma_n}([A,B]) \quad = \quad & \left\{ \sum_{j\in\mathbb{Z}} a_j B_{j,M,K_n} : |a_j| \le \beta_n, |a_j - a_{j-1}| \le \gamma_n/K_n, \right. \\ & \left. \qquad a_j = 0 \text{ if } supp(B_{j,M,K_n}) \cap [A,B] = \emptyset \quad (j \in \mathbb{Z}) \right\} \end{aligned} \tag{8}$$

for some $\beta_n, \gamma_n > 0$. By standard results on B-splines and its derivatives (cf., e.g., Lemmas 14.4 and 14.6 in Györfi et al. (2002)) it can be shown that each function in $S_{K_n,M,\beta_n,\gamma_n}([A, B])$ is bounded in absolute value by β_n and Lipschitz continuous with Lipschitz constant γ_n, which we will need later in the proofs.

In case of higher dimensions we will use tensor product B-splines. Let e_i be the i-th unit vector $(i = 1, \ldots, d)$. For a multi-index $k = (k_1, \ldots, k_d) \in \mathbb{Z}^d$, we define the multivariate B-spline $B_k : \mathcal{R}^d \to \mathbb{R}$ of degree $\mathbf{M} = (M_1, \ldots, M_d) \in \mathbb{N}^d$ by

$$B_{k,\mathbf{M},\mathbf{K}_n}(x^{(1)}, \ldots, x^{(d)}) = \prod_{i=1}^{d} B_{k_i,M_i,K_{i,n}} \left(x^{(i)} \right) \qquad (x^{(1)}, \ldots, x^{(d)} \in \mathbb{R}),$$

where $\mathbf{K}_n = (K_{1,n}, \ldots, K_{d,n}) \in \mathbb{N}^d$.

Accordingly we define

$$\begin{aligned} & S_{K_n,M,\beta_n,\gamma_n}(\bigtimes_{i=1}^{d} [A_i, B_i]) \\ & = \left\{ \sum_{j \in \mathbb{Z}^d} a_j B_{j,M,K_n} : |a_j| \le \beta_n, |a_j - a_{j-e_i}| \le \frac{\gamma_n}{\sqrt{d} K_{n,i}} \; (i = 1, \ldots, d) \right. \\ & \qquad \left. a_j = 0 \text{ if } supp(B_{j,M,K_n}) \cap \bigtimes_{i=1}^{d} [A_i, B_i] = \emptyset \quad (j \in \mathbb{Z}^d) \right\}. \end{aligned} \tag{9}$$

The definition of the B-splines implies that $S_{\mathbf{K}_n,\mathbf{M},\beta_n,\gamma_n}([\times_{i=1}^d [A_i, B_i])$ is a subset of a linear vector space of dimension $\prod_{i=1}^d (K_i + M_i + 1)$. Furthermore it follows as above that the functions in $S_{\mathbf{K}_n,\mathbf{M},\beta_n,\gamma_n}(\times_{i=1}^d [A_i, B_i])$ are bounded in absolute value by β_n and are Lipschitz continuous with Lipschitz constant γ_n.

Given the above estimates of the continuation values, we can estimate the price of the option by

$$\hat{V}_{0,n} = \max\{f_0(x_0), \hat{q}_{n,0}(x_0)\}.$$

Since

$$\begin{aligned} \left| \hat{V}_{0,n} - V_0 \right| &= |\max\{f_0(x_0), \hat{q}_{n,0}(x_0)\} - \max\{f_0(x_0), q_0(x_0)\}| \le |\hat{q}_{n,0}(x_0) - q_0(x_0)| \\ &= \left(\int |\hat{q}_{n,0}(u) - q_0(u)|^2 \, \mathbf{P}_{X_0}(du) \right)^{1/2} \end{aligned} \tag{10}$$

(where the last equality follows from $X_0 = x_0$ $a.s.$) the error of this estimate tends to zero whenever the so-called L_2 error of our estimate of q_0 tends to zero.

Alternatively, we can estimate the price of the option by so-called lower estimates defined by Monte Carlo estimates of the expected payoff of a plug-in version of the stopping rule (4), cf., e.g., Subsection 8.6.1 in Glasserman (2004). It follows from proposition 1.3 in Belomestny (2011) that in this case the error of our estimated price of the option tends also to zero if the L_2 errors of the above estimates of the continuation values tend to zero.

3 Main results

In this section we consider three different models for the stock values and present for each model a consistency result of our estimation procedure applied to paths of versions of the model where the parameter values converge to the true values.

3.1 A Black-Scholes model

In this subsection the stock values are modelled via Black-Scholes theory, and the log-prices are given by $X_{k,t} = \log Z_{k,t}$, where

$$Z_{k,t} = z_{k,0} \cdot e^{r \cdot t} \cdot e^{\sum_{j=1}^{d} (\sigma_{k,j} \cdot W_j(t) - \frac{1}{2}\sigma_{k,j}^2 t)} \quad (k = 1, \ldots, d). \tag{11}$$

Here $r > 0$ is the riskless interest rate, $\sigma_k = (\sigma_{k,1}, \ldots, \sigma_{k,d})^T$ is the volatility of the k-th stock, $z_{k,0}$ is the initial stock price of the k-th stock, and $\{W_j(t) : t \in \mathbb{R}_+\}$ $(j = 1, \ldots, d)$ are independent Wiener processes. Since $W_j(1)$, $W_j(2) - W_j(1)$, $\ldots$, $W_j(T) - W_j(T-1)$ are independent standard normally distributed random variables, we can define $Z_{k,t}$ $(k = 1, \ldots, m, t = 0, \ldots, T)$ also by

$$Z_{k,0} = z_{k,0} \quad (k = 1, \ldots, d)$$

and by

$$Z_{k,t+1} = Z_{k,t} \cdot e^{r} \cdot e^{\sum_{j=1}^{d} (\sigma_{k,j} \cdot \epsilon_{t+1,j} - \frac{1}{2}\sigma_{k,j}^2)} \quad (k = 1, \ldots, d, t = 0, \ldots, T-1)$$

where $(\epsilon_{t,j})_{t \in \{1,\ldots,T\}, j \in \{1,\ldots,d\}}$ are independent standard normally distributed random variables.

In the sequel we assume that instead of sample paths from

$$(X_t)_{t=0,\ldots,T} = ((X_{1,t}, \ldots, X_{d,t}))_{t=0,\ldots,T}$$

we observe

$$\left(\bar{X}_{t,i}^{(n)}\right)_{t=0,\ldots,T} \quad (i = 1, \ldots, n)$$

where

$$\bar{X}_{t,i}^{(n)} = (\bar{X}_{1,t,i}^{(n)}, \ldots, \bar{X}_{d,t,i}^{(n)})^T = (\log \bar{Z}_{1,t,i}^{(n)}, \ldots, \log \bar{Z}_{d,t,i}^{(n)})^T$$

is given by

$$\bar{Z}^{(n)}_{k,0,i} = z_{k,0} \quad (k = 1, \dots, d)$$

and by

$$\bar{Z}^{(n)}_{k,t+1,i} = \bar{Z}^{(n)}_{k,t,i} \cdot e^r \cdot e^{\sum_{j=1}^d (\hat{\sigma}^{(n)}_{k,j} \cdot \epsilon_{t+1,j,i} - \frac{1}{2}(\hat{\sigma}^{(n)}_{k,j})^2)} \quad (k = 1, \dots, d, t = 0, \dots, T-1)$$

for some independent standard normally distributed random variables $\epsilon_{t+1,j,i}$.

Theorem 1. *Assume that the discounted payoff function f_t with respect to the above defined log price process $(X_t)_{t=0,\dots,T}$ is bounded and Lipschitz continuous. Let V_0 and q_t be the corresponding price of the option and continuation values. Let $M \in \mathbb{N}, \mathbf{M} = \sum_{i=1}^d e_i M, K_n \in \mathbb{N}, \mathbf{K}_n = \sum_{i=1}^d e_i K_n, \beta_n > 0, \gamma_n > 0$ and $A_n > 0$ and let the estimate be defined as in Section 2 with*

$$\mathcal{F}_n = S_{\mathbf{K}_n,\mathbf{M},\beta_n,\gamma_n}([-A_n, A_n]^d).$$

Assume that the parameters of the function spaces satisfy

$$A_n \to \infty, \beta_n \to \infty, \gamma_n \to \infty, \frac{A_n}{K_n} \to 0, \frac{\beta_n^5 \cdot A_n^d K_n^d}{n} \to 0 \qquad (n \to \infty).$$

If the parameters of the estimated model converge to the true parameter values in the sense that

$$\gamma_n \cdot (\hat{\sigma}^{(n)}_{k,j} - \sigma_{k,j}) \to 0 \quad (n \to \infty)$$

for all $k, j \in \{1, \dots, d\}$ then we have for $t = 0, \dots, T$

$$\int |\hat{q}_{n,t}(x) - q_t(x)|^2 \, \mathbf{P}_{X_t}(dx) \to 0 \qquad (n \to \infty) \text{ in probability}$$

and, in addition,

$$\hat{V}_{0,n} \to V_0 \qquad (n \to \infty) \text{ in probability.}$$

3.2 A GARCH model

Next we present results for a price process, where the volatility is modelled by a GARCH time series, which was introduced by Bollerslev (1986). We consider this in the form proposed in Duan (1995), where the GARCH process is modified in such a way that the discounted price process is a martingale.

