Analysis of least squares regression estimates in case of additional errors in the variables

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

September 30, 2009

Abstract

Estimation of a regression function from independent and identical distributed data is considered. The L_2 error with integration with respect to the design measure is used as error criterion. Upper bounds on the L_2 error of least squares regression estimates are presented, which bound the error of the estimate in case that in the sample given to the estimate the values of the independent and the dependent variables are pertubated by some arbitrary procedure. The bounds are applied to analyze regression-based Monte Carlo methods for pricing American options in case of errors in modelling the price process.

AMS classification: Primary 62G08; secondary 91B28, 60G40.

Key words and phrases: American option, consistency, least squares estimates, nonparametric regression, errors in variables, rate of convergence, regression-based Monte Carlo methods.

1 Introduction

Let (X, Y) be a $\mathbb{R}^d \times \mathbb{R}$ valued random vector with $\mathbf{E}\{Y^2\} < \infty$. In nonparametric regression we are interested in predicting Y after observing the value of X. More precisely,

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

Running title: *Analysis of LS regression estimates* we want to find a function f^* such that

$$\mathbf{E}\left\{|f^*(X)-Y|^2\right\}=\min_f \mathbf{E}\left\{|f(X)-Y|^2\right\}. \tag{1}$$

Denote the distribution of X by μ. For the regression function $m(x):=\mathbf{E}\{Y|X=x\}$ we have for each measurable function $f:\mathbb{R}^d\to\mathbb{R}$, that

$$\mathbf{E}\left\{|f(X)-Y|^2\right\}=\mathbf{E}\left\{|m(X)-Y|^2\right\}+\int|f(x)-m(x)|^2\,\mu(dx), \tag{2}$$

which implies that m is the solution of the minimization problem (1), $\mathbf{E}\left\{|m(X)-Y|^2\right\}$ is the minimum of (2) and the so called L_2 error $\int|f(x)-m(x)|^2\mu(dx)$ is the diffence between $\mathbf{E}\left\{|f(X)-Y|^2\right\}$ and $\mathbf{E}\left\{|m(X)-Y|^2\right\}$.

In the regression estimation problem the distribution of (X,Y) (and consequently m) is unknown. Given a sequence $\mathcal{D}_n=\{(X_1,Y_1),...,(X_n,Y_n)\}$ of independent observations of (X,Y), the goal is to construct an estimate $m_n(x)$=$m_n(x,\mathcal{D}_n)$ of $m(x)$ such that the L_2 error $\int|m_n(x)-m(x)|^2\mu(dx)$ is small. For a general introduction to regression estimation see, e.g., Györfi et al. (2002).

In this article we assume that we observe $\mathcal{D}_n$ only with some additional errors in the variables, i.e., the values of the variables are pertubated by some arbitrary procedure. In this context usually the problem is considered that the independent variable X can be observed only with additional random errors which have mean zero. More precisely, instead of X_i one observes $W_i=X_i+U_i$ for some random variables U_i which satisfy $\mathbf{E}\{U_i|X_i\}=0$, and the problem is to estimate the regression function from

$$\{(W_1,Y_1),\ldots,(W_n,Y_n)\}.$$

In the literature often estimates for the distribution of U_i are constructed and estimates of the regression function are defined by using the estimated distribution of U_i (see, e.g., Fan and Truong (1993), Caroll, Maca and Ruppert (1999), Delaigle and Meister (2007), Delaigle, Fan and Caroll (2009) and the references therein).

In this paper we consider a setting, where measurement errors occur simultaneously in the dependent and in the independent variables and where basically nothing is assumed on the nature of the measurement errors. In particular, the measurement errors do not have to be independent or identically distributed, and they do not need to have expectation zero. The only assumption we are making is that these measurement errors are somehow "small". Related results can be found in Kohler (2006) (where additional measurement errors occur only in the dependent variable) and in Kohler and Mehnert (2009) (where additional measurement errors occur only in the independent variable and the rate of convergence of least squares neural network regression estimates is analyzed in case of "small" measurement errors).

In the sequel we assume that we have given an arbitrary data set

$$\bar{\mathcal{D}}_n = \left\{ (\bar{X}_{1,n}, \bar{Y}_{1,n}), \ldots, (\bar{X}_{n,n}, \bar{Y}_{n,n}) \right\}, \tag{3}$$

where the average squared measurement errors

$$\frac{1}{n} \sum_{i=1}^{n} |Y_i - \bar{Y}_{i,n}|^2 \tag{4}$$

and

$$\frac{1}{n} \sum_{i=1}^{n} |X_i - \bar{X}_{i,n}|^2 \tag{5}$$

are small, and where $Y_1, \ldots, Y_n$ and $\bar{X}_{1,n}, \ldots, \bar{X}_{n,n}$ are independent given $X_1, \ldots, X_n$. Besides that we do not assume anything on the distribution of $\bar{\mathcal{D}}_n$, in particular the random variables in $\bar{\mathcal{D}}_n$ need not to be independent or identically distributed.

The basic idea behind the definition of our estimate is as follows: Since we assume that the measurement errors (4) and (5) are small, it is reasonable to estimate the L_2 risk $\mathbf{E}\{|f(X) - Y|^2\}$ of a Lipschitz continuous function f by the so-called empirical L_2 risk

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

computed with the aid of the data with measurement error, and to define least squares estimates as if no measurement errors are present by

$$\bar{m}_n(\cdot) = \arg \min_{f \in \mathcal{F}_n} \frac{1}{n} \sum_{i=1}^{n} |f(\bar{X}_{i,n}) - \bar{Y}_{i,n}|^2 \tag{6}$$

for some set $\mathcal{F}_n$ of Lipschitz continuous functions $f : \mathbb{R}^d \to \mathbb{R}$. Here $z = \arg\min_{x \in A} G(x)$ is an abbreviation for $z \in A$ and $G(z) = \min_{x \in A} G(x)$ and we assume for simplicity that the minimum in (6) exists, however we do not require it to be unique.

It is not clear, how the L_2 error of an arbitrary regression estimate is influenced by the additional errors in the data (3). Due to the fact that we do not assume that these additonal errors are in some sense in the mean zero, there is no chance to get rid of these errors, so these errors will necessarily increase the L_2 error of the estimate. Intuitively one can expect that measurement errors influence the L_2 error of the estimate not much as long as these measurement errors are small. In this paper we prove that as long as the underlying function space consists of Lipschitz continuous functions, this is indeed true for least squares estimates. This result is used to analyze regression-based Monte Carlo methods for pricing American options in case of errors in modelling the price process.

1.1 Notation

Throughout this paper we will use the following notations: $\mathbb{R}$, $\mathbb{Z}$, $\mathbb{N}$ denote the sets of real numbers, of integers, of positive integers, resp., $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$, $[a, b)$ denotes the half–open interval from a to b, $\log(x)$ is the natural logarithm of $x > 0$, $|u|$ is the Euclidean norm of $u \in \mathbb{R}^d$. For a function $f : \mathbb{R}^d \to \mathbb{R}$

$$\|f\|_\infty = \sup_{x \in \mathbb{R}^d} |f(x)|$$

denotes its supremum norm. $z = \arg\min_{x \in A} G(x)$ is an abbreviation for $z \in A$ and $G(z) = \min_{x \in A} G(x)$.

For $x_1, \dots, x_n \in \mathbb{R}^d$ set $x_1^n = (x_1, \dots, x_n)$. For x_1^n fixed we define the (pseudo-) distance $d_2(f, g)$ between to functions $f, g : \mathbb{R}^d \to \mathbb{R}$ by

$$d_2(f, g) = d_{2,x_1^n}(f, g) = \sqrt{\frac{1}{n} \sum_{i=1}^n |f(x_i) - g(x_i)|^2}.$$

An ϵ-cover of a set $\mathcal{F}$ of functions $f : \mathbb{R}^d \to \mathbb{R}$ (w.r.t. the distance d_2) is a finite collection of functions $f_1, \dots, f_k : \mathbb{R}^d \to \mathbb{R}$ with the property

$$\min_{1 \le j \le k} d_2(f, f_j) \le \epsilon \text{ for all } f \in \mathcal{F}.$$

Let $\mathcal{N}_2(\epsilon, \mathcal{F}, x_1^n)$ denote the size k of the smallest ϵ-cover of $\mathcal{F}$ w.r.t. the distance d_2, and set $\mathcal{N}_2(\epsilon, \mathcal{F}, x_1^n) = \infty$ if no finite sized ϵ-cover of $\mathcal{F}$ exists. $\mathcal{N}_2(\epsilon, \mathcal{F}, x_1^n)$ is called $L_2 - \epsilon$-covering number of $\mathcal{F}$ on x_1^n.

We say that random variables a_n, b_n satisfy $a_n = O_{\mathbf{P}}(b_n)$ if $\limsup_{n\to\infty} \mathbf{P}(a_n > c \cdot b_n) = 0$ for some finite constant c.

In order to avoid measurability problems in the case of uncountable collections of functions, we assume throughout this paper that the function classes in the definition of our least squares estimates are permissible in the sense of Pollard (1984), Appendix C. This mild measurability condition is satisfied for most classes of functions used in application.

