

Birke, Melanie; Pilz, Kay F.

**Working Paper**

## Nonparametric option pricing with no-arbitrage constraints

Technical Report, No. 2007,30

**Provided in Cooperation with:**

Collaborative Research Center 'Reduction of Complexity in Multivariate Data Structures' (SFB 475), University of Dortmund

*Suggested Citation:* Birke, Melanie; Pilz, Kay F. (2007) : Nonparametric option pricing with no-arbitrage constraints, Technical Report, No. 2007,30, Universität Dortmund, Sonderforschungsbereich 475 - Komplexitätsreduktion in Multivariaten Datenstrukturen, Dortmund

This Version is available at:

<https://hdl.handle.net/10419/25013>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# Nonparametric Option Pricing with No-arbitrage Constraints

Melanie Birke  
Ruhr-Universität Bochum  
Fakultät für Mathematik  
44780 Bochum, Germany  
e-mail: melanie.birke@rub.de

Kay F. Pilz  
Sal. Oppenheim jr. & Cie. KGaA  
Trading & Derivatives  
60329 Frankfurt, Germany  
e-mail: KayFrederik.Pilz@oppenheim.de

September 6, 2007

## Abstract

We propose a completely kernel based method of estimating the call price function or the state price density of options. The new estimator of the call price function fulfills the constraints like monotonicity and convexity given in Breeden and Litzenberger (1978) without necessarily estimating the state price density for an underlying asset price from its option prices. It can be shown that the estimator is pointwise consistent and asymptotically normal. In a simulation study we compare the new estimator to the unconstrained kernel estimator and to the estimator given in Aït-Sahalia and Duarte (2003).

**Keywords:** Call pricing function, constrained nonparametric estimation, monotone rearrangements, state price density

## 1 Introduction

Research about estimating probability density functions (PDF) implied in option prices has been an ongoing topic since the pioneering work of Breeden and Litzenberger (1978). They observed that by expressing the call price  $C$  as the discounted expectation under the equivalent martingale measure  $Q$  of the payoff function at expiration time  $T$  of the option, i.e.

$$\begin{aligned} C(t, T, S_t, K, r_{t,T}, d_{t,T}) &= e^{-r_{t,T}(T-t)} \mathbb{E}_Q[\max(S_T - K, 0)] \\ &= e^{-r_{t,T}(T-t)} \int_K^{+\infty} (S_T - K) p_Q(S_T) dS_T, \end{aligned} \quad (1)$$

and differentiating this equation two times with respect to  $K$ , the following relationship between call price and PDF holds,

$$\frac{\partial^2 C(t, T, S_t, K, r_{t,T}, d_{t,T})}{\partial K^2} = e^{-r_{t,T}(T-t)} p_Q(K). \quad (2)$$

Here,  $K$  denotes the strike of the option,  $t$  the valuation time and  $p_Q$  the PDF related to the equivalent martingale measure, which is also named state price density (SPD). The risk free rate  $r_{t,T}$  and the dividend yield  $d_{t,T}$  are assumed to be deterministic continuous rates, dependent on the valuation time  $t$  and the maturity  $T$ .

Many approaches perform a parametric fit to the SPD or to the implied volatility function from which the call prices can be computed by (1) or by the Black-Scholes equation, respectively. In the latter case, which was first introduced by Shimko (1993), the Black-Scholes equation is used only for transformation, i.e. the returns of the underlying asset price process are not assumed to be log-normally distributed. For a comprehensive review of parametric methods we refer to Jackwerth (2004). These approaches have in common that they assume a functional form at some stage of the estimation procedure. Since there are no generally accepted parametric forms for asset price dynamics, for volatility surfaces or for call and put price functions, the use of parametric models for estimating the implied PDF bears the risk of introducing systematic errors. The nonparametric approach of estimating SPDs tries to overcome this risk of misspecification. Aït-Sahalia and Lo (1998) use a nonparametric kernel estimator of Nadaraya-Watson type for the SPD with the 5-dimensional design space consisting of spot, strike, time to maturity, interest rate and dividend yield. Since kernel estimators become very data intensive when the number of dependent variables is large (so called "curse of dimensionality", cf. Fan and Gijbels (1996)) this approach requires sample sizes which can only be achieved by using historical data, which in turn requires the questionable assumption of i.i.d. sample points for the considered period. Several dimension reduction techniques are also discussed by Aït-Sahalia and Lo (1998) to overcome this problem, e.g. the semiparametric estimator based on the Black-Scholes formula except that the implied volatility function is a nonparametric function of forward price, strike and time to maturity.