Here the price process $\{S_t\}_{t=0,1,\dots}$ of a stock is modelled by

$$S_t = x_0 \cdot \exp\left(r \cdot t - \frac{1}{2} \sum_{j=1}^{t} h_j + \sum_{j=1}^{t} \sqrt{h_j} \cdot \epsilon_j \right)$$

where $x_0 \in \mathbb{R}_+$ is the value of the stock at time zero, $r > 0$ is the riskless interest rate, $(\epsilon_j)_{j \in \mathbb{Z}}$ are independent standard normally distributed random variables and where the (random) volatility h_t of the process satisfies

$$h_t = 0 \quad \text{for } t \leq 0$$

and

$$h_t = a_0 + \sum_{j=1}^{q} a_j \cdot h_{t-j} \left(\epsilon_{t-j} - \lambda\right)^2 + \sum_{j=1}^{p} b_j \cdot h_{t-j} \quad (t \in \mathbb{N})$$

for some $p, q \in \mathbb{N}_0$ and parameters $\lambda > 0$, $a_j > 0$ ($j = 0, \dots, p$) and $b_j > 0$ ($j = 1, \dots, p$).

In applications the parameters $\lambda, a_0, \dots, a_q, b_1, \dots, b_p$ are unknown. Therefore it is only possible to generate Monte Carlo samples where these parameters are estimated. Given sequences $(\hat{a}_{i,n})_{n \in \mathbb{N}}$ ($i = 0, \dots, q$), $(\hat{b}_{i,n})_{n \in \mathbb{N}}$ ($i = 1, \dots, p$), $(\hat{\lambda}_n)_{n \in \mathbb{N}}$ of nonegative real numbers, where $\hat{a}_{0,n} > 0$ for all $n \in \mathbb{N}$, and independent standard normally distributed random variables $\epsilon_{j,i}$, the samples of the error-behaved logarithmic returns are given by

$$\bar{Z}_{t,i}^{(n)} = \log(x_0) + r \cdot t - \frac{1}{2} \sum_{j=1}^{t} \bar{h}_{j,i}^{(n)} + \sum_{j=1}^{t} \sqrt{\bar{h}_{j,i}^{(n)}} \cdot \epsilon_{j,i} \tag{12}$$

$$\bar{h}_{t,i}^{(n)} = \hat{a}_{0,n} + \sum_{j=1}^{q} \hat{a}_{j,n} \cdot \bar{h}_{t-j,i}^{(n)} \left(\epsilon_{t-j,i} - \hat{\lambda}_n\right)^2 + \sum_{j=1}^{p} \hat{b}_{j,n} \cdot \bar{h}_{t-j,i}^{(n)} \tag{13}$$

($i = 1, \dots, n$), where we set again $\bar{h}_{t,j}^{(n)} = 0$ for $t \leq 0$.

To ensure that the price process is a Markov process, we have to extend the state space. So instead of only $\bar{Z}_{t,i}^{(n)}$ we consider

$$\bar{X}_{t,i}^{(n)} = \left(\bar{Z}_{t,i}^{(n)}, \epsilon_{t,i}, \dots, \epsilon_{t+1-q,i}, \bar{h}_{t,i}^{(n)}, \dots, \bar{h}_{t+1-\max\{p,q\},i}^{(n)} \right)^T. \tag{14}$$

Theorem 2. *Assume that the payoff-function f_t is bounded and Lipschitz continuous. Let V_0 and q_t be the price of the option and the continuation values corresponding to the above defined log price process $\log(S_t)$. Let $\bar{X}_{t,i}^{(n)}$ be defined by (14). Let $M \in \mathbb{N}, \mathbf{M} = \sum_{i=1}^{1+q+\max\{p,q\}} e_i M, K_n \in \mathbb{N}, \mathbf{K}_n = \sum_{i=1}^{1+q+\max\{p,q\}} e_i K_n, \beta_n > 0, \gamma_n > 0$ and $A_n > 0$. Let* *the estimates $\hat{q}_{n,t}$ be defined as in Section 2, where the function space is given by*

$$\mathcal{F}_n = S_{\mathbf{K}_n,\mathbf{M},\beta_n,\gamma_n}\left(\bigtimes_{i=1}^{1+q+\max\{p,q\}} [A_i, B_i]\right).$$

Assume

$$A_{i,n} \to -\infty, B_{i,n} \to \infty, \frac{\gamma_n}{B_{i,n} - A_{i,n}} \to \infty, \gamma_n \frac{B_{i,n} - A_{i,n}}{K_n} \to 0,$$

for $i = 1, \dots, 1 + q + \max\{p, q\}$,

$$\beta_n \to \infty$$

and

$$\frac{\beta_n^5 \prod_{i=1}^{1+q+\max\{p,q\}} K_{i,n}}{n} \to 0 \quad (n \to \infty).$$

Then

$$\gamma_n\left(\hat{a}_{i,n} - a_i\right) \to 0 \quad (n \to \infty), \quad \gamma_n\left(\hat{b}_{j,n} - b_j\right) \to 0 \quad (n \to \infty)$$

for all $i \in \{0, \dots, q\}$, $j \in \{1, \dots, p\}$ and

$$\gamma_n^2(\lambda - \hat{\lambda}_n) \to 0 \quad (n \to \infty)$$

imply

$$\int |\hat{q}_{n,t}(x) - q_t(x)|^2 \, \mathbf{P}_{X_{t,1}}(dx) \to 0 \quad \textit{in probability}$$

for all $t = 0, 1 \dots, T$ and, in addition,

$$\hat{V}_{0,n} \to V_0 \quad \textit{in probability.}$$

3.3 A Levy model

Finally we present a result for a Lévy Processes. Here we consider the following Merton Model (cf. Merton (1976)):

$$S_t = x_0 \cdot \exp\left(\mu \cdot t + \sigma \cdot W_t + \sum_{j=1}^{N_t} Y_j\right),$$

where $W = (W_t)_{t \in \mathbb{R}_+}$ is a Wiener process, $N = (N_t)_{t \in \mathbb{R}_+}$ is a Poisson process with parameter λ independent from W and Y_1, Y_2, …are independent normally distributed random variables with mean m and variance δ^2 independent from W and N. By defining

$$\mu = r - \frac{\sigma^2}{2} - \lambda\left(\exp\left(m + \frac{\delta^2}{2}\right) - 1\right) \tag{15}$$

the price process with respect to the martingale measure in the sense of Merton (1976) can be written as

$$S_t = s_0 \exp\left(\mu t + \sum_{s=1}^{t} \sigma\epsilon_s + \sum_{i=1}^{N_t}(m + \delta\xi_i)\right), \tag{16}$$

where ϵ_s, ξ_i with $i, s \in \mathbb{N}$ are independent standard normal distributed and N_t is Poisson distributed with parameter λt. Again we consider the corresponding log price process

$$X_t = \log s_0 + \mu t + \sum_{s=1}^{t} \sigma\epsilon_s + \sum_{i=1}^{N_t}(m + \delta\xi_i).$$

As in the Black-Scholes and the GARCH case we we estimate the parameters σ, λ, m and δ by $\hat{\sigma}_n$, $\hat{\lambda}_n$, $\hat{m}_n$ and $\hat{\delta}_n$, and consider the logarithm of the returns. According to Merton (1976) we define $\hat{\mu}_n$ by

$$\hat{\mu}_n = r - \frac{\hat{\sigma}_n^2}{2} - \hat{\lambda}_n\left(\exp\left(\hat{m}_n + \frac{\hat{\delta}_n^2}{2}\right) - 1\right)$$

and generate data

$$\bar{X}_{t,i}^{(n)} = \log(x_0) + \hat{\mu}_n t + \sum_{s=1}^{t} \hat{\sigma}_n\epsilon_{s,i} + \sum_{i=1}^{N_{t,i}^{(n)}}(\hat{m}_n + \hat{\delta}_n\xi_i) \qquad (t = 0, \dots, T, i = 1, \dots, n), \tag{17}$$

where $\epsilon_{i,j}, \xi_{i,j}$ are independent standard normal distributed and $N_t^{(i,n)}$ is a Poisson distributed random variable with parameter $\hat{\lambda}_n t$ independent from $\epsilon_{i,j}, \xi_{i,j}$ for all $i, j \in \mathbb{N}$. Note that above definition of μ_n ensures the martingale property of the discounted price process.

Theorem 3. *Assume that the payoff-function f_t is bounded and Lipschitz continuous. Let V_0 and q_t be the price of the option and the continuation values corresponding to the above defined log price process X_t. Let $\bar{X}_{t,i}^{(n)}$ be defined by (17). Let $M \in \mathbb{N}, K_n \in \mathbb{N}, \beta_n > 0, \gamma_n > 0$ and $A_n > 0$. Let the estimates $\hat{q}_{n,t}$ be defined as in Section 2, where the function space $S_{K_n,M,\beta_n,\gamma_n}([-A_n, A_n])$ is defined by (8). Assume that*

$$A_n \to \infty, \beta_n \to \infty, \gamma_n \to \infty, \frac{A_n}{K_n} \to 0, \frac{\beta_n^5 \cdot A_n^d K_n^d}{n} \to 0 \qquad (n \to \infty).$$

Then

$$\gamma_n \cdot (\hat{\sigma}_n - \sigma) \to 0, \quad \gamma_n \cdot (m_n - m) \to 0, \quad \gamma_n \cdot (\hat{\delta}_n - \delta) \to 0$$

and

$$\gamma_n^2 \cdot (\hat{\lambda}_n - \lambda) \to 0 \quad (n \to \infty)$$

imply

$$\int |\hat{q}_{n,t}(x) - q_t(x)|^2 \, \mathbf{P}_{X_{t,1}}(dx) \to 0 \qquad \textit{in probability}$$

for all $t \in \{0, \ldots, T\}$ *and, in addition,*

$$\hat{V}_{0,n} \to V_0 \qquad \textit{in probability.}$$

4 Application to simulated data

To demonstrate the finite sample performance of our estimate we first apply it in case of a Duan-GARCH model. With respect to the martingale measure of Duan (cf. Duan (1995)) the logarithm of the price process has the form

$$\log(S_t) = \log(s_0) + rt - \frac{1}{2} \sum_{s=}^{t} h_s + \sum_{s=1}^{t} \sqrt{h_s} \epsilon_s,$$

with

$$h_t = a0 + a_1 |h_{t-1}| (\epsilon_{t-1} - \lambda)^2 + b_1 h_{t,1}.$$

with independent standard normal distributed random variables $\epsilon_0, \ldots, \epsilon_t$. As parameters of the real price process we take

$$a_0 = 0.0000166, a_1 = 0.144, b_1 = 0.776 \text{ and } \lambda = 0.7138.$$

These values are taken from an example in Duan (1995). To ensure the markov property of the considered process we take

$$X_t = \begin{pmatrix} \log(S_t) \\ \exp(\epsilon_t) \\ h_t \end{pmatrix}.$$

The riskless rate is assumed to be known as

$$r = 0.05/T \approx 0.001389.$$

As start vector of this process we choose

$$X_0 = \begin{pmatrix} \log(100) \\ 1 \\ 1 \end{pmatrix}$$

In applications the parameters a_0, a_1, b_1 and λ are unknown and have to be estimated from historical data. To demonstrate the influence of such an estimate, we consider different error levels for each parameter.