1.2 Outline

The main result is fomulated in Section 2. An application in financial mathematics in the context of regression-based Monte Carlo methods for pricing American options is described in Section 3. The proofs are given in Section 4.

2 Main results

Our main result is the following theorem.

Theorem 1. *Let $(X,Y),(X_1,Y_1),\ldots,(X_n,Y_n)$ be independent and identically distributed $\mathbb{R}^d \times \mathbb{R}$ valued random vectors with $\mathbf{E}\left\{Y^2\right\} < \infty$. Assume that $Y - m(X)$ is sub-Gaussian in the sense that*

$$K^2 \mathbf{E}\left\{ \exp\left(\frac{(Y-m(X))^2}{K^2} - 1 \right) \middle| X \right\} \leq \sigma_0^2 \text{ almost surely} \tag{7}$$

for some $K, \sigma_0 > 0$. Let $\beta_n \geq \beta > 1$ and assume that the regression function is bounded in absolute value by β. Let $\mathcal{F}_n$ be a set of functions $f : \mathbb{R}^d \to [-\beta_n, \beta_n]$ that are Lipschitz-continuous with Lipschitz-constant L_n in the sense, that for all $f \in \mathcal{F}_n$ and for all $x, y \in \mathbb{R}^d$ we have

$$|f(x) - f(y)| \leq L_n |x - y| . \tag{8}$$

Given an arbitrary dataset

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

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

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

Then there exist constants $c_1, c_2 > 0$ depending only on σ_0 and K such that for any δ_n which satisfies

$$\delta_n \to 0 \quad (n \to \infty) \quad \textit{and} \quad \frac{n \cdot \delta_n}{\beta_n^2} \to \infty \quad (n \to \infty) \tag{9}$$

and

$$\begin{aligned} \sqrt{n}\frac{\delta}{\beta_n^2} \quad \geq \quad c_1 \int_{c_2\delta/\beta_n^2}^{\sqrt{\delta}} \Bigg(\log \mathcal{N}_2 \Bigg(\frac{u}{4\beta_n}, \{f-g : f \in \mathcal{F}_n, \\ \frac{1}{n}\sum_{i=1}^n |f(x_i) - g(x_i)|^2 \leq \delta\Bigg\}, x_1^n\Bigg)\Bigg)^{\frac{1}{2}} du \end{aligned} \tag{10}$$

for all $\delta \geq \frac{\delta_n}{4}$, all $x_1, \ldots, x_n \in \mathbb{R}^d$ and all $g \in \mathcal{F}_n \cup \{m\}$ we have

$$\int |\bar{m}_n(x) - m(x)|^2 \mu(dx) = O_{\mathbf{P}}(Z_n) \quad \textit{and} \quad \frac{1}{n}\sum_{i=1}^n |\bar{m}_n(\bar{X}_{i,n}) - m(X_i)|^2 = O_{\mathbf{P}}(Z_n)$$

where

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

Remark 1. The proof of the theorem (which can be found in Section 4) shows that if we have $X_i = \bar{X}_{i,n}$ for alle $i \in \{1, \ldots, n\}$, we don't need the Lipschitz condition anymore. In this case we get exactly Theorem 1 in Kohler (2006).

Remark 2. In the proof we show a stronger result, namely that for c_3 sufficiently large we have for any $n \in \mathbb{N}$

$$\mathbf{P}\left\{\int |\bar{m}_n(x) - m(x)|^2 \mu(dx) > c_3 \cdot Z_n\right\} \leq c_3 \cdot \exp\left(-c_3 \cdot \frac{n \cdot \delta_n}{\beta_n^2}\right)$$

and

$$\mathbf{P}\left\{\frac{1}{n}\sum_{i=1}^n |\bar{m}_n(\bar{X}_{i,n}) - m(X_i)|^2 > c_3 \cdot Z_n\right\} \leq c_3 \cdot \exp\left(-c_3 \cdot \frac{n \cdot \delta_n}{\beta_n^2}\right)$$

If we choose in the above theorem $\mathcal{F}_n$ as a subset of a finite dimensional linear vector space, the entropy condition (10) can be simplified and we get

Corollary 1. *Assume that* (X,Y), $(X_1,\bar{X}_{1,n},Y_1,\bar{Y}_{1,n})$, $\dots (X_n,\bar{X}_{n,n},Y_n,\bar{Y}_{n,n})$ *satisfy the assumptions of Theorem 1, let* $\mathcal{F}_n$ *be a set of functions* $f : \mathbb{R}^d \to [-\beta_n,\beta_n]$ *that are Lipschitz-continuous with Lipschitz-constant* L_n *and assume that* $\mathcal{F}_n$ *is a subset of a linear vector space of dimension* K_n*. Let the estimate be defined as in Theorem 1. Then*

$$\begin{aligned}\int |\bar{m}_n(x)-m(x)|^2\mu(dx) &= O_{\mathbf{P}}\Bigg\{\frac{1}{n}\sum_{i=1}^n \left|Y_i-\bar{Y}_{i,n}\right|^2 + L_n^2\cdot\frac{1}{n}\sum_{i=1}^n\left|X_i-\bar{X}_{i,n}\right|^2 \\ &\qquad +\beta_n^5\cdot\frac{K_n}{n}+\inf_{f\in\mathcal{F}_n}\int|f(x)-m(x)|^2\mu(dx)\Bigg\}.\end{aligned}$$

Proof. The result follows from Theorem 1 and the bound

$$\mathcal{N}_2\left(v,\left\{f\in\mathcal{F}:\frac{1}{n}\sum_{i=1}^n|f(z_i)|^2\le R^2\right\},z_1^n\right)\le\left(\frac{4R+v}{v}\right)^{\dim(\mathcal{F})}.$$

for the covering number of a linear vector space $\mathcal{F}$ of dimension $dim(\mathcal{F})$ (cf. Corollary 2.6 and Example 9.3.1 in van de Geer (2000), or Lemma 9.3 and proof of Lemma 19.1 in Györfi et al. (2002)). □

In the sequel we demonstrate the usefullness of the above corollary by applying it to least squares spline estimates. Choose $M\in\mathbb{N}_0$, $K_n\in\mathbb{N}$, $A,B\in\mathbb{R}$ with $A<B$ and set $u_k=k\cdot(B-A)/K_n$ for $k\in\mathbb{Z}$. Let $\{B_{j,M,K} : j=1,...,K+M\}$ 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 will be defined as subspaces of

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

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 Lipschitz continuous. More precisely, we set

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

By standard results on B-splines and its derivatives (cf., e.g., Lemmas 14.4 and 14.6 in Györfi et al. (2002)) we have that each function in $S_{K_n,M,\beta_n}([A,B])$ is bounded in absolute value by β_n and Lipschitz continuous with Lipschitz constant γ_n.

Let L_M be the norm of the quasi interpoland in the proof of Theorem 14.4 in Györfi et al. (2002) in case of equidistant knots and $Q_{j,k}$ chosen independent of k. In our next corollary we will define the parameter β_n of the spline space by $\beta_n = L_M \cdot \|m\|_\infty$, which implies that the quasi interpoland of the regression function is contained in our spline space provided the regression function is Lipschitz continuous.

Using this function space in Corollary 1 we get

Corollary 2. *Assume that* (X,Y), $(X_1,\bar{X}_{1,n},Y_1,\bar{Y}_{1,n})$, $\dots(X_n,\bar{X}_{n,n},Y_n,\bar{Y}_{n,n})$ *satisfy the assumptions of Theorem 1, and that, in addition,* $X\in[0,1]$ *a.s. and that the regression functions is p–times continuously differentiable on* $[0,1]$ *for some* $p\geq 1$. *Set*

$$K_n = \lceil n^{1/(2p+1)} \rceil$$

and

$$\beta_n = L_M \cdot B,$$

where L_M *is defined as above and* B *is a bound on the supremum norm of the regression function, and assume*

$$\gamma_n \to \infty \quad (n\to\infty).$$

Set $\mathcal{F}_n = S_{K_n,M,\beta_n,\gamma_n}([0,1])$ *and let the estimate be defined as in Theorem 1. Then*

$$\int |\bar{m}_n(x)-m(x)|^2\mu(dx) \quad = \quad O_{\mathbf{P}}\left\{\frac{1}{n}\sum_{i=1}^n \left|Y_i-\bar{Y}_{i,n}\right|^2 + \gamma_n^2\cdot\frac{1}{n}\sum_{i=1}^n\left|X_i-\bar{X}_{i,n}\right|^2 + n^{-\frac{2p}{2p+1}}\right\}.$$

Proof. It follows from Theorem 14.3 and the proof of Theorem 14.4 in Györfi et al. (2002) that for n sufficiently large we have

$$\inf_{f\in S_{K_n,M,\beta_n,\gamma_n}([0,1])}\int |f(x)-m(x)|^2\mu(dx) \leq c_4\cdot\left(\frac{1}{K_n}\right)^{2p}.$$

From this we get the assertion by an application of Corollary 1. □

Remark 3. In case $X_i=\bar{X}_{i,n}$ and $Y_i=\bar{Y}_{i,n}$ for all $i\in\{1,\dots,n\}$ it follows from Stone (1982) that the rate of convergence in Corollary 2 is optimal.