Aït-Sahalia and Duarte (2003) consider the call price (1) as a 1-dimensional function of the strike alone, by using only options with equal time to maturity and assuming that interest rate and dividend yield are deterministic functions of  $t$  and  $T$ . This reduction to only one dimension has two major advantages: First, the method becomes even applicable to marked-to-market option valuation, because no historical data is needed in the estimation procedure. Second, no-arbitrage conditions (in the direction of the strike) can be incorporated in the estimation procedure more easily (see section 2.2). On the other hand, there are basically two consequences of the imposed restrictions in relation to the full nonparametric approach given in Aït-Sahalia and Lo (1998). First, since every SPD related to a certain option maturity has to be estimated separately, one does not get a density surface, but only slices of particular SPDs or call price functions respectively. This is not a major disadvantage because even in the case of time to maturity as a dependent variable the values for times between two market maturities are interpolated by the kernel method and it is not assured that the corresponding call price function is arbitrage-free. Second, interest rates and dividend yields are assumed to be deterministic functions of valuation time  $t$  and time to maturity  $T$ , which is apparently not a realistic assumption. This implies that an estimated SPD is only valid as long as the functions  $r_{t,T}$  and  $d_{t,T}$  have not changed.

Aït-Sahalia and Duarte (2003) estimate the call pricing function by applying a constrained least

squares procedure in a first step and smoothing the result in a second step using local polynomials. Recently, Yatchew and Härdle (2006) introduced a further constrained nonparametric least squares estimator. They also assume the call price to be a function only depending on the strike price. Different from the method described in Aït-Sahalia and Duarte (2003) they use a nonparametric least squares approach which directly yields a two times differentiable estimator. There exist many other nonparametric approaches different from kernel methods and for a more complete overview we refer to Fengler (2005) and Jackwerth (2004).

In this paper we propose a further nonparametric estimator of the 1-dimensional call pricing function (1) with strike as dependend variable and the other arguments fixed. Kernel methods – regression as well as density estimation – are the building blocks of a method for convex regression estimation proposed by Birke and Dette (2007). This method relies on a monotonizing procedure described in Dette, Neumeyer and Pilz (2006) which is applied to the derivative of an unconstrained estimator. We slightly change this method to satisfy further restrictions imposed by the no-arbitrage condition. Our approach allows to estimate the call price function without calculating the corresponding SPD, i.e. the second derivative of the call price function. When only an interpolation between observed call prices is needed this technique might be advantageous, because estimation becomes more cumbersome when higher order derivatives have to be estimated. Furthermore it is also possible to estimate the SPD from our approach (see Remark 3.1).

The rest of this paper is organized as follows: In section 2 we review a few details about nonparametric kernel estimators which are required for our approach. We also discuss there how the no-arbitrage conditions can be formulated in an adequate way for our estimator. In section 3, we define our estimator of the call pricing function and discuss its asymptotic properties. As a main result Theorem 3.3 states asymptotic normality of our estimator. A finite sample simulation is given in section 4. Further, we compare our estimator to the estimator proposed by Aït-Sahalia and Duarte (2003), which is comparable to our estimator in the sense that they also use kernel methods for construction. Finally, section 5 gives a short conclusion of the results. The proofs are deferred to the appendix.

## 2 Preliminaries: Kernel estimators and no-arbitrage conditions

### 2.1 Local polynomial estimation

In this subsection we give an introduction to the concepts of kernel estimators since they are building blocks for our approach given in section 3. For a detailed and comprehensive treatment of nonparametric kernel methods we refer to Fan and Gijbels (1996).

Consider a nonparametric regression model

$$Y = C(X) + \sigma(X)\varepsilon$$

and a bivariate sample  $\{X_i, Y_i\}_{1 \leq i \leq n}$  of independent observations from that model. It is assumed that  $X$  has a positive twice continuously differentiable density  $f$  with compact support  $D$ .

The random variables  $\varepsilon_i$  are i.i.d. with  $E[\varepsilon_i] = 0$  and  $\text{Var}(\varepsilon_i^2) = 1$ ,  $i = 1, \dots, n$ , and finite fourth moment. The variance function  $\sigma : D \rightarrow \mathbb{R}^+$  and the regression function  $C : D \rightarrow \mathbb{R}$  are assumed to be continuous and twice continuously differentiable, respectively. Therefore, the equation  $E[Y | X = x] = C(x)$  holds and approximating this value locally for  $z$  in a neighbourhood of  $x$  gives

$$C(z) \approx \sum_{k=0}^p \frac{C^{(k)}(x)}{k!} \cdot (z - x)^k.$$

The unknown values  $\alpha_k(x) \equiv C^{(k)}(x)/k!$  of the regression function can be estimated from the sampled data by a weighted least squares method, that is by minimizing

$$\sum_{i=1}^n \left\{ Y_i - \sum_{k=0}^p \alpha_k(x) \cdot (X_i - x)^k \right\}^2 \cdot K_h(X_i - x), \quad (3)$$

where the kernel function  $K_h(X - x) \equiv K((X - x)/h)/h$  integrates to one and the expansion of the neighbourhood around  $x$  included in the estimate is controlled by the bandwidth  $h$ . If  $\hat{\alpha}_k \equiv \hat{\alpha}_k(x)$  denotes the estimated coefficients, estimated values for the regression function and their derivatives are given by

$$\hat{C}^{(k)}(x) \equiv \hat{C}_{k,p}(x) \equiv k! \hat{\alpha}_k(x). \quad (4)$$