We price a Bermudan capped-straddle option with exercise prices $70, 100$ and 130. The payoff function is shown in Figure 1. As maturity we choose $T = 36$, so exercising is

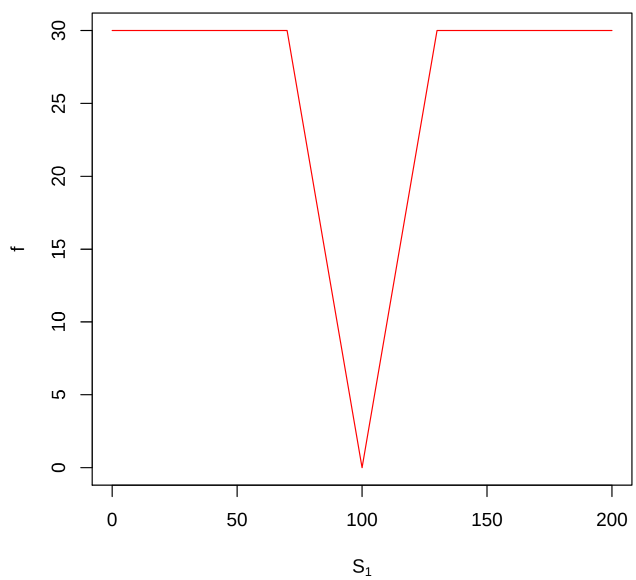

Figure 1: Payoff function of a Capped-Straddle with strike prices 70, 100 und 130

possible at the timepoints $t = 0, 1, \ldots, 36$.

As parameters of the spline space we take $A = (0, 0, 0)^T, B = (200, 2, 120)^T, M = (1, 1, 1)^T$. The parameter $K_1 \in \{4, 10\}$ is choosen by splitting of the sample, K_2 and K_3 are set to 2. To estimate the continuation values we use $n_l = 1000$ paths as learning data and $n_t = 1000$ testing data. The so computed estimates for continuation values are taken as plug-in estimates to evaluate on $n_a = 10000$ newly generated paths. The arithemic mean of these n_a paths is the estimate of one option price. This procedure is repeated 50 times for each parameter constellation. So we get for every chosen parameter set 50 option prices.

Figures 2, 3, 4 and 5 show the simulation results in form of boxplots. Each of these simulations consider different parameter levels for one of the parameter while all other parameters are set to the value of the real price process defined above.

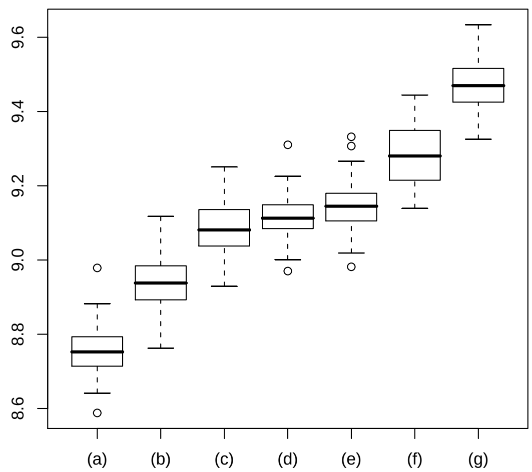

Figure 2: Simulated option prices in case of a Duan GARCH model with (a) $a_0 = 0.0001494$, (b) $a_0 = 0.0001577$, (c) $a_0 = 0.00016434$, (d) $a_0 = 0.000166$, (e) $a_0 = 0.00016766$, (f) $a_0 = 0.0001743$, (g) $a_0 = 0.0001826$

The next simulation series considers the pricing of a Bermudan bear spread option with strikes 60 and 140. The corresponding payoff function is illustrated as Figure 6. The maturity is set to $T = 36$, the riskless rate is $r = 0.05/T$.

This time we use a Merton modell for simulating the asset prices. We start with the log price process (16). Here we assume, that the real parameters are $\sigma = 0.02/\sqrt{T}, m = 0.2, \delta = 0.5, \lambda = 0.2$. The parameter μ is computed by (15).

The parameters of the spline space are $K_n \in \{4, 10\}, A = 0, B = 250, M_n \in \{1, 3\}$. The number of learning data for choosing the spline space parameters is $n_l = 1000$. The corresponding testing data is $n_t = 1000$. The so computed estimates for continuation

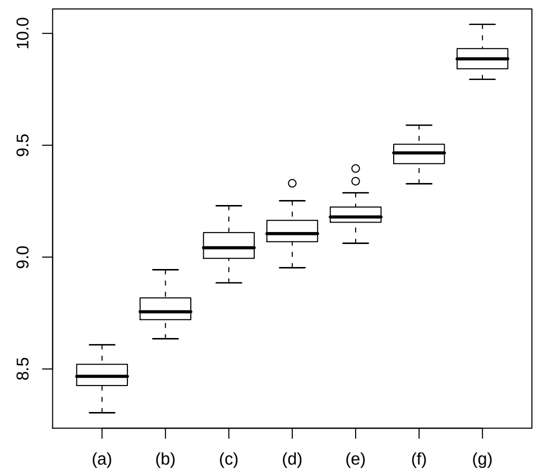

Figure 3: Simulated option prices in case of a Duan GARCH model with (a) $a_1 = 0.1296$, (b) $a_1 = 0.1368$, (c) $a_1 = 0.14256$, (d) $a_1 = 0.144$, (e) $a_1 = 0.14544$, (f) $a_1 = 0.1512$, (g) $a_1 = 0.1584$

values are taken as plug-in estimates to evaluate on $n_a = 10000$ new generated paths. The arithemic mean of these n_a paths is the estimate of one option price. This procedure is repeated 50 times for each parameter constellation. So we get for every chosen parameter set 50 option prices. Startvalue of the price process is 100.

Figure 7 shows the results, if we choose m as

$$0.18, 0.19, 0.198, 0.2, 0.202, 0.21, 0.22$$

and keep all other parameters fixed.

Figure 8 shows the results, if we choose δ as

$$0.45, 0.475, 0.495, 0.5, 0.505, 0.525, 0.55$$

and keep all other parameters fixed.

Figure 9 shows the results, if we choose λ as

$$0.18, 0.19, 0.198, 0.2, 0.202, 0.21, 0.22$$

and keep all other parameters fixed.

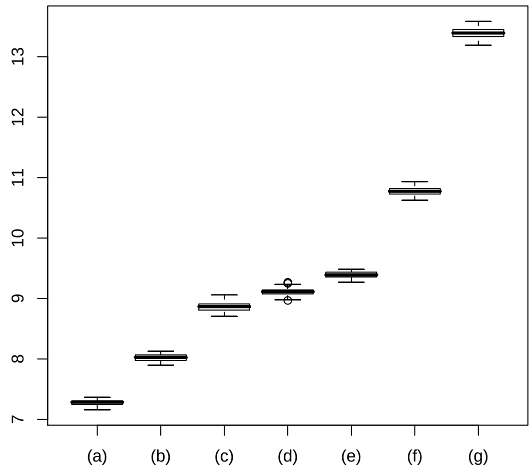

Figure 4: Simulated option prices in case of a Duan GARCH model with (a) $b_1 = 0.69840$, (b) $b_1 = 0.7372$, (c) $b_1 = 0.76824$, (d) $b_1 = 0.776$, (e) $b_1 = 0.78376$, (f) $b_1 = 0.8148$, (g) $b_1 = 0.85360$

We do not consider an error in the remaining parameters, due to the similarity in case of the Black-Scholes model.