Remark 4. In any application the parameters of the spline space have to be chosen using the given data only. This can be done e.g. by splitting of the sample. Here the sample $\bar{\mathcal{D}}_n$ is divided into a learning sample consisting of the first n_l data points and a testing sample of size $n_t = n - n_l$ (e.g. with $n_l \approx n/2 \approx n_t$). Given a finite list $\mathcal{P}_n$ of parameters and for each parameter $k \in \mathcal{P}_n$ a set of functions $\mathcal{F}_{n,k}$ (which we assume to be bounded in absolute value by β_n and to be Lipschitz continuous with Lipschitz constant L_n) we define estimates

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

and choose the value of the parameter by minimizing the error on the testing data, i.e. we set

$$\bar{m}_n(\cdot) = \bar{m}_{n,\hat{k}}(\cdot)$$

where

$$\hat{k} = \arg \min_{k \in \mathcal{P}_n} \frac{1}{n_t} \sum_{i=n_l+1}^{n_l+n_t} |\bar{m}_{n,k}(\bar{X}_{i,n}) - \bar{Y}_{i,n}|^2.$$

If we assume that the learning data $\{(\bar{X}_{i,n}, \bar{Y}_{i,n}) : i = 1, \ldots, n_l\}$ is independent from $\{(X_j, Y_j) : i = n_l + 1, \ldots, n\}$ then we can apply Theorem 1 conditioned on the learning data, bound the covering number in (10) by the finite cardinality of the parameter set $\mathcal{P}_n$ and conclude

$$\int |\bar{m}_n(x) - m(x)|^2 \mu(dx) = O_{\mathbf{P}} \Bigg\{ \frac{1}{n_t} \sum_{i=n_l+1}^{n} \left|Y_i - \bar{Y}_{i,n}\right|^2 + L_n^2 \cdot \frac{1}{n_t} \sum_{i=n_l+1}^{n} \left|X_i - \bar{X}_{i,n}\right|^2$$
$$+ \beta_n^4 \cdot \frac{\log |\mathcal{P}_n|}{n_t} + \min_{k \in \mathcal{P}_n} \int |\bar{m}_{n,k}(x) - m(x)|^2 \mu(dx) \Bigg\}.$$

3 Application in option pricing

In the sequel we describe how our main result can be used to analyze regression-based Monte Carlo methods for pricing American options in discrete time in case of errors in modelling the price process of the underlying asset.

An American option can be excercised at any time up to maturity. In complete and arbitrage-free markets the price of an American option with maturity T is given by the value of the optimal stopping problem (cf., e.g., Karatzas and Shreve (1998))

$$V_0 = \sup_{t \in \mathcal{T}[0,T]} \mathbf{E} \left\{ f_\tau \left(X_\tau \right) \right\}.$$

Here f_t denotes the discounted payoff function at time $t \in [0, T]$ (e.g.,

$$f_t(x) = e^{-r \cdot t} \cdot \max\{K - x, 0\}$$

in case of a put option with strike K and disounting factor $e^{-r \cdot t}$) and the $\mathbb{R}^d$-valued stochastic process $(X_t)_{0 \le t \le T}$ models the underlying risk factors such as the stock value of the underlying. $\mathcal{T}[0, T]$ is the class of all $[0, T]$-valued stopping times, i.e. $\tau \in \mathcal{T}[0, T]$ is a measurable function of $(X_t)_{0 \le t \le T}$ with values in $[0, T]$ with the property that for any $r \in [0, T]$ the event $[\tau \le r]$ is contained in the sigma algebra $\mathcal{F}_r = \mathcal{F}((X_s)_{0 \le s \le r})$ generated by $(X_s)_{0 \le s \le r}$.

The first step to treat this problem numerically is to consider only discrete time steps. In terms of finance this means that we approximate the price of an American option by a Bermudan option. In the sequel we assume that $X_0, X_1, \ldots, X_T$ is a discrete Markov process (maybe with augmented space state in order to ensure the Markovian property) and the price of our Bermudan option is now given by

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

where $\mathcal{T}(0, \ldots, T)$ is the class of alle $\{0, \ldots, T\}$-valued stopping times, and τ^* is the optimal stopping time.

One way to compute the price of such an option numerically, which is especially useful in case of option based on several underlying assets, is to compute so-called continuation values $q_t(x)$ which describe the value of the option at time t in case that $X_t = x$ has been observed subject to the constraint of holding the option rather than exercising it.

More precisely,

$$q_t(x) = \sup_{\tau \in \mathcal{T}(t+1, t+2, \ldots, T)} \mathbf{E}\left\{f_\tau(X_\tau)\right\}$$

where $\mathcal{T}(t + 1, \ldots, T)$ is the set of all stopping times with values in $\{t + 1, \ldots, T\}$, and $q_T(x) = 0$ $(x \in \mathbb{R}^d)$. The general theory of optimal stopping (cf., eg., Shiryayev (1978)) implies that once we know the contnuation values q_t, we can compute the optimal stopping time τ^* via

$$\tau^* = \min\left\{t \in \{0, \ldots, T\} : f_t(X_t) \ge q_t(X_t)\right\}.$$

In order to compute the continuation values a regression representation of $q_t(x)$ like

$$q_t(x) = \mathbf{E}\left\{\max\left\{f_{t+1}(X_{t+1}), q_{t+1}(X_{t+1})\right\}\middle|\, X_t = x\right\} \tag{11}$$

(cf. Tsitsiklis and van Roy (1999), see also Longstaff and Schwartz (2001) or Egloff (2005) for additional regression representations) can be used. Typically in applications, the underlying distributions are rather complicated and therefore it is not clear how the conditional expectation in (11) can be calculated. The idea behind regression-based Monte Carlo methods is that the conditional expectations in (11) can be computed numerically by applying recursively a regression estimate to a sample of

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

where $\hat{q}_{n,t+1}$ is an estimate of q_{t+1} computed in the step before and $\hat{q}_{n,T} = 0$. In the context of linear regression this was proposed by Tsitsiklis an van Roy (1999) and Longstaff and Schwartz (2001), and based on a regression representation for the so-called value function $v_t(x) = \max\{f_t(x), q_t(x)\}$ this was proposed in Carrier (1996). Nonparametric regression estimates of continuation values have been investigated in Egloff (2005), Egloff, Kohler and Todorovich (2007), Kohler, Krzyżak and Todorovich (2006), Kohler (2008), Belomestny (2009) and Kohler and Krzyżak (2009).

The estimates there are applied to a sample

$$\left\{\left(X_{t,i}, \max\left\{f_{t+1}(X_{t+1,i}), \hat{q}_{n,t+1}(X_{t+1,i})\right\}\right) : i = 1, \ldots, n\right\}$$

which can be considered as a sample of

$$\left(X_t, \max\left\{f_{t+1}(X_{t+1}), q_{t+1}(X_{t+1})\right\}\right)$$

with additional measurement errors in the dependent variable.

In order to apply such methods in practice a model for the price process has to be chosen. The most simple case is a Black-Scholes model, where (in case $d = 1$)

$$X_t = x_0 \cdot \exp\left(\left(r - \frac{1}{2}\sigma^2\right) t + \sigma W_t\right). \tag{12}$$

Here $\{W_t : t \in [0, T]\}$ is a Wiener process, r is the (riskless) interest rate and $\sigma > 0$ is the volatility of the asset. As long as the interest rate r is the same as the interest rate used for discounting (which is necessary in order to get an arbitrage free market) the corresponding price will not depend on r. But critical for the price is the choice of the volatility σ. Its value has to be estimated from observed data in the past, so in that model we use in fact

$$\bar{X}_{t,n} = x_0 \cdot \exp\left(\left(r - \frac{1}{2}\hat{\sigma}_n^2\right) t + \hat{\sigma}_n W_t\right) \tag{13}$$

for some estimate $\hat{\sigma}_n$ of σ.

Clearly, in an application the value of $\hat{\sigma}_n$ will be not equal to σ. This rises the question how robust the estimation procedure is with respect to errors in σ. In the sequel we show for suitably defined regression-based Monte Carlo methods that the price computed with $\hat{\sigma}_n$ instead of σ tends to the true price in case of $\hat{\sigma}_n$ tending to σ.

More precisely, assume that we have estimates $\hat{\sigma}_n$ of σ available. On the probability space where these estimates are defined there exists independent Wiener processes $(W_{t,i})_{t\in[0,T]}$ for $i \in \mathbb{N}$, which are independent of all data used in the estimate $\hat{\sigma}_n$ (i.e., the Wiener processes are independent of the estimates). Set

$$X_{t,i} = x_0 \cdot \exp\left(\left(r - \frac{1}{2}\sigma^2\right) t + \sigma W_{t,i}\right) \tag{14}$$

and

$$\bar{X}_{t,i} = x_0 \cdot \exp\left(\left(r - \frac{1}{2}\hat{\sigma}_n^2\right) t + \hat{\sigma}_n W_{t,i}\right). \tag{15}$$

We consider $\bar{X}_{t,i}$ as an observation of $X_{t,i}$ with additional measurement errors.

In the sequel we define regression-based Monte Carlo estimates of the continuation values depending on $(\bar{X}_{t,i})_{t\in\{0,1,\dots,T\}}$ $(i \in \{1,\dots,n\})$.

We start with

$$\hat{q}_{n,T}(x) = 0 \quad (x \in \mathbb{R}).$$

Given an estimate $\hat{q}_{n,t+1}$ of q_{t+1} for some $t \in \{0,\dots,T-1\}$ we define an estimate $\hat{q}_{n,t}$ of q_t as follows:

Set

$$\bar{X}_{i,n} = \bar{X}_{t,i}$$

and

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

$(i = 1, \ldots, n)$ and define $\hat{q}_{n,t}$ by

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

where

$$\mathcal{F}_n = S_{K_n, M, \beta_n, \gamma_n}([-A_n, A_n])$$

is the spline space introduced in Corollary 2 and $K_n \in \mathbb{N}, M \in \mathbb{N}_0, \beta_n > 0, \gamma_n > 0$ and $A_n > 0$ are parameters of the estimate.

Finally we estimate the price

$$V_0 = \mathbf{E}\left\{\max\{f_0(X_0), q_0(X_0)\}\right\} = \int \max\{f_0(x_0), q_0(x_0)\} d\mathbf{P}_{X_0}(x_0)$$

of the option by

$$\hat{V}_{0,n} = \int \max\{f_0(x_0), \hat{q}_{n,0}(x_0)\} d\mathbf{P}_{X_0}(x_0).$$

If we consider for fixed $t \in \{0, \ldots, T-1\}$ the sample

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

as a sample of $(X_t, \max\{f_{t+1}(X_{t+1}), q_{t+1}(X_{t+1})\})$ with additional measurement errors in the variables we can conclude from Theorem 1:

Theorem 2. *Let $K_n \in \mathbb{N}, M \in \mathbb{N}_0, \beta_n > 0, \gamma_n > 0$ and $A_n > 0$ such that*

$$K_n \to \infty \quad (n \to \infty), \tag{16}$$

$$A_n \to \infty \quad (n \to \infty), \tag{17}$$

$$\beta_n \to \infty \quad (n \to \infty), \tag{18}$$

$$\gamma_n \to \infty \quad (n \to \infty), \tag{19}$$

and

$$\frac{A_n \cdot \beta_n^5 \cdot K_n}{n} \to 0 \quad (n \to \infty) \tag{20}$$

for some $\delta > 0$. Assume that the discounted payoff function f_t is bounded and Lipschitz continuous and that the price process of the underlying stock is given by (12). Let $\hat{\sigma}_n$ be an *estimate of the volatility σ in the model (12) (based on data observed in the past, which we assume to be independent of all data used in the Monte Carlo simulation) which satisfies*

$$\gamma_n \cdot (\hat{\sigma}_n - \sigma_0) \to 0 \quad a.s. \tag{21}$$

Let the estimates of the continuation values and the price of the option be defined as above. Then we have for all $t \in \{0, \ldots, T-1\}$

$$\int |\hat{q}_{n,t}(x) - q_t(x)|^2 \mathbf{P}_{X_t}(dx) \to 0 \quad a.s. \tag{22}$$

and, in addition, we have

$$\hat{V}_{0,n} \to V_0 \quad a.s. \tag{23}$$

4 Proofs

Throughout the proofs we will use the abbreviation $\bar{X}_i = \bar{X}_{i,n}$ and $\bar{Y}_i = \bar{Y}_{i,n}$.

4.1 Preliminarien to the proof of Theorem 1

We start with a deterministic lemma. Let $x_1, \ldots, x_n, \bar{x}_1, \ldots, \bar{x}_n \in \mathbb{R}^d, y_1, \ldots, y_n, \bar{y}_1, \ldots, \bar{y}_n \in \mathbb{R}$. Let $\mathcal{G}$ be a set of functions $g : \mathbb{R}^d \to \mathbb{R}$ and for $g \in \mathcal{G}$ define

$$\bar{g}_n = \arg\min_{g \in \mathcal{G}} \left(\frac{1}{n} \sum_{i=1}^{n} |\bar{y}_i - g(\bar{x}_i)|^2 \right).$$

Let $m : \mathbb{R}^d \to \mathbb{R}$ be a fixed function and let $h \in \mathcal{G}$.

Lemma 1. *Assume*

$$\frac{1}{n} \sum_{i=1}^{n} |\bar{g}_n(\bar{x}_i) - m(x_i)|^2 \geq 3 \cdot \frac{1}{n} \sum_{i=1}^{n} |h(\bar{x}_i) - m(x_i)|^2 + 128 \cdot \frac{1}{n} \sum_{i=1}^{n} |y_i - \bar{y}_i|^2 + \delta \tag{24}$$

for some $\delta \geq 0$. Then

$$\frac{1}{n} \sum_{i=1}^{n} (y_i - m(x_i)) \cdot (\bar{g}_n(\bar{x}_i) - h(\bar{x}_i)) \geq \frac{1}{16} \cdot \left(\frac{1}{n} \sum_{i=1}^{n} |\bar{g}_n(\bar{x}_i) - h(\bar{x}_i)|^2 \right) + \frac{\delta}{8}. \tag{25}$$

Proof. By definition of the estimate and because of $h \in \mathcal{G}$ we have

$$\frac{1}{n} \sum_{i=1}^{n} |\bar{y}_i - \bar{g}_n(\bar{x}_i)|^2 \leq \frac{1}{n} \sum_{i=1}^{n} |\bar{y}_i - h(\bar{x}_i)|^2$$

hence

$$
\begin{aligned}
& \frac{1}{n}\sum_{i=1}^n |\bar{y}_i - m(x_i)|^2 + \frac{2}{n}\sum_{i=1}^n (\bar{y}_i - m(x_i)) \cdot (m(x_i) - \bar{g}_n(\bar{x}_i)) + \frac{1}{n}\sum_{i=1}^n |m(x_i) - \bar{g}_n(\bar{x}_i)|^2 \\
\leq \; & \frac{1}{n}\sum_{i=1}^n |\bar{y}_i - m(x_i)|^2 + \frac{2}{n}\sum_{i=1}^n (\bar{y}_i - m(x_i)) \cdot (m(x_i) - h(\bar{x}_i)) + \frac{1}{n}\sum_{i=1}^n |m(x_i) - h(\bar{x}_i)|^2,
\end{aligned}
$$

which implies

$$
\begin{aligned}
& \frac{1}{n}\sum_{i=1}^n |\bar{g}_n(\bar{x}_i) - m(x_i)|^2 - \frac{1}{n}\sum_{i=1}^n |h(\bar{x}_i) - m(x_i)|^2 \\
\leq \; & \frac{2}{n}\sum_{i=1}^n (\bar{y}_i - m(x_i)) \cdot (\bar{g}_n(\bar{x}_i) - h(\bar{x}_i)) \\
= \; & \frac{2}{n}\sum_{i=1}^n (\bar{y}_i - y_i) \cdot (\bar{g}_n(\bar{x}_i) - h(\bar{x}_i)) + \frac{2}{n}\sum_{i=1}^n (y_i - m(x_i)) \cdot (\bar{g}_n(\bar{x}_i) - h(\bar{x}_i)) \\
=: \; & T_1 + T_2.
\end{aligned}
$$

We show next that $T_1 \leq T_2$. Assume to the contrary that this is not true. Then

$$
\begin{aligned}
& \frac{1}{n}\sum_{i=1}^n |\bar{g}_n(\bar{x}_i) - m(x_i)|^2 - \frac{1}{n}\sum_{i=1}^n |h(\bar{x}_i) - m(x_i)|^2 \\
< \; & \frac{4}{n}\sum_{i=1}^n (\bar{y}_i - y_i) \cdot (\bar{g}_n(\bar{x}_i) - h(\bar{x}_i)) \\
\leq \; & 4 \cdot \sqrt{\frac{1}{n}\sum_{i=1}^n (\bar{y}_i - y_i)^2} \cdot \sqrt{\frac{1}{n}\sum_{i=1}^n (\bar{g}_n(\bar{x}_i) - h(\bar{x}_i))^2} \\
\leq \; & 4 \cdot \sqrt{\frac{1}{n}\sum_{i=1}^n (\bar{y}_i - y_i)^2} \cdot \sqrt{2\frac{1}{n}\sum_{i=1}^n |\bar{g}_n(\bar{x}_i) - m(x_i)|^2 + 2\frac{1}{n}\sum_{i=1}^n |h(\bar{x}_i) - m(x_i)|^2}.
\end{aligned}
$$