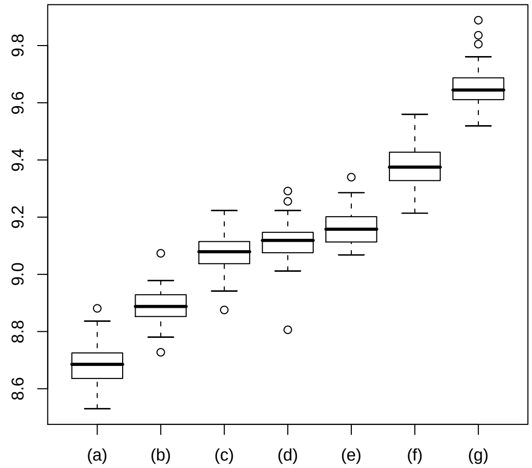

Figure 5: Simulated option prices in case of a Duan GARCH model with (a) $\lambda = 0.642420$, (b) $\lambda = 0.678110$, (c) $\lambda = 0.706662$, (d) $\lambda = 0.713800$, (e) $\lambda = 0.720938$, (f) $\lambda = 0.749490$, (g) $\lambda = 0.785180$

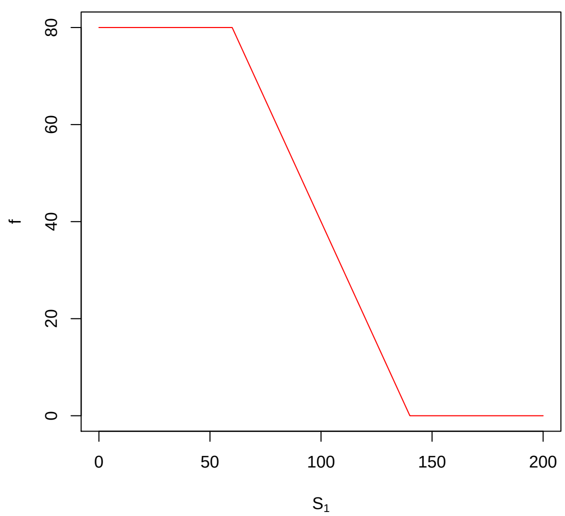

Figure 6: Payoff function of a bear spread with strikes 60 and 140

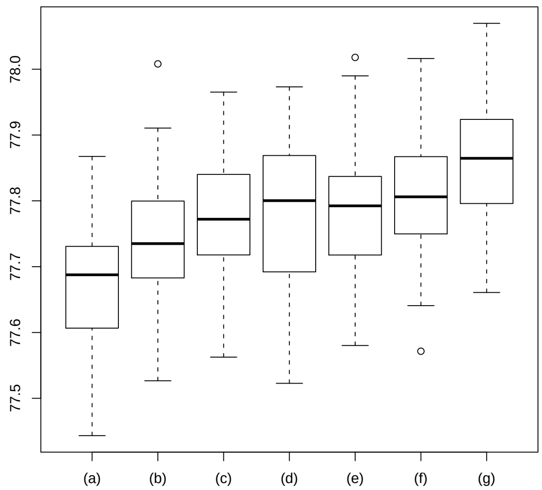

Figure 7: Simulated option prices in case of a Bermudan bear spread option considering a Merton model with (a) $m = 0.18$, (b) $m = 0.19$, (c) $m = 0.198$, (d) $m = 0.2$, (e) $m = 0.202$, (f) $m = 0.21$, (g) $m = 0.22$

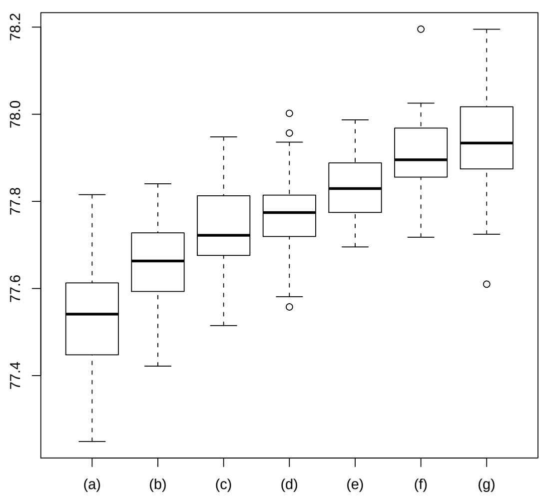

Figure 8: Simulated option prices in case of a Bermudan bear spread option considering a Merton model with (a) $\delta = 0.45$, (b) $\delta = 0.475$, (c) $\delta = 0.495$, (d) $\delta = 0.5$, (e) $\delta = 0.505$, (f) $\delta = 0.525$, (g) $\delta = 0.55$

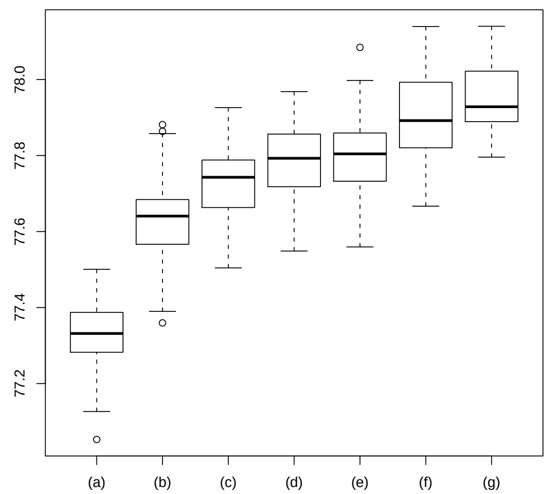

Figure 9: Simulated option prices in case of a Bermudan bear spread option considering a Merton model with (a) $\lambda = 0.18$, (b) $\lambda = 0.19$, (c) $\lambda = 0.198$, (d) $\lambda = 0.2$, (e) $\lambda = 0.202$, (f) $\lambda = 0.21$, (g) $\lambda = 0.22$

5 Proofs

5.1 An auxiliary lemma

In the proofs we will need the following lemma.

Lemma 1. *Let $f_t, q_t, \bar{q}_t : \mathcal{R}^d \to \mathbb{R}$ be functions ($t = 0, \ldots, T$). For given $\mathcal{R}^d$-valued stochastic processes $(X_t)_{t=0,\ldots,T}, (\bar{X}_t)_{t=0,\ldots,T}$ and $t \in \{0, \ldots, T-1\}$ define*

$$Y_t = \max\{f_{t+1}(X_{t+1}), q_{t+1}(X_{t+1})\} \text{ and } \bar{Y}_t = \max\{f_{t+1}(\bar{X}_{t+1}), \bar{q}_{t+1}(\bar{X}_{t+1})\}.$$

Let $s \in \{0, \ldots, T-1\}$ and assume that f_{s+1} is Lipschitz continuous with Lipschitz constant L. Then

$$\left|Y_s - \bar{Y}_s\right|^2 \leq 2L^2 \left\|X_{s+1} - \bar{X}_{s+1}\right\|^2 + 2\left|q_{s+1}(X_{s+1}) - \bar{q}_{s+1}(\bar{X}_{s+1})\right|^2.$$

Proof. Using $(a+b)^2 \leq 2a^2 + 2b^2$ and $|\max\{a,b\} - \max\{a,c\}| \leq |b-c|$ for $a, b, c \in \mathbb{R}$ we get

$$\begin{aligned} \left|Y_s - \bar{Y}_s\right|^2 &\leq 2\left|\max\{f_{s+1}(X_{s+1}), q_{s+1}(X_{s+1})\} - \max\{f_{s+1}(X_{s+1}), \bar{q}_{s+1}(\bar{X}_{s+1})\}\right|^2 \\ &\quad +2\left|\max\{f_{s+1}(X_{s+1}), \bar{q}_{s+1}(\bar{X}_{s+1})\} - \max\{f_{s+1}(\bar{X}_{s+1}), \bar{q}_{s+1}(\bar{X}_{s+1})\}\right|^2 \\ &\leq 2\left|q_{s+1}(X_{s+1}) - \bar{q}_{s+1}(X_{s+1})\right|^2 + 2\left|f_{s+1}(X_{s+1}) - f_{s+1}(\bar{X}_{s+1})\right|^2. \end{aligned}$$

By using the Lipschitz property of f_{s+1} we get the desired result. □

5.2 A general consistency result

In this subsection we formulate and proof a general consistency result. Here we assume that instead of independent copies

$$(X_{t,i})_{t=0,\ldots,T} \quad (i = 1, \ldots, n)$$

of the underlying $\mathbb{R}^d$-valued Markov process $(X_t)_{t=0,\ldots,T}$ we have given paths

$$(\bar{X}_{t,i}^{(n)})_{t=0,\ldots,T} \quad (i = 1, \ldots, n)$$

such that

$$\frac{1}{n}\sum_{i=1}^{n} \|\bar{X}_{t,i}^{(n)} - X_{t,i}\|^2$$

is small for all t. We define our estimates $\hat{q}_{n,t}$ and $\hat{V}_{0,n}$ as in Section 2 with a general function space $\mathcal{F}_n$. Then the following result holds.

Theorem 4. *Let the discounted payoff function f_t be bounded in absolute value by some $M > 1$ and Lipschitz continuous with Lipschitz constant $L > 0$. Let $\mathcal{F}_n$ be a subspace of a linear vector space of dimension D_n consisting of functions which are bounded in absolute value by some $\beta_n > 0$ and which are Lipschitz continuous with respect to some Lipschitz constant γ_n. Assume that*

$$X_{1,t+1}, \dots, X_{n,t+1} \text{ and } \bar{X}_{1,t}^{(n)}, \dots, \bar{X}_{n,t}^{(n)} \text{ are independent given } X_{1,t}, \dots, X_{n,t} \tag{18}$$

for all $t = 0, \dots, T-1$. Then

$$\frac{D_n \beta_n^5}{n} \to 0, \quad D_n \beta_n^3 \to \infty, \quad \delta_n \to \infty \quad (n \to \infty),$$

$$\inf_{f \in \mathcal{F}_n} \int |f(x) - q_t(x)|^2 \, \mathbf{P}_{X_t}(dx) \to 0 \quad (n \to \infty) \tag{19}$$

for all $t = 0, \dots, T-1$ and

$$\frac{\gamma_n^2}{n} \sum_{i=1}^{n} \left\| \bar{X}_{t,i}^{(n)} - X_{t,i} \right\|^2 \to 0 \quad \text{in probability} \quad (n \to \infty)$$

for all $t = 0, \dots, T$ imply

$$\int |\hat{q}_{n,t}(x) - q_t(x)|^2 \, \mathbf{P}_{X_t}(dx) \to 0 \quad \text{ in probability } (n \to \infty) \tag{20}$$

for $t = 0, \dots, T$ and, in addition

$$\hat{V}_{0,n} \to V_0 \quad \text{in probability.} \tag{21}$$

In the proof we will apply Theorem 1 of Fromkorth and Kohler (2011). In case of bounded Y, the Sub-Gaussian condition there is trivially fullfilled. Using Lemma 9.3 in Györfi et al. (2002) we can conclude from the result there the following lemma:

Lemma 2. *Let $(X, Y), (X_1, Y_1), \dots, (X_n, Y_n)$ be independent and identically distributed $\mathcal{R}^d \times \mathbb{R}$ valued random vectors with $|Y| \leq \beta$ a.s. for some $\beta > 0$. For each n let $\mathcal{F}_n$ be a subset of a linear vector space of dimension D_n consisting of functions $f : \mathbb{R}^d \to \mathbb{R}$ which are bounded in absolute value by β_n and which are Lipschitz continuous with Lipschitz constant γ_n. Given an arbitrary data set*

$$\bar{\mathcal{D}}_n = \left\{ (\bar{X}_{1,n}, \bar{Y}_{1,n}), \dots, (\bar{X}_{n,n}, \bar{Y}_{n,n}) \right\}$$

with the property, that $Y_1, \ldots, Y_n$ and $\bar{X}_{1,n}, \ldots, \bar{X}_{n,n}$ are independent given $X_1, \ldots, X_n$, define the estimate $\bar{m}_n$ by

$$\bar{m}_n(\cdot) = \arg\min_{f \in \mathcal{F}_n} \frac{1}{n} \sum_{i=1}^{n} \left| f\left(\bar{X}_{i,n}\right) - \bar{Y}_{i,n} \right|^2 .$$

If

$$\frac{D_n \beta_n^5}{n} \to 0, \quad D_n \beta_n^3 \to \infty \quad (n \to \infty)$$

then it holds for c sufficiently large that we have for any $n \in \mathbb{N}$

$$\mathbf{P}\left\{ \int \left| \bar{m}_n(x) - m(x) \right|^2 \mu(dx) > c \cdot Z_n \right\} \leq c \cdot \exp\left(-c \cdot D_n \cdot \beta_n^3\right)$$

and

$$\mathbf{P}\left\{ \frac{1}{n} \sum_{i=1}^{n} \left| \bar{m}_n(\bar{X}_{i,n}) - m(X_i) \right|^2 > c \cdot Z_n \right\} \leq c \cdot \exp\left(-c \cdot D_n \cdot \beta_n^3\right)$$

where

$$Z_n = \frac{1}{n} \sum_{i=1}^{n} \left| Y_i - \bar{Y}_{i,n} \right|^2 + \gamma_n^2 \cdot \frac{1}{n} \sum_{i=1}^{n} \left\| X_i - \bar{X}_{i,n} \right\|^2 + \frac{D_n \beta_n^5}{n} + \inf_{f \in \mathcal{F}_n} \int |f(x) - m(x)|^2 \mu(dx).$$

Proof. The result follows directly from Theorem 1 in Fromkorth and Kohler (2011) and Lemma 9.3 in Györfi et al. (2002). □

Proof of Theorem 4. We prove the theorem by backward induction. We start with $t = T$, in which case we have $\hat{q}_{n,T}(x) = q_T(x) = 0$, which implies

$$\int |\hat{q}_{n,s}(x) - q_s(x)|^2 \mathbf{P}_{X_s}(dx) \to 0 \quad \text{in probability} \tag{22}$$

and

$$\frac{1}{n} \sum_{i=1}^{n} |\hat{q}_{n,s}(\bar{X}_{s,i}) - q_s(\bar{X}_{s,i}^{(n)})|^2 \to 0 \quad \text{in probability} \tag{23}$$

for $s = T$.

Let $t \in \{0, \ldots, T-1\}$ be arbitrary and assume that (22) and (23) hold for $s = t+1$. In the sequel we show (22) and (23) for $s = t$. To do this we apply Lemma 2 with

$$X_i = X_{i,t}, \quad \bar{X}_{i,n} = \bar{X}_{i,t}^{(n)}, \quad Y_i = \max\{f_{t+1}(X_{i,t+1}), q_{t+1}(X_{i,t+1})\}$$

and

$$\bar{Y}_{i,n} = \max\{f_{t+1}(\bar{X}_{i,t+1}^{(n)}), \hat{q}_{n,t+1}(\bar{X}_{i,t+1}^{(n)}\}.$$

Here (18) implies that $Y_1, \ldots, Y_n$ and $\bar{X}_{1,n}, \ldots, \bar{X}_{n,n}$ are independent given $X_1, \ldots, X_n$, so we can conclude

$$\mathbf{P}\left\{\int |\hat{q}_{n,t}(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) > c_{11} \cdot Z_n\right\} \to 0 \quad (n \to \infty)$$

and

$$\mathbf{P}\left\{\frac{1}{n}\sum_{i=1}^n |\hat{q}_{n,t}(\bar{X}_{t,i}) - q_t(X_{t,i})|^2 > c_{11} \cdot Z_n\right\} \to 0 \quad (n \to \infty)$$

where

$$\begin{aligned} Z_n &= \frac{1}{n}\sum_{i=1}^n \left|\max\{f_{t+1}(X_{i,t+1}), q_{t+1}(X_{i,t+1})\} - \max\{f_{t+1}(\bar{X}_{i,t+1}^{(n)}), \hat{q}_{n,t+1}(\bar{X}_{i,t+1}^{(n)})\}\right|^2 \\ &\quad + \gamma_n^2 \cdot \frac{1}{n}\sum_{i=1}^n \left\|X_{i,t} - \bar{X}_{i,t}^{(n)}\right\|^2 + \frac{D_n \cdot \beta_n^5}{n} + \inf_{f \in \mathcal{F}_n} \int |f(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx). \end{aligned}$$

Lemma 1 implies that for n sufficiently large (i.e., in case $\gamma_n \geq L$) we have

$$\begin{aligned} Z_n &\leq 2\frac{1}{n}\sum_{i=1}^n |\hat{q}_{n,t+1}(\bar{X}_{i,t+1}^{(n)}) - q_{t+1}(X_{i,t+1})|^2 \\ &\quad + 3 \cdot \gamma_n^2 \cdot \frac{1}{n}\sum_{i=1}^n \left\|X_{i,t+1} - \bar{X}_{i,t+1}^{(n)}\right\|^2 + \frac{D_n \cdot \beta_n^5}{n} + \inf_{f \in \mathcal{F}_n} \int |f(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx). \end{aligned}$$

By the induction hypothesis and the assumptions of the theorem, we have

$$Z_n \to 0 \quad \text{in probability.}$$

The proof of (20) is complete, and using (10) we also get (21). □

We reformulate Theorem 4 in case of choosing the function space as a spline space.

Corollary 1. *Let the discounted payoff function f_t be bounded in absolute value by some $\beta > 1$ and Lipschitz continuous with Lipschitz constant $L > 0$. For $n \in \mathbb{N}$ let $A_n, \beta_n, \gamma_n > 0$, $K_n \in \mathbb{N}, \mathbf{K}_n = (K_n, \ldots, K_n)$, $\mathbf{M} \in \mathbb{N}_0^2$ and set*

$$\mathcal{F}_n = S_{\mathbf{K}_n, \mathbf{M}, \beta_n, \gamma_n}\left([-A_n, A_n]^d\right).$$

Assume that

$$X_{1,t+1}, \ldots, X_{n,t+1} \text{ and } \bar{X}_{1,t}^{(n)}, \ldots, \bar{X}_{n,t}^{(n)} \text{ are independent given } X_{1,t}, \ldots, X_{n,t} \tag{24}$$

for all $t = 1, \ldots, T-1$. Then

$$A_n \to \infty, \beta_n \to \infty, \gamma_n \to \infty, \frac{A_n}{K_n} \to 0, \frac{\beta_n^5 \cdot A_n^d K_n^d}{n} \to 0 \qquad (n \to \infty)$$

and

$$\frac{\gamma_n^2}{n}\sum_{i=1}^n \left\|\bar{X}_{t,i}^{(n)} - X_{t,i}\right\|^2 \to 0 \quad \textit{in probability}$$

for all $t = 0, \ldots, T$ *imply*

$$\int \left|\hat{q}_t^{(n)}(x) - q_t(x)\right|^2 \mathbf{P}_{X_t}(dx) \to 0 \quad \textit{in probability}$$

for $t = 0, \ldots, T$ *and, in addition*

$$\hat{V}_{0,n} \to V_0 \quad \textit{in probability.}$$

Proof. Corollary 1 follows directly from Theorem 4 if we observe that (19) follows from $q_t \in \L_2(\mathbf{P}_{X_t}$ (which is implied by the boundaries of the payoff function) and approximation properties of spline spaces (cf., e.g., proof of Corollary 2 in Fromkorth and Kohler (2011). □

5.3 Proof of Theorem 1

Set $X_{t,i} = (X_{1,t,i}, \ldots, X_{d,t,i})'$, where

$$X_{k,t,i} = \log(z_{k,0}) + r \cdot t + \sum_{s=1}^t \sum_{j=1}^d \left(\sigma_{k,j} \cdot \epsilon_{s,j,i} - \frac{1}{2} \cdot \sigma_{k,j}^2\right).$$

This can be interpreted as an artificial sample of the logarithm of asset values with the real (but unknown) distribution.