Using (24) we see that the left-hand side of the above inequality is bounded from below by

$$
\begin{aligned}
& \frac{1}{2} \cdot \left(\frac{1}{n}\sum_{i=1}^n |\bar{g}_n(\bar{x}_i) - m(x_i)|^2 \right) \\
& + \frac{1}{2} \cdot \left(3 \cdot \frac{1}{n}\sum_{i=1}^n |h(\bar{x}_i) - m(x_i)|^2 + 128 \cdot \frac{1}{n}\sum_{i=1}^n |y_i - \bar{y}_i|^2 + \delta \right) \\
& - \frac{1}{n}\sum_{i=1}^n |h(\bar{x}_i) - m(x_i)|^2 \\
\geq \; & \frac{1}{2} \cdot \left(\frac{1}{n}\sum_{i=1}^n |\bar{g}_n(\bar{x}_i) - m(x_i)|^2 + \frac{1}{n}\sum_{i=1}^n |h(\bar{x}_i) - m(x_i)|^2 \right) + \frac{\delta}{2}, \qquad (26)
\end{aligned}
$$

which implies

$$
\begin{aligned}
& \frac{1}{n}\sum_{i=1}^{n}|\bar{g}_n(\bar{x}_i)-m(x_i)|^2+\sum_{i=1}^{n}|h(\bar{x}_i)-m(x_i)|^2 \\
< \quad & 8\cdot\sqrt{2}\cdot\sqrt{\frac{1}{n}\sum_{i=1}^{n}|\bar{g}_n(\bar{x}_i)-m(x_i)|^2+\sum_{i=1}^{n}|h(\bar{x}_i-m(x_i)|^2}\cdot\sqrt{\frac{1}{n}\sum_{i=1}^{n}|y_i-\bar{y}_i|^2}
\end{aligned}
$$

i.e.,

$$
\frac{1}{n}\sum_{i=1}^{n}|\bar{g}_n(\bar{x}_i)-m(x_i)|^2+\frac{1}{n}\sum_{i=1}^{n}|h(\bar{x}_i-m(x_i)|^2<128\cdot\frac{1}{n}\sum_{i=1}^{n}|y_i-\bar{y}_i|^2.
$$

But this is a contradiction to (24), so we have indeed proved $T_1\leq T_2$.

As a consequence we can conclude

$$
\begin{aligned}
& \frac{4}{n}\sum_{i=1}^{n}(y_i-m(x_i))\cdot(\bar{g}_n(\bar{x}_i)-h(\bar{x}_i)) \\
\geq \quad & \frac{1}{n}\sum_{i=1}^{n}|\bar{g}_n(\bar{x}_i)-m(x_i)|^2-\frac{1}{n}\sum_{i=1}^{n}|h(\bar{x}_i)-m(x_i)|^2. \qquad (27)
\end{aligned}
$$

As before we can bound the right-hand side of (27) from below by (26) and get

$$
\begin{aligned}
& \frac{4}{n}\sum_{i=1}^{n}(y_i-m(x_i))(\bar{g}_n(\bar{x}_i)-h(\bar{x}_i)) \\
\geq \quad & \frac{1}{2}\cdot\left(\frac{1}{n}\sum_{i=1}^{n}|\bar{g}_n(\bar{x}_i)-m(x_i)|^2+\frac{1}{n}\sum_{i=1}^{n}|h(\bar{x}_i)-m(x_i)|^2\right)+\frac{\delta}{2}. \qquad (28)
\end{aligned}
$$

Because of $a^2/2-b^2\leq(a-b)^2$ $(a,b\in\mathbb{R})$ we have

$$
\begin{aligned}
& \frac{1}{2}\cdot\frac{1}{n}\sum_{i=1}^{n}|\bar{g}_n(\bar{x}_i)-h(\bar{x}_i)|^2-\frac{1}{n}\sum_{i=1}^{n}|h(\bar{x}_i)-m(x_i)|^2 \\
\leq \quad & \frac{1}{n}\sum_{i=1}^{n}\left(|\bar{g}_n(\bar{x}_i)-h(\bar{x}_i)|-|h(\bar{x}_i)-m(x_i)|\right)^2 \\
\leq \quad & \frac{1}{n}\sum_{i=1}^{n}|\bar{g}_n(\bar{x}_i)-m(x_i)|^2.
\end{aligned}
$$

Using this we can bound the right-hand side of (28) from below by

$$
\frac{1}{2}\cdot\left(\frac{1}{2}\cdot\frac{1}{n}\sum_{i=1}^{n}|\bar{g}_n(\bar{x}_i)-h(\bar{x}_i)|^2\right)+\frac{\delta}{2}=\frac{1}{4}\frac{1}{n}\sum_{i=1}^{n}|\bar{g}_n(\bar{x}_i)-m(\bar{x}_i)|^2+\frac{\delta}{2}
$$

Summing up the above results we get the desired inequality. $\square$

Next we work conditionally on $X_1,\ldots,X_n,\bar{X}_1,\ldots,\bar{X}_n$ and measure the error by the empirical L_2 error. To formulate the result we use a fixed design regression model.

Let

$$Y_i = m(x_i) + W_i \qquad (i = 1, \dots, n)$$

for some $x_1, \dots, x_n \in \mathbb{R}^d, m : \mathbb{R}^d \to \mathbb{R}$ and some random variables $W_1, \dots, W_n$ which are independent and have expectation zero. Additionally we assume that the W_i's are sub-Gaussian, i.e.

$$\max_{i=1,\dots,n} K^2 \mathbf{E} \left\{ \exp\left(\frac{W_i^2}{K^2} \right) - 1 \right\} \leq \sigma_0^2 \tag{29}$$

for some $K, \sigma_0 > 0$. Instead of $x_1, \dots, x_n$, we observe only $\bar{x}_1, \dots, \bar{x}_n \in \mathbb{R}^d$. Our goal is to estimate m from $(\bar{x}_1, \bar{Y}_1), \dots, (\bar{x}_n, \bar{Y}_n)$, where $\bar{Y}_1, \dots, \bar{Y}_n$ are arbitrary random variables with the property that the average squared measurement error

$$\frac{1}{n} \sum_{i=1}^{n} \left| Y_i - \bar{Y}_i \right|^2$$

is "small".

Let $\mathcal{F}_n$ be a set of functions $f : \mathbb{R}^d \to \mathbb{R}$ and consider the least squares estimate

$$\hat{m}_n = \arg \min_{f \in \mathcal{F}_n} \frac{1}{n} \sum_{i=1}^{n} |f(\bar{x}_i) - \bar{Y}_i|^2. \tag{30}$$

To shorten some of the expressions in the next proof, we use the following notations: For $\mathbb{R}^d$-valued random variables $Z_1, \dots, Z_n$ we write

$$\mathbf{P}_{Z_1^n} = \frac{1}{n} \sum_{i=1}^{n} \delta_{Z_i},$$

where δ_{Z_i} denotes the point mass at Z_i. Then for a function $g : \mathbb{R}^d \to \mathbb{R}$ we have

$$\int g d\mathbf{P}_{Z_1^n} = \frac{1}{n} \sum_{i=1}^{n} g(Z_i).$$

If $z_1, \dots, z_n \in \mathbb{R}^d$, we can define in the same way $\mathbf{P}_{z_1^n}$ and $\int g d\mathbf{P}_{z_1^n}$.

Lemma 2. *Assume that the sub-Gaussian condition (29) holds and let the estimate be defined by (30). Then there exists a constant* $c_5 > 0$ *which depends only on* σ_0 *and* K *such that for any* δ_n *with*

$$\delta_n \to 0 \quad (n \to \infty) \text{ and } n\delta_n \to \infty \quad (n \to \infty)$$

and

$$\sqrt{n} \cdot \delta \geq c_5 \int\limits_{\frac{\delta}{8 \cdot \sqrt{2} \cdot \sigma_0}}^{\sqrt{\delta}} \sqrt{\log \mathcal{N}_2 \left(u, \left\{ f - h : f \in \mathcal{F}_n, \int (f-h)^2 d\mathbf{P}_{x_1^n} \leq \delta \right\}, x_1^n \right)} du \tag{31}$$

for all $\delta \geq \delta_n$, all $x_1, \ldots, x_n \in \mathbb{R}^d$ and all $h \in \mathcal{F}_n$, we have for a sufficiently large constant c_6

$$\mathbf{P}\left\{ \frac{1}{n} \sum_{i=1}^{n} |\hat{m}_n(\bar{x}_i) - m(x_i)|^2 > 3 \frac{1}{n} \sum_{i=1}^{n} |h(\bar{x}_i) - m(x_i)|^2 + 128 \frac{1}{n} \sum_{i=1}^{n} |Y_i - \bar{Y}_i|^2 + \delta_n \right\}$$
$$\leq c_6 \cdot \exp(-c_6 \cdot n \cdot \delta_n) \to 0 \qquad (n \to \infty).$$