For all $t = 0, \ldots, T$ it holds

$$\begin{aligned}
&\left\|X_{t,1} - \bar{X}_{t,1}^{(n)}\right\|^2 \\
= &\sum_{k=1}^d \left|X_{k,t,1} - \bar{X}_{k,t,1}^{(n)}\right|^2 \\
= &\sum_{k=1}^d \left|\frac{t}{2} \cdot \sum_{j=1}^d \left((\hat{\sigma}_{k,j}^{(n)})^2 - \sigma_{k,j}^2\right) + \sum_{s=1}^t \sum_{j=1}^d \left((\sigma_{k,j} - \hat{\sigma}_{k,j}^{(n)}) \cdot \epsilon_{s,j,1}\right)\right|^2 \\
\leq &\sum_{k=1}^d d \cdot (t+1) \cdot \left(\frac{t^2}{4} \cdot \sum_{j=1}^d \left((\hat{\sigma}_{k,j}^{(n)})^2 - \sigma_{k,j}^2\right)^2 + \sum_{s=1}^t \sum_{j=1}^d (\sigma_{k,j} - \hat{\sigma}_{k,j}^{(n)})^2 \cdot \epsilon_{s,j,1}^2\right)
\end{aligned}$$

where we have used the inequality of Jensen. From this we get

$$\mathbf{E}\left\{\gamma_n^2 \left\|X_{t,1} - \bar{X}_{t,1}^{(n)}\right\|\right\}$$

$$
\begin{aligned}
&\le \sum_{k=1}^{d} t\cdot(d+1)\left(\frac{t^2}{4}\cdot\sum_{j=1}^{d}\gamma_n^2\left((\hat{\sigma}_{k,j}^{(n)})^2-\sigma_{k,j}^2\right)^2+\sum_{s=1}^{t}\sum_{j=1}^{d}\gamma_n^2(\sigma_{k,j}-\hat{\sigma}_{k,j}^{(n)})^2\cdot\mathbf{E}(\epsilon_{s,j,1}^2)\right)\\
&\le \sum_{k=1}^{d} t\cdot(d+1)\left(\frac{t^2}{4}\cdot\sum_{j=1}^{d}\gamma_n^2\left(\hat{\sigma}_{k,j}^{(n)}-\sigma_{k,j}\right)^2\cdot\left(\hat{\sigma}_{k,j}^{(n)}+\sigma_{k,j}\right)^2+t\cdot\sum_{j=1}^{d}\gamma_n^2(\sigma_{k,j}-\hat{\sigma}_{k,j}^{(n)})^2\right)\\
&\to 0 \quad (n\to\infty),
\end{aligned}
$$

where the last step follows from the assumptions of the theorem. For arbitrary $\epsilon > 0$ the Markov inequality implies now

$$
\begin{aligned}
\mathbf{P}\left\{\gamma_n^2\cdot\frac{1}{n}\sum_{i=1}^{n}\left\|X_{t,i}-\bar{X}_{t,i}^{(n)}\right\|^2>\epsilon\right\} &\le \frac{\mathbf{E}\left\{\gamma_n^2\cdot\frac{1}{n}\sum_{i=1}^{n}\left\|X_{t,i}-\bar{X}_{t,i}^{(n)}\right\|^2\right\}}{\epsilon}\\
&= \frac{\mathbf{E}\left\{\gamma_n^2\left\|X_{t,1}-\bar{X}_{t,1}^{(n)}\right\|\right\}}{\epsilon}\to 0 \quad (n\to\infty),
\end{aligned}
$$

this means

$$
\gamma_n^2\cdot\frac{1}{n}\sum_{i=1}^{n}\left\|X_{t,i}-\bar{X}_{t,i}^{(n)}\right\|^2\to 0 \quad (n\to\infty) \text{ in probability.}
$$

Finally we note, that $X_{1,t+1},\ldots,X_{n,t+1}$ depend only on random variables independent from all random variables used up to time t provided we fix $X_{t,1},\ldots,X_{t,n}$. So the independence assumption in Corollary 1 is trivially fulfilled.

Corollary 1 implies the assertion. □

5.4 Proof of Theorem 2

Again we introduce the artificial sample $X_{t,i}$ of the real distribution, i.e. we set

$$
X_{t,i}=\left(Z_{t,i},\epsilon_{t,i},\ldots,\epsilon_{t-q+1,i},h_{t,i},\ldots,h_{t-\max\{p,q\}+1,i}\right)^T
$$

where

$$
\begin{aligned}
Z_{t,i} &= \log(x_0)+rt-\frac{1}{2}\sum_{j=1}^{t}h_{j,i}+\sum_{j=1}^{t}\sqrt{h_{j,i}}\cdot\epsilon_{j,i}\\
h_{t,i} &= a_0+\sum_{j=1}^{q}a_j\cdot h_{t-j,i}\left(\epsilon_{t-j,i}-\lambda\right)^2+\sum_{j=1}^{p}b_j\cdot h_{t-j,i},
\end{aligned}
$$

for $t>0$ and $h_{s,i}=0$ for $s\le 0$.

As in the proof of Theorem 1 it suffices to show

$$\mathbf{E}\left\{\gamma_n^2 \left\|X_{t,1} - \bar{X}_{t,1}^{(n)}\right\|^2\right\} \to 0 \quad (n \to \infty).$$

By the definition of the stochastic processes and the inequality of Jensen we get

$$\begin{aligned}
&\left\|X_{t,1} - \bar{X}_{t,1}^{(n)}\right\|^2 \\
= & \left|Z_{t,1} - \bar{Z}_{t,1}^{(n)}\right| + \sum_{j=1}^{\max\{p,q\}} \left|h_{t+1-j,1} - \bar{h}_{t+1-j,1}^{(n)}\right|^2 \\
= & \left|\sum_{s=1}^{t} \frac{1}{2}\left(\bar{h}_{s,1}^{(n)} - h_{s,1}\right) + \sum_{s=1}^{t}\left(\sqrt{h_{s,1}} - \sqrt{\bar{h}_{s,1}^{(n)}}\right) \cdot \epsilon_{s,1}\right|^2 + \sum_{j=1}^{\max\{p,q\}} \left|h_{t+1-j,1} - \bar{h}_{t+1-j,1}^{(n)}\right|^2 \\
\leq & \; 2t \cdot \left(\sum_{s=1}^{t} \frac{1}{4}\left(\bar{h}_{s,1}^{(n)} - h_{s,1}\right)^2 + \sum_{s=1}^{t}\left(\sqrt{h_{s,1}} - \sqrt{\bar{h}_{s,1}^{(n)}}\right)^2 \cdot \epsilon_{s,1}^2\right) \\
& + \sum_{j=1}^{\max\{p,q\}} \left|h_{t+1-j,1} - \bar{h}_{t+1-j,1}^{(n)}\right|^2 .
\end{aligned}$$

Because of $a_0 > 0$ and a_j, $\hat{a}_{j,n}$, b_i, $\hat{b}_{i,n}$ nonnegative we know $\bar{h}_{s,1}^{(n)} \geq 0$ and $h_{s,1} \geq a_0$ for all $s \in \{0, \ldots, t\}$. Therefore

$$\left(\sqrt{h_{s,1}} - \sqrt{\bar{h}_{s,1}^{(n)}}\right)^2 = \left(\frac{h_{s,1} - \bar{h}_{s,1}^{(n)}}{\sqrt{h_{s,1}} + \sqrt{\bar{h}_{s,1}^{(n)}}}\right)^2 \leq \frac{1}{a_0}\left(h_{s,1} - \bar{h}_{s,1}^{(n)}\right)^2 .$$

Using this and the independence of $\epsilon_{s,1}$ from $h_{s,1}$ and $\bar{h}_{s,1}^{(n)}$ we get

$$\begin{aligned}
&\mathbf{E}\left\{\gamma_n^2 \cdot \left\|X_{t,1} - \bar{X}_{t,1}^{(n)}\right\|^2\right\} \\
&\leq 2t \cdot \left(\sum_{s=1}^{t} \frac{1}{4}\mathbf{E}\left\{\gamma_n^2 \cdot \left(\bar{h}_{s,1}^{(n)} - h_{s,1}\right)^2\right\} + \sum_{s=1}^{t} \mathbf{E}\left\{\gamma_n^2 \cdot \left(\sqrt{h_{s,1}} - \sqrt{\bar{h}_{s,1}^{(n)}}\right)^2\right\} \cdot \mathbf{E}\epsilon_{s,1}^2\right) \\
&\quad + \sum_{j=1}^{\max\{p,q\}} \mathbf{E}\left\{\gamma_n^2 \cdot \left|h_{t+1-j,1} - \bar{h}_{t+1-j,1}^{(n)}\right|^2\right\} \\
&\leq \left(\frac{t}{2} + \frac{2t}{a_0} + 1\right) \cdot \sum_{s=1}^{t} \mathbf{E}\left\{\gamma_n^2 \cdot \left(\bar{h}_{s,1}^{(n)} - h_{s,1}\right)^2\right\}.
\end{aligned}$$

So it remains to proof

$$\gamma_n^2 \cdot \mathbf{E}\left\{\left(h_{s,1} - \bar{h}_{s,1}^{(n)}\right)^2\right\} \to 0 \quad (n \to \infty) \tag{25}$$

for $s \in \{0, \dots, t\}$. We will do this by induction over s. By definition we have for $s = 0$ that

$$\gamma_n^2 \cdot \mathbf{E}\left\{ \left(h_{s,1} - \bar{h}_{s,1}^{(n)} \right)^2 \right\} = 0.$$

Let $s \in \mathbb{N}_0$ and assume (25) holds for s (and all smaller indices). Then we have for $s+1$, that

$$\begin{aligned}
& \gamma_n^2 \cdot \mathbf{E}\left\{ \left| h_{s+1,1} - \bar{h}_{s+1,1}^{(n)} \right|^2 \right\} \\
= \; & \gamma_n^2 \cdot \mathbf{E}\left\{ \left| a_0 - \hat{a}_{0,n} + \sum_{j=1}^{q} \left(a_j h_{s+1-j,1} (\epsilon_{s+1-j,1} - \lambda)^2 - \hat{a}_{j,n} \bar{h}_{s+1-j,1}^{(n)} (\epsilon_{s+1-j,1} - \hat{\lambda}_n)^2 \right) \right.\right. \\
& \qquad \left.\left. + \sum_{j=1}^{p} \left(b_j h_{s+1-j,1} - \hat{b}_{j,n} \bar{h}_{s+1-j,1}^{(n)} \right) \right|^2 \right\} \\
\leq \; & (1+p+q) \Bigg(\gamma_n^2 \left| a_0 - \hat{a}_{0,n} \right|^2 \\
& \qquad + \sum_{j=1}^{q} \gamma_n^2 \mathbf{E}\left\{ \left| a_j h_{s+1-j,1} (\epsilon_{s+1-j,1} - \lambda)^2 - \hat{a}_{j,n} \bar{h}_{s+1-j,1}^{(n)} (\epsilon_{s+1-j,1} - \hat{\lambda}_n)^2 \right|^2 \right\} \\
& \qquad + \sum_{j=1}^{p} \gamma_n^2 \mathbf{E}\left\{ \left| b_j h_{s+1-j,1} - \hat{b}_{j,n} \bar{h}_{s+1-j,1}^{(n)} \right|^2 \right\} \Bigg),
\end{aligned}$$

where the last inequality follows from the inequality of Jensen.

By the assumptions of the theorem we know

$$\gamma_n^2 \left| a_0 - \hat{a}_{0,n} \right|^2 \to 0 \quad (n \to 0).$$

Using $(a+b+c)^2 \leq 3a^2 + 3b^2 + 3c^2$ $(a, b, c \in \mathbb{R})$ and the independence of $\epsilon_{s+1-j,1}$ from $h_{s+1-j,1}$ and $\bar{h}_{s+1-j,1}^{(n)}$ we get

$$\begin{aligned}
& \mathbf{E}\left\{ \left| a_j h_{s+1-j,1} (\epsilon_{s+1-j,1} - \lambda)^2 - \hat{a}_{j,n} \bar{h}_{s+1-j,1}^{(n)} (\epsilon_{s+1-j,1} - \hat{\lambda}_n)^2 \right|^2 \right\} \\
& \leq 3a_j^2 \cdot \mathbf{E} h_{s+1-j,1}^2 \cdot \mathbf{E}\left\{ (\epsilon_{s+1-j,1} - \lambda)^2 - (\epsilon_{s+1-j,1} - \hat{\lambda}_n)^2 \right\} \\
& \quad + 3a_j^2 \cdot \mathbf{E}\left\{ \left(h_{s+1-j,1} - \bar{h}_{s+1-j,1}^{(n)} \right)^2 \right\} \cdot \mathbf{E}\left\{ (\epsilon_{s+1-j,1} - \hat{\lambda}_n)^2 \right\} \\
& \quad + 3 \cdot (a_j - \hat{a}_{j,n})^2 \cdot \mathbf{E}\left\{ \left(\bar{h}_{s+1-j,1}^{(n)} \right)^2 \right\} \cdot \mathbf{E}\left\{ (\epsilon_{s+1-j,1} - \hat{\lambda}_n)^2 \right\} \\
& = 3a_j^2 \cdot \mathbf{E} h_{s+1-j,1}^2 \cdot (\lambda^2 - \hat{\lambda}_n^2) + 3a_j^2 \cdot \mathbf{E}\left\{ \left(h_{s+1-j,1} - \bar{h}_{s+1-j,1}^{(n)} \right)^2 \right\} \cdot (1 + \hat{\lambda}_n^2)
\end{aligned}$$

$$+3 \cdot (a_j - \hat{a}_{j,n})^2 \cdot \mathbf{E}\left\{ \left(\bar{h}^{(n)}_{s+1-j,1} \right)^2 \right\} \cdot (1 + \hat{\lambda}_n^2),$$

where the last equality follows from $\mathbf{E}\epsilon_{s+1-j,1} = 0$ and $\mathbf{E}\epsilon^2_{s+1-j,1} = 1$.

From this, the induction hypothesis, $\mathbf{E}\{h^2_{s,1}\} < \infty$ (which follows by induction) and

$$\mathbf{E}\left\{ \left(\bar{h}^{(n)}_{s+1-j,1} \right)^2 \right\} \leq 2 \cdot \mathbf{E}\left\{ \left(h_{s+1-j,1} - \bar{h}^{(n)}_{s+1-j,1} \right)^2 \right\} + 2 \cdot \mathbf{E}\{h^2_{s+1-j,1}\}$$

for all s, j and the assumptions of the theorem we see

$$\gamma_n^2 \mathbf{E}\left\{ \left| a_j h_{s+1-j,1}(\epsilon_{s+1-j,1} - \lambda)^2 - \hat{a}_{j,n} \bar{h}^{(n)}_{s+1-j,1}(\epsilon_{s+1-j,1} - \hat{\lambda}_n)^2 \right|^2 \right\} \to 0 \quad (n \to \infty).$$

Similar arguments lead to

$$\gamma_n^2 \mathbf{E}\left\{ \left| b_j h_{s+1-j,1} - \hat{b}_{j,n} \bar{h}^{(n)}_{s+1-j,1} \right|^2 \right\} \to 0 \quad (n \to \infty),$$

from which we conclude the assertion. □

5.5 Proof of Theorem 3

The assertion depends only on the joint distribution of the random variables describing the discrete the price process used in the regression-based Monte Carlo method sampled at discrete points, so in order to prove the theorem we may assume w.l.o.g. that the random variables are generated in some special way. We do this in the same way for random variables describing the logarithms of the returns of the price process using the true parameter values. In both cases we use that values of a Poisson process sampled at discrete points can be generated as partial sums to a sequence of independent Poisson distributed random variables.

Let

$$\varepsilon_{t,i}, \xi_{t,i}, N_{t,i}, \hat{N}^{(n)}_{t,i} \qquad (i, t \in \mathbb{N}),$$

be independent random variables, where $\epsilon_{t,i}$ and $\xi_{t,i}$ are standard normally distributed, $N_{t,i}$ is Poisson distributed with parameter λt and $\hat{N}^{(n)}_{t,i}$ is Poisson distributed with parameter $|\lambda - \hat{\lambda}_n| t$.

At first we consider the case $\lambda < \lambda_n$. Because of the folding property of the Poisson distribution we can $\bar{N}^{(n)}_{t,i}$ write as

$$\bar{N}^{(n)}_{t,i} = N_{t,i} + \hat{N}^{(n)}_{t,i}.$$

For the logarithm of the price processes $X_{t,i}$ and $\bar{X}_{t,i}^{(n)}$ this means

$$X_{t,i} = \log(x_0) + \mu t + \sum_{s=1}^{t} \sigma \varepsilon_{s,i} + \sum_{j=1}^{N_{t,i}} (m + \delta \xi_{t,j})$$

and

$$\bar{X}_{t,i} = \log(x_0) + \hat{\mu}_n t + \sum_{s=1}^{t} \hat{\sigma}_n \varepsilon_{s,i} + \sum_{j=1}^{N_{t,i}+\hat{N}_{t,i}^{(n)}} \left(\hat{m}_n + \hat{\delta}_n \xi_{t,j}\right).$$

As in the proofs of Theorem 1 and Theorem 2 it's enough to show

$$\mathbf{E}\left\{\gamma_n^2 \left|X_{t,1} - \bar{X}_{t,1}^{(n)}\right|^2\right\} \to 0 \qquad (n \to \infty).$$

Jensen's inequality implies

$$\begin{aligned}
& \left|X_{t,1} - \bar{X}_{t,1}^{(n)}\right|^2 \\
= & \left|(\mu - \hat{\mu}_n)t + (\sigma - \hat{\sigma}_n) \sum_{s=1}^{t} \varepsilon_{s,1} + N_{t,1}(m - \hat{m}_n) + (\delta - \hat{\delta}_n) \sum_{j=1}^{N_{t,1}} \xi_{t,j}\right. \\
& \left. - \sum_{j=N_{t,1}+1}^{N_{t,1}+\hat{N}_{t,1}^{(n)}} \left(\hat{m}_n + \hat{\delta}_n \xi_{t,j}\right)\right|^2 \\
\leq & \; 5 \cdot \left((\mu - \hat{\mu}_n)^2 t^2 + (\sigma - \hat{\sigma}_n)^2 \left(\sum_{s=1}^{t} \varepsilon_{s,1}\right)^2 + (N_{t,1})^2 (m - \hat{m}_n)^2 \right. \\
& \left. + \left(\delta - \hat{\delta}_n\right)^2 \left(\sum_{j=1}^{N_{t,1}} \xi_{t,j}\right)^2 + \left(\sum_{j=N_{t,1}+1}^{N_{t,1}+\hat{N}_{t,1}^{(n)}} \left(\hat{m}_n + \hat{\delta}_n \xi_{t,j}\right)\right)^2 \right)
\end{aligned}$$

and therfore

$$\begin{aligned}
& \mathbf{E}\left\{\gamma_n^2 \left|X_{t,1} - \bar{X}_{t,1}^{(n)}\right|^2\right\} \\
\leq & \; 5 \cdot \gamma_n^2 \left((\mu - \hat{\mu}_n)^2 t^2 + (\sigma - \hat{\sigma}_n)^2 \mathbf{E}\left\{\left(\sum_{s=1}^{t} \varepsilon_{s,1}\right)^2\right\} + \mathbf{E}\left\{(N_{t,1})^2\right\} (m - \hat{m}_n)^2 \right. \\
& \left. + \left(\delta - \hat{\delta}_n\right)^2 \mathbf{E}\left\{\left(\sum_{j=1}^{N_{t,1}} \xi_{t,j}\right)^2\right\} + \mathbf{E}\left\{\left(\hat{N}_{t,1}^{(n)} \hat{m}_n + \hat{\delta}_n \sum_{j=N_{t,1}+1}^{N_{t,1}+\hat{N}_{t,1}^{(n)}} \xi_{t,j}\right)^2\right\} \right).
\end{aligned}$$