Proof. Let $h \in \mathcal{F}_n$ be arbitrary. It follows from

$$2 \frac{1}{n} \sum_{i=1}^{n} |\hat{m}_n(\bar{x}_i) - h(\bar{x}_i)|^2 + 2 \frac{1}{n} \sum_{i=1}^{n} |h(\bar{x}_i) - m(x_i)|^2 \geq \frac{1}{n} \sum_{i=1}^{n} |\hat{m}_n(\bar{x}_i) - m(x_i)|^2,$$

that

$$\frac{1}{n} \sum_{i=1}^{n} |\hat{m}_n(\bar{x}_i) - m(x_i)|^2 > 3 \frac{1}{n} \sum_{i=1}^{n} |h(\bar{x}_i) - m(x_i)|^2 + 128 \frac{1}{n} \sum_{i=1}^{n} |Y_i - \bar{Y}_i|^2 + \delta_n$$

implies

$$\frac{\delta_n}{2} < \frac{1}{n} \sum_{i=1}^{n} |\hat{m}_n(\bar{x}_i) - h(\bar{x}_i)|^2.$$

By combining this with Lemma 1 we get

$$\begin{aligned}
& \mathbf{P}\left\{ \frac{1}{n} \sum_{i=1}^{n} |\hat{m}_n(\bar{x}_i) - m(x_i)|^2 > 3 \frac{1}{n} \sum_{i=1}^{n} |h(\bar{x}_i) - m(x_i)|^2 + 128 \frac{1}{n} \sum_{i=1}^{n} |Y_i - \bar{Y}_i|^2 + \delta_n \right\} \\
\leq \; & \mathbf{P}\left\{ \frac{\delta_n}{2} \leq \int |\hat{m}_n - h|^2 d\mathbf{P}_{\bar{x}_1^n} \leq \frac{16}{n} \sum_{i=1}^{n} (\hat{m}_n(\bar{x}_i) - h(\bar{x}_i)) \cdot W_i \right\} \\
\leq \; & \underbrace{\mathbf{P}\left\{ \frac{1}{n} \sum_{i=1}^{n} W_i^2 > 2\sigma_0^2 \right\}}_{=:P_1} \\
& + \underbrace{\mathbf{P}\left\{ \frac{1}{n} \sum_{i=1}^{n} W_i^2 \leq 2\sigma_0^2, \frac{\delta_n}{2} \leq \int |\hat{m}_n - h|^2 d\mathbf{P}_{\bar{x}_1^n} \leq \frac{16}{n} \sum_{i=1}^{n} (\hat{m}_n(\bar{x}_i) - h(\bar{x}_i)) \cdot W_i \right\}}_{=:P_2}.
\end{aligned}$$

Application of Chernoff's exponential bounding method (cf. Chernoff (1952)) yields

$$\begin{aligned}
P_1 &= \mathbf{P}\left\{\sum_{i=1}^n \frac{W_i^2}{K^2} \geq \frac{2n\sigma_0^2}{K^2}\right\} \\
&\leq \mathbf{P}\left\{\exp\left(\sum_{i=1}^n \frac{W_i^2}{K^2}\right) > \exp\left(\frac{2n\sigma_0^2}{K^2}\right)\right\} \\
&\leq \exp\left(-\frac{2n\sigma_0^2}{K^2}\right) \cdot \mathbf{E}\left\{\exp\left(\sum_{i=1}^n \frac{W_i^2}{K^2}\right)\right\} \\
&\stackrel{(29)}{\leq} \exp\left(-\frac{2n\sigma_0^2}{K^2}\right) \cdot \left(1 + \frac{\sigma_0^2}{K^2}\right)^n \\
&\leq \exp\left(-\frac{2n\sigma_0^2}{K^2}\right) \cdot \exp\left(n \cdot \frac{\sigma_0^2}{K^2}\right) \leq \exp(-c_7 \cdot n \cdot \delta_n) \to 0 \qquad (n \to \infty).
\end{aligned}$$

To bound P_2, we observe first that $\frac{1}{n}\sum_{i=1}^n W_i^2 \leq 2\sigma_0^2$ together with the Cauchy-Schwarz inequality implies

$$\frac{16}{n}\sum_{i=1}^n (\hat{m}_n(\bar{x}_i) - h(\bar{x}_i)) \cdot W_i \leq 16 \cdot \sqrt{\int |\hat{m}_n - h|^2 d\mathbf{P}_{\bar{x}_1^n}} \cdot \sqrt{2\sigma_0^2},$$

hence inside of P_2 we have $\int |\hat{m}_n - h|^2 d\mathbf{P}_{\bar{x}_1^n} \leq 512\sigma_0^2$.

Define

$$S = \min\left\{s \in \mathbb{N}_0 : 2^s \delta_n \geq 512\sigma_0^2\right\}.$$

Application of the peeling device (cf., e.g., Section 5.3 in van de Geer (2000)) yields

$$\begin{aligned}
& P_2 \qquad (32) \\
\leq & \sum_{s=0}^S \mathbf{P}\Bigg\{\frac{1}{n}\sum_{i=1}^n W_i^2 \leq 2\sigma_0^2, 2^{s-1}\delta_n < \int |\hat{m}_n - h|^2 d\mathbf{P}_{\bar{x}_1^n} \leq 2^s\delta_n, \\
& \qquad \int |\hat{m}_n - h|^2 d\mathbf{P}_{\bar{x}_1^n} \leq \frac{16}{n}\sum_{i=1}^n (\hat{m}(\bar{x}_i) - h(\bar{x}_i))W_i\Bigg\} \\
\leq & \sum_{s=0}^S \mathbf{P}\left\{\frac{1}{n}\sum_{i=1}^n W_i^2 \leq 2\sigma_0^2, \int |\hat{m}_n - h|^2 d\mathbf{P}_{\bar{x}_1^n} \leq 2^s\delta_n, \frac{1}{n}\sum_{i=1}^n (\hat{m}(\bar{x}_i) - h(\bar{x}_i))W_i > \frac{2^s\delta_n}{32}\right\}.
\end{aligned}$$

Because of $n\delta_n \to \infty \quad (n \to \infty)$, we may assume that $\sqrt{n\delta_n} > 32c_5$. So the probabilities in above sums can be bounded by Corollary 8.3 in van de Geer (2000) (use there $2C = c_5, R = \sqrt{2^s\delta_n}, \delta = \frac{2^s\delta_n}{32}, \sigma = \sqrt{2}\sigma_0$ and $\mathcal{G} = \left\{f - h : f \in \mathcal{F}_n, \int |f - h|^2 d\mathbf{P}_{x_1^n} \leq 2^s\delta_n\right\}$).

This yields

$$\begin{aligned}
P_2 &\leq \sum_{s=0}^S c_8 \cdot \exp\left(-\frac{n\left(\frac{2^s\delta_n}{32}\right)^2}{4c_8 2^s\delta_n}\right) \leq \sum_{s=0}^S c_8 \cdot \exp\left(-\frac{n2^s}{4 \cdot 32^2 c_8}\delta_n\right) \\
&\leq c_9 \cdot \exp(-c_9 \cdot n \cdot \delta_n) \to 0
\end{aligned}$$

for $n \to \infty$, where c_9 is a constant which does only depend on K and σ_0. $\square$

Finally we need a lemma which enables us to bound the L_2 error by some constant times the empirical L_2 error.

Lemma 3. *Let $\beta \geq 1$, let $m : \mathbb{R}^d \to [-\beta, \beta]$ and let $\mathcal{F}_n$ be a class of functions $f : \mathbb{R}^d \to [-\beta, \beta]$. Let $0 < \epsilon < 1$ and $\alpha > 0$. Assume that*

$$\sqrt{n}\epsilon\sqrt{\alpha} \geq 1152\beta$$

and that, for all $x_1, \ldots, x_n \in \mathbb{R}$ and all $\delta > 2\beta^2\alpha$, we have

$$\frac{\sqrt{n}\epsilon\delta}{768\sqrt{2}\beta^2} \geq \int_{\frac{\epsilon\delta}{128\beta^2}}^{\sqrt{\delta}} \sqrt{\log \mathcal{N}_2\left(\frac{u}{4\beta}, \left\{f - m : f \in \mathcal{F}_n, \int |f-m|^2 d\mathbf{P}_{x_1^n} \leq \frac{\delta}{\beta^2}\right\}, x_1^n\right)} du. \quad (33)$$

Then

$$\mathbf{P}\left\{\sup_{f \in \mathcal{F}_n} \frac{\left|\int |f-m|^2 d\mu - \int |f-m|^2 d\mathbf{P}_{X_1^n}\right|}{\alpha + \int |f-m|^2 d\mu + \int |f-m|^2 d\mathbf{P}_{X_1^n}} > \epsilon\right\} \leq 15 \exp\left(-\frac{n\alpha\epsilon^2}{512 \cdot 2304\beta^2}\right).$$

Proof. See Lemma 5 in Kohler (2006) and the literatur cited there. $\square$

Remark 5. By Lemma 3 we can bound the L_2 error by some constant times the empirical L_2 error:

$$\mathbf{E}\left\{|f(X) - m(X)|^2\right\} > \alpha + 2\int |f-m| \, d\mathbf{P}_{X_1^n}$$

is equivalent to

$$\frac{\mathbf{E}\left\{|f(X) - m(X)|^2\right\} - \int |f-m|^2 \, d\mathbf{P}_{X_1^n}}{2\alpha + \mathbf{E}\left\{|f(X) - m(X)|^2\right\} + \int |f-m|^2 \, d\mathbf{P}_{X_1^n}} > \frac{1}{3}.$$

Hence

$$\begin{aligned} &\mathbf{P}\left\{\exists f \in \mathcal{F}_n : \mathbf{E}\left\{|f(X) - m(X)|^2\right\} > \alpha + 2 \cdot \int |f-m|^2 d\mathbf{P}_{X_1^n}\right\} \\ \leq \;& \mathbf{P}\left\{\sup_{f \in \mathcal{F}_n} \frac{\left|\mathbf{E}\left\{|f(X) - m(X)|^2\right\} - \int |f-m|^2 d\mathbf{P}_{X_1^n}\right|}{2\alpha + \mathbf{E}\left\{|f(X) - m(X)|^2\right\} + \int |f-m|^2 d\mathbf{P}_{X_1^n}} > \frac{1}{3}\right\}. \end{aligned}$$

Similary one can show

$$\begin{aligned}
&\mathbf{P}\left\{\exists f \in \mathcal{F}_n : \int |f-m|^2 d\mathbf{P}_{X_1^n} > \alpha + 2\cdot \mathbf{E}\left\{|f(X)-m(X)|^2\right\}\right\} \\
\leq\ &\mathbf{P}\left\{\sup_{f\in\mathcal{F}_n} \frac{\left|\int |f-m|^2 d\mathbf{P}_{X_1^n} - \mathbf{E}\left\{|f(X)-m(X)|^2\right\}\right|}{2\alpha + \int |f-m|^2 d\mathbf{P}_{X_1^n} + \mathbf{E}\left\{|f(X)-m(X)|^2\right\}} > \frac{1}{3}\right\}.
\end{aligned}$$

4.2 Proof of Theorem 1

Set

$$h_n(\cdot) = \arg\min_{f\in\mathcal{F}_n} \int |f(x)-m(x)|^2 \, \mu(dx).$$

In the proof we use the error decomposition

$$\begin{aligned}
&\int |\bar{m}_n - m|^2 d\mu \\
= &\int |\bar{m}_n - m|^2 d\mu - 2\int |\bar{m}_n - m|^2 d\mathbf{P}_{X_1^n} \\
+ \ & 2\int |\bar{m}_n - m|^2 d\mathbf{P}_{X_1^n} - \frac{4}{n}\sum_{i=1}^n |\bar{m}_n(X_i) - \bar{m}_n(\bar{X}_i)|^2 - \frac{4}{n}\sum_{i=1}^n |\bar{m}_n(\bar{X}_i) - m(X_i)|^2 \\
+ \ & \frac{4}{n}\sum_{i=1}^n |\bar{m}_n(\bar{X}_i) - m(X_i)|^2 - \frac{24}{n}\sum_{i=1}^n |h_n(\bar{X}_i) - h_n(X_i)|^2 \\
& -24\int |h_n - m|^2 d\mathbf{P}_{X_1^n} - \frac{512}{n}\sum_{i=1}^n |Y_i - \bar{Y}_i|^2 \\
+ \ & \frac{24}{n}\sum_{i=1}^n |h_n(X_i) - h_n(\bar{X}_i)|^2 + \frac{4}{n}\sum_{i=1}^n |\bar{m}_n(X_i) - \bar{m}_n(\bar{X}_i)|^2 + \frac{512}{n}\sum_{i=1}^n |\bar{Y}_i - Y_i|^2 \\
+ \ & 24\int |h_n - m|^2 d\mathbf{P}_{X_1^n} - 48\int |h_n - m|^2 d\mu + 48\int |h_n - m|^2 d\mu
\end{aligned}$$

By Remark 5 we can conclude

$$\begin{aligned}
&\mathbf{P}\left\{\int |\bar{m}_n - m|^2 d\mu - 2\int |\bar{m}_n - m|^2 d\mathbf{P}_{X_1^n} > \delta_n\right\} \\
\leq\ &\mathbf{P}\left\{\sup_{f\in\mathcal{F}_n} \frac{\left|\mathbf{E}\left\{|f(X)-m(X)|^2\right\} - \int |f-m|^2 d\mathbf{P}_{X_1^n}\right|}{2\delta_n + \mathbf{E}\left\{|f(X)-m(X)|^2\right\} + \int |f-m|^2 d\mathbf{P}_{X_1^n}} > \frac{1}{3}\right\}.
\end{aligned}$$

Condition (33) of Lemma 3 is implied by

$$\sqrt{n}\frac{\delta_n}{\beta_n^2} \ \geq \ c_1 \int_{c_2\delta/\beta_n^2}^{\frac{\sqrt{\delta}}{4}} \sqrt{\log \mathcal{N}_2\left(\frac{u}{4\beta_n^2}, \left\{f - m : f\in\mathcal{F}_n, \int |f-m|^2 d\mathbf{P}_{x_1^n} \leq \delta\right\}, x_1^n\right)}$$

which in turn is implied by condition (10) of the theorem. Because of (9) we may assume that

$$\sqrt{n\delta_n} > \frac{3 \cdot 1152 \cdot \beta_n}{\sqrt{2}}.$$

Now Lemma 3 implies

$$\mathbf{P}\left\{\int |\bar{m}_n - m|^2 d\mu - 2\int |\bar{m}_n - m|^2 d\mathbf{P}_{X_1^n} > \delta_n\right\} \leq 15 \exp\left(-\frac{n \cdot 2 \cdot \delta_n}{512 \cdot 2304 \cdot 9 \cdot \beta_n^2}\right).$$

So by using again (9) we get

$$\int |\bar{m}_n - m|^2 d\mu - 2\int |\bar{m}_n - m|^2 d\mathbf{P}_{X_1^n} = O_{\mathbf{P}}\left(\delta_n\right).$$

In the same way one can show that

$$24\int |h_n - m|^2 d\mathbf{P}_{X_1^n} - 48\int |h_n - m|^2 d\mu = O_{\mathbf{P}}\left(\delta_n\right). \tag{34}$$

Because of $(a+b)^2 \leq 2a^2 + 2b^2$ we have

$$2\int |\bar{m}_n - m|^2 d\mathbf{P}_{X_1^n} - \frac{4}{n}\sum_{i=1}^n |\bar{m}_n(X_i) - \bar{m}_n(\bar{X}_i)|^2 - \frac{4}{n}\sum_{i=1}^n |\bar{m}_n(\bar{X}_i) - m(X_i)|^2 \leq 0,$$

and therefore

$$2\int |\bar{m}_n - m|^2 d\mathbf{P}_{X_1^n} - \frac{4}{n}\sum_{i=1}^n |\bar{m}_n(X_i) - \bar{m}_n(\bar{X}_i)|^2 - \frac{4}{n}\sum_{i=1}^n |\bar{m}_n(\bar{X}_i) - m(X_i)|^2 = O_{\mathbf{P}}(0).$$

Application of Lemma 2 (conditioned on $(X_i, \bar{X}_i)$ $(i = 1, \ldots, n)$) implies

$$\begin{aligned}
& \mathbf{P}\left\{\frac{4}{n}\sum_{i=1}^n |\bar{m}_n(X_i) - m(X_i)|^2 - \frac{24}{n}\sum_{i=1}^n |h_n(\bar{X}_i) - h_n(X_i)|^2\right. \\
& \qquad \left. - \frac{24}{n}\sum_{i=1}^n |h_n(X_i) - m(X_i)|^2 - \frac{512}{n}\sum_{i=1}^n |Y_i - \bar{Y}_i|^2 > \delta_n\right\} \\
\leq \quad & \mathbf{P}\left\{\frac{1}{n}\sum_{i=1}^n |\bar{m}_n(X_i) - m(X_i)|^2 - \frac{3}{n}\sum_{i=1}^n |h_n(\bar{X}_i) - m(X_i)|^2 - \frac{128}{n}\sum_{i=1}^n |Y_i - \bar{Y}_i|^2 > \frac{\delta_n}{4}\right\} \\
\to \quad & 0 \qquad (n \to \infty).
\end{aligned}$$

This means

$$\begin{aligned}
& \frac{4}{n}\sum_{i=1}^n |\bar{m}_n(X_i) - m(X_i)|^2 - \frac{24}{n}\sum_{i=1}^n |h_n(\bar{X}_i) - h_n(X_i)|^2 \\
& \qquad - \frac{24}{n}\sum_{i=1}^n |h_n(X_i) - m(X_i)|^2 - \frac{512}{n}\sum_{i=1}^n |Y_i - \bar{Y}_i|^2 = O_{\mathbf{P}}(\delta_n).
\end{aligned} \tag{35}$$