The random variables $\varepsilon_{1,1}, \ldots, \varepsilon_{t,1}$ are independent and standard normal distributed, so it holds

$$\mathbf{E}\left\{\left(\sum_{s=1}^{t} \varepsilon_{s,1}\right)^2\right\} = t,$$

and by definition of $N_{t,1}$ we have

$$\mathbf{E}\left\{(N_{t,1})^2\right\} = (\lambda t)^2 + \lambda t.$$

Because of the independence of $N_{t,1}, \xi_{t,1}, \xi_{t,2}, \ldots$ and the identical distribution of $\xi_{t,1}, \xi_{t,2}, \ldots$ this implies

$$\mathbf{E}\left\{\left(\sum_{j=1}^{N_{t,1}} \xi_{t,j}\right)^2\right\} = \mathbf{E}\left\{N_{t,1}\right\} \mathbf{Var}\left\{\xi_{t,1}\right\} + \mathbf{E}\left\{(N_{t,1})^2\right\} (\mathbf{E}\xi_{t,1})^2 = \lambda t$$

and so

$$\left(\delta - \hat{\delta}_n\right)^2 \mathbf{E}\left\{\left(\sum_{j=1}^{N_{t,1}} \xi_{t,j}\right)^2\right\} = \left(\delta - \hat{\delta}_n\right)^2 \lambda t.$$

From the independence of $N_{t,1}, \hat{N}_{t,1}^{(n)}, \xi_{t,1}, \xi_{t,2}, \ldots$ and because of $\mathbf{E}\left(\xi_{t,j}\right) = 0$ for all i, j one gets

$$\begin{aligned} \mathbf{E}\left\{\left(\hat{N}_{t,1}^{(n)} \hat{m}_n + \hat{\delta}_n \sum_{j=N_{t,1}+1}^{N_{t,1}+\hat{N}_{t,1}^{(n)}} \xi_{t,j}\right)^2\right\} &= \mathbf{E}\left\{\left(\hat{N}_{t,1}^{(n)} \hat{m}_n\right)^2\right\} + \mathbf{E}\left\{\left(\hat{\delta}_n \sum_{j=N_{t,1}+1}^{N_{t,1}+\hat{N}_{t,1}^{(n)}} \xi_{t,j}\right)^2\right\} \\ &= \left(|\lambda - \hat{\lambda}_n| t + |\lambda - \hat{\lambda}_n|^2 t^2\right) \cdot \hat{m}_n^2 + |\lambda - \hat{\lambda}_n| \cdot \hat{\delta}_n^2 \cdot t. \end{aligned}$$

To conclued this means in case of $\lambda < \lambda_n$, that

$$\begin{aligned} & \mathbf{E}\left\{\gamma_n^2 \left|X_{t,1} - \bar{X}_{t,1}^{(n)}\right|^2\right\} \\ \leq\ & 5 \cdot \gamma_n^2 \Bigg(\left(\mu - \hat{\mu}_n\right)^2 t^2 + \left(\sigma - \hat{\sigma}_n\right)^2 t + (\lambda^2 t^2 + \lambda t) \cdot (m - \hat{m}_n)^2 \\ & + \left(\delta - \hat{\delta}_n\right)^2 \lambda t + \left(|\lambda - \hat{\lambda}_n| t + |\lambda - \hat{\lambda}_n|^2 t^2\right) \cdot \hat{m}_n^2 + |\lambda - \hat{\lambda}_n| t \cdot \hat{\delta}_n^2 \Bigg). \end{aligned}$$

Similar argumentation implies for $\lambda \geq \lambda_n$, that

$$N_{t,i} = \bar{N}_{t,i}^{(n)} + \hat{N}_{t,i}^{(n)}.$$

So we get in this case

$$\begin{aligned} & \mathbf{E}\left\{\gamma_n^2 \left|X_{t,1} - \bar{X}_{t,1}^{(n)}\right|^2\right\} \\ \leq\ & 5 \cdot \gamma_n^2 \Bigg(\left(\mu - \hat{\mu}_n\right)^2 t^2 + \left(\sigma - \hat{\sigma}_n\right)^2 t + (\hat{\lambda}_n^2 t^2 + \hat{\lambda}_n t) \cdot (m - \hat{m}_n)^2 \end{aligned}$$

$$+ \left(\delta - \hat{\delta}_n\right)^2 \hat{\lambda}_n t + |\lambda - \hat{\lambda}_n| t \cdot \hat{\delta}_n^2 + \left(|\lambda - \hat{\lambda}_n| t + |\lambda - \hat{\lambda}_n|^2 t^2\right) \cdot m^2 \Bigg).$$

Alltogether this means

$$\begin{aligned}
& \mathbf{E}\left\{\gamma_n^2 \left\| X_{t,1} - \bar{X}_{t,1}^{(n)} \right\|^2 \right\} \\
\leq\ & 5 \cdot \gamma_n^2 \Bigg(\left(\mu - \hat{\mu}_n\right)^2 t^2 + \left(\sigma - \hat{\sigma}_n\right)^2 t + \max\left\{\hat{\lambda}_n^2 t^2 + \hat{\lambda}_n t, \lambda^2 t^2 + \lambda t\right\} \cdot \left(m - \hat{m}_n\right)^2 \\
& + \left(\delta - \hat{\delta}_n\right)^2 \max\left\{\lambda t, \hat{\lambda}_n t\right\} + |\lambda - \hat{\lambda}_n| t \cdot \hat{\delta}_n^2 \\
& + \left(|\lambda - \hat{\lambda}_n| t + |\lambda - \hat{\lambda}_n|^2 t^2\right) \cdot \max\left\{\hat{m}_n^2, m^2\right\} \Bigg) \\
\leq\ & 5 \cdot \gamma_n^2 \Bigg(\left(\mu - \hat{\mu}_n\right)^2 t^2 + \left(\sigma - \hat{\sigma}_n\right)^2 t + \left(\hat{\lambda}_n^2 t^2 + \hat{\lambda}_n t + \lambda^2 t^2 + \lambda t\right) \cdot \left(m - \hat{m}_n\right)^2 \\
& + \left(\delta - \hat{\delta}_n\right)^2 \left(\lambda t + \hat{\lambda}_n t\right) + |\lambda - \hat{\lambda}_n| t \cdot \hat{\delta}_n^2 \\
& + \left(|\lambda - \hat{\lambda}_n| t + |\lambda - \hat{\lambda}_n|^2 t^2\right) \cdot \left(\hat{m}_n^2 + m^2\right) \Bigg)
\end{aligned}$$

By the assumptions of the theorem this expression converges to zero for $n \to \infty$. Here we have used that by the mean-value-theorem we have

$$\left| \exp\left(\hat{m}_n + \frac{\hat{\delta}_n^2}{2}\right) - \exp\left(m + \frac{\delta^2}{2}\right) \right| \leq \left| \hat{m}_n + \frac{\hat{\delta}_n}{2} - m - \frac{\delta^2}{2} \right| \exp\left(|\hat{m}_n| + \frac{\hat{\delta}_n^2}{2} + m + \frac{\delta^2}{2}\right)$$

The result follows as in theorem 2. $\square$

References

[1] Belomestny, D. (2011). *Pricing Bermudan options using regression: optimal rates of convergence for lower estimates.* Finance and Stochastics **15**, pp. 655-683.

[2] Bollerslev, T. (1986). *Generalized autoregressive conditional heteroscedasticity.* Journal of Economics **31**, pp. 307-327

[3] Chow, Y.T., Robbins, H., and Siegmund, D. (1971). *Great expectations: The theory of optimal stopping.* Boston etc.: Houghton Mifflin Company.

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

[5] de Boor, C. (1987). *A Practical Guide to Splines*. Springer.

[6] Cont, R., and Tankov, P. (2004). *Financial modelling with jump processes*. Chapman & Hall / CRC .

[7] Duan, Jin-Chuan (1995). *The GARCH option pricing model*. Mathematical Finance **5**, pp. 13 - 32.

[8] Egloff, D. (2005). *Monte Carlo Algorithms for Optimal Stopping and Statistical Learning*. Annals of Applied Probability **15**, pp. 1-37.

[9] Egloff, D., Kohler, M., and Todorovic, N. (2007). *A dynamic look-ahead Monte Carlo algorithm for pricing American options*. Ann. Appl. Probab. **17**, No. 4, pp. 1138-1171.

[10] Fromkorth, A., and Kohler, M. (2011). *Analysis of least squares regression estimates in case of additional errors in the variables*. Journal of Statistical Planing and Inference **141**, pp. 172 - 188.

[11] Glasserman, P. (2004). *Monte Carlo methods in financial engineering*. Applications of Mathematics 53. New York, NY: Springer.

[12] Györfi, L., Kohler, M., Krzyżak, A., and Walk, H. (2002). *A Distribution-Free Theory of Nonparametric Regression*. Springer Series in Statistics, Springer.

[13] Karatzas, I., and Shreve, E.S. (1998). *Methods of Mathematical Finance*. Application of Mathematics - Stochastic Modelling and Applied Probability, Springer.

[14] Kohler, M. (2008). *A regression-based smoothing spline Monte Carlo Algorithm for pricing American options*. Advances in Statistical Analysis **92**, pp. 153-178.

[15] Kohler, M. (2010). *A review on regression-based Monte Carlo methods for pricing American options*. Devroye, Luc et al. (ed.), Recent developments in applied probability and statistics. Dedicated to the memory of Jürgen Lehn. Heidelberg: Physica-Verlag. 37-58 pp.

[16] Kohler, M., Krzyżak, A. (2009). *Pricing of American options in discrete time using least squares estimates with complexity penalties*. Preprint

[17] Kohler, M., Krzyżak, A., and Todorovich, N. (2010). *Pricing of high-dimensional american options by neural networks.* Mathematical Finance **20**, pp. 383-410.

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

[19] Merton, R. (1976). *Option pricing when underlying stock returns are discontinuous.* J. Financial Economics, **3**, pp. 125 - 144.

[20] Shiryayev, A.N. (1978). *Optimal Stopping Rules.* Applications of Mathematics, Springer.

[21] Tsitsiklis, J.N. and van Roy, B. (1999). *Optimal stopping of Markov processes: Hilbert space theory, approximation algorithms, and an application to pricing high-dimensional financial derivatives.* IEEE Transactions On Automatic Control **44**, pp. 1840-1851.