The functions $\bar{m}_n$ and h_n are Lipschitz-continuous, which means

$$\begin{aligned}
& \frac{24}{n}\sum_{i=1}^{n}|h_n(X_i)-h_n(\bar{X}_i)|^2+\frac{4}{n}\sum_{i=1}^{n}|\bar{m}_n(X_i)-\bar{m}_n(\bar{X}_i)|^2 \\
\leq \; & \frac{24}{n}L_n^2\sum_{i=1}^{n}|X_i-\bar{X}_i|^2+\frac{4}{n}L_n^2\sum_{i=1}^{n}|X_i-\bar{X}_i|^2 \\
= \; & O_{\mathbf{P}}\left(L_n^2\frac{1}{n}\sum_{i=1}^{n}|\bar{X}_i-X_i|^2\right).
\end{aligned}$$

Obviously it holds

$$\frac{512}{n}\sum_{i=1}^{n}|\bar{Y}_i-Y_i|^2=O_{\mathbf{P}}\left(\frac{1}{n}\sum_{I=1}^{n}\left|Y_i-\bar{Y}_i\right|^2\right).$$

Finally we have

$$\int |h_n-m|^2d\mu=O_{\mathbf{P}}\left(\inf_{f\in\mathcal{F}_n}\int |f(x)-m(x)|^2\mu(dx)\right).$$

The additivity of the $O_{\mathbf{P}}$–symbol completes the proof of the first assertion. The second assertion follows from (34), (35) and the Lipschitz continuity of h_n. □

4.3 Preliminarien to the proof of Theorem 2

In this subsection we prove two auxiliary results, which we need in the proof of Theorem 2. We start with

Lemma 4. *Let $f_t, q_t, \bar{q}_t : \mathbb{R}^d \to \mathbb{R}$ and assume that f_t are Lipschitz continuous with Lipschitz constant L $(t=1,\ldots,T)$. For given $\mathbb{R}^d$-valued stochastic processes $(X_t)_{t=0,\ldots,T}$ and, $(\bar{X}_t)_{t=0,\ldots,T}$ we define*

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

Then we have for all $s\in\{0,\ldots,T-1\}$

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

Proof. $(a+b)^2\leq 2a^2+2b^2$ and $|\max\{a,b\}-\max\{a,c\}|\leq|b-c|$ imply

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

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

In our second auxiliary result we bound the measurement error occuring in the x-variables in estimation of a Black-Scholes model.

Lemma 5. *Let $X_{t,i}$ and $\bar{X}_{t,i}$ be defined as in (14) and (15). Assume that $\hat{\sigma}_n$ satisfies (21). Then we have for any $t \in \{0, \ldots, T\}$*

$$\gamma_n^2 \cdot \frac{1}{n} \sum_{i=1}^{n} \left| \bar{X}_{t,i} - X_{t,i} \right|^2 \to 0 \quad a.s.$$

Proof. It holds

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^{n} \left| \bar{X}_{t,i} - X_{t,i} \right|^2 \\ = \quad & x_0^2 \exp(2rt) \frac{1}{n} \sum_{i=1}^{n} \left| \exp\left(\hat{\sigma}_n W_{t,i} - \frac{\hat{\sigma}_n^2}{2} t \right) - \exp\left(\sigma W_{t,i} - \frac{\sigma^2}{2} t \right) \right|^2 . \end{aligned}$$

By the mean value theorem it exists $\xi_i \in [\min\{a_i, b_i\}, \max\{a_i, b_i\}]$, where

$$a_i = \sigma W_{t,i} - \frac{\sigma^2}{2} t \text{ and } b_i = \hat{\sigma}_n W_{t,i} - \frac{\hat{\sigma}_n^2}{2} t$$

such that

$$\begin{aligned} & \left| \exp\left(\sigma W_{t,i} - \frac{\sigma^2}{2} t \right) - \exp\left(\hat{\sigma}_n W_{t,i} - \frac{\hat{\sigma}_n^2}{2} t \right) \right|^2 \\ \leq \quad & \left| \left(\sigma - \hat{\sigma}_n \right) W_{t,i} + \frac{\sigma_n^2 - \sigma^2}{2} t \right|^2 \exp\left(2\xi_i \right) \\ \leq \quad & \left| \left(\sigma - \hat{\sigma}_n \right) W_{t,i} + \frac{\sigma_n^2 - \sigma^2}{2} t \right|^2 \exp\left((|\sigma| + |\hat{\sigma}|) \cdot |W_{t,i}| + (\frac{\sigma^2}{2} + \frac{\hat{\sigma}_n^2}{2}) \cdot t \right) \\ = \quad & |\sigma - \hat{\sigma}_n|^2 \cdot \left| W_{t,i} + \frac{\hat{\sigma}_n + \sigma}{2} \cdot t \right|^2 \exp\left((|\sigma| + |\hat{\sigma}|) \cdot |W_{t,i}| + (\frac{\sigma^2}{2} + \frac{\hat{\sigma}_n^2}{2}) \cdot t \right) . \end{aligned}$$

Since W_t is $N(0, t)$-distributed we have for any $c_{10} > 0$

$$\mathbf{E} \left\{ \left| |W_{t,1}| + c_{10} \right|^2 \exp\left(c_{10} \cdot |W_{t,1}| + c_{10} \right) \right\} < \infty.$$

Together with the strong law of large numbers and (21) we get the assertion. $\square$

4.4 Proof of Theorem 2

We proof 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

$$\int |\hat{q}_{n,s}(x) - q_s(x)|^2 \mathbf{P}_{X_s}(dx) \to 0 \quad a.s. \tag{36}$$

and

$$\frac{1}{n}\sum_{i=1}^{n}|\hat{q}_{n,s}(\bar{X}_{s,i})-q_s(X_{s,i})|^2 \to 0 \quad a.s. \tag{37}$$

for $s = T$.

Let $t \in \{0,\dots,T-1\}$ be arbitrary and assume that (36) and (37) hold for $s = t+1$. In the sequel we show (36) and (37) for $s = t$. To do this we apply Theorem 1 in the form described in Remark 2 and (by using the bounds on the covering number from the proof of Corollaries 1 and 2) we get

$$\sum_{n=1}^{\infty}\mathbf{P}\left\{\int |\hat{q}_{n,t}(x)-q_t(x)|^2\mathbf{P}_{X_t}(dx) > c_{11}\cdot Z_n\right\} < \infty$$

and

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

where

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

By an application of the Borel-Cantelli lemma we get that it suffices to show

$$Z_n \to 0 \quad a.s.$$

in order to prove (36) and (37) for $s = t$. But this in turn follows from the assumptions of Theorem 2, Lemma 4 and Lemma 5. To see

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

we approximate q_t by a smooth function with compact support (cf., e.g., Theorem A.1 in Györfi et al. (2002)) and observe that we can approximate this smooth function arbitrarily exact by spline functions in $\mathcal{F}_n$. This completes the proof of (22).

By an easy application of Cauchy-Schwarz inequality together with

$$|\max\{a,b\}-\max\{a,c\}| \le |a-c| \quad (a,b,c\in\mathbb{R})$$

we get

$$\hat{V}_{0,n} = \int \max\{f_0(x_0),\hat{q}_{n,0}(x_0)\}d\mathbf{P}_{X_0}(x_0) \to \int \max\{f_0(x_0),q_0(x_0)\}d\mathbf{P}_{X_0}(x_0) = V_0 \quad a.s.$$

The proof is complete. $\square$

References

[1] Belomestny, D. (2009). *Pricing Bermudan options using regression: optimal rates of convergence for lower estimates.* Preprint.

[2] Caroll, R. J., Maca, J. D. and Ruppert, D. (1999). *Nonparametric regression in the presence of measurement error.* Biometrika **86**, pp. 541-554.

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

[4] Chernoff, H. (1952). *A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations.* Ann. Math. Stat. **23**, pp. 493-507.

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

[6] Delaigle, A. and Meister, A. (2007). *Nonparametric regression estimation in the heteroscedastic errors-in-variables problem.* Journal of the American Statistical Association **102**, pp. 1416-1426.

[7] Delaigle, A., Fan, J. and Carroll, R.J. (2009). *A design-adaptive local polynomial estimator for the errors-in-variables problem.* Journal of the American Statistical Association **104**, pp. 348-359.

[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] Fan,. J., Truong, Y.K. (1993). *Nonparametric regression with errors in variables.* Annals of Statistics **21**, pp. 1900-1925.

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

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

[13] Kohler, M. (2006). *Nonparametric regression with additional measurement errors in the dependent variable.* Journal of Statistical Planning and Inference **136**, pp. 3339-3361.

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

[16] Kohler, M., Krzyżak, A., Todorovich, N. (2006). *Pricing of high-dimensional american options by neural networks.* To appear in *Mathematical Finance.*

[17] Kohler, M. and Mehnert, J. (2009). *Analysis of the rate of convergence of least squares neural network regression estimates in case of measurement errors.* Preprint.

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

[19] Pollard (1984). *Convergence of Stochastic Processes.* Springer Series in Statistics, Springer.

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

[21] Stone, C. J. (1982). *Optimal global Rates of Convergence for Nonparametric Regression.* Annals of Statistics **10**, pp. 1040-1053.

[22] 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.

[23] van de Geer, S. (2000). *Empirical Processes in M–estimation.* Cambridge University Press.