Let  $A$  denote the design matrix of (3),  $y$  the vector of observed function values and  $\hat{\alpha}_p(x)$  the vector of estimated coefficients,

$$A = \begin{pmatrix} 1 & (X_1 - x) & \cdots & (X_1 - x)^p \\ \vdots & \vdots & & \vdots \\ 1 & (X_n - x) & \cdots & (X_n - x)^p \end{pmatrix}, \quad y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \quad \hat{\alpha} = \begin{pmatrix} \hat{\alpha}_0 \\ \vdots \\ \hat{\alpha}_p \end{pmatrix}.$$

Further, let  $W$  be the diagonal matrix of weights determined by the bandwidth and kernel function,

$$W = \text{diag}(K_h(X_1 - x), \dots, K_h(X_n - x)).$$

Then the general solution of (3) is given by weighted least squares theory,

$$\hat{\alpha} = (A^T W A)^{-1} A^T W y,$$

where the superscript  $T$  denotes transposition of a matrix. In particular, for  $p = 0$  the function fitted locally around  $x$  in (3) is constant and the resulting estimator

$$\hat{C}_{0,0}(x) = \frac{\sum_{i=1}^n K_h(X_i - x) Y_i}{\sum_{i=1}^n K_h(X_i - x)}$$

is called Nadaraya-Watson estimator.

**Remark 2.1** Note that by equation (4) for a local polynomial estimate with order  $p$  all derivatives of the regression function up to order  $p$  are part of the estimate. But they are usually different from the derivatives of the estimate  $\hat{C}_{0,p}$ , i.e. in general

$$\frac{\partial^k \hat{C}_{0,p}}{\partial x^k}(x) \neq \hat{C}_{k,p}(x).$$

## 2.2 Monotonicity and convexity as no-arbitrage constraints

The representation (1) of the call price function by the expectation under the equivalent martingale measure is established by the fundamental work of Cox and Ross (1976) and Harrison and Kreps (1979). Differentiation in (1) with respect to the strike  $K$  yields

$$\frac{\partial C(t, T, S_t, K, r_{t,T}, d_{t,T})}{\partial K} = -e^{-r_{t,T}(T-t)} \int_K^{+\infty} p_Q(S_T) dS_T.$$

By setting  $K$  to 0 and  $+\infty$ , respectively, the boundaries

$$-e^{-r_{t,T}(T-t)} \leq \frac{\partial C(t, T, S_t, K, r_{t,T}, d_{t,T})}{\partial K} \leq 0$$

follow. Taking the partial derivative with respect to  $K$  again results in equation (2) from which the condition

$$\frac{\partial^2 C(t, T, S_t, K, r_{t,T}, d_{t,T})}{\partial K^2} \geq 0$$

is derived. Since we deal only with derivatives with respect to  $K$  we write these two conditions for short by

$$-e^{-r_{t,T}(T-t)} \leq C'(K) \leq 0, \tag{5}$$

$$C''(K) \geq 0. \tag{6}$$

For the option pricing function no-arbitrage bounds

$$\max(0, S_t e^{-d_{t,T}(T-t)} - K e^{-r_{t,T}(T-t)}) \leq C(K) \leq S_t e^{-d_{t,T}(T-t)} \tag{7}$$

have to be satisfied. When the constraints (5) and (6) are already satisfied, (7) reduces to

$$C(0) = S_t e^{-d_{t,T}(T-t)} \tag{8}$$

$$C(K) \geq 0 \quad \forall K \in [0, +\infty). \tag{9}$$

Indeed, (5) and (6) imply that  $C$  is a convex and monotone decreasing function of  $K$ , hence together with (8) and (9) the inequalities  $0 \leq C(K) \leq S_t e^{-d_{t,T}(T-t)}$  hold. To see, that also

$$S_t e^{-d_{t,T}(T-t)} - K e^{-r_{t,T}(T-t)} \leq C(K)$$

is true, note that (8) can be written as

$$C(0) - e^{-r_{t,T}(T-t)}(K - 0) + Ke^{-r_{t,T}(T-t)} = S_te^{-d_{t,T}(T-t)}.$$

Because of (5) we get for any  $\xi > 0$

$$C(0) + C'(\xi)(K - 0) + Ke^{-r_{t,T}(T-t)} \geq S_te^{-d_{t,T}(T-t)},$$

this also holds for that  $\xi$  with

$$C(K) = C(0) + C'(\xi)(K - 0)$$

and therefore yields

$$C(K) + Ke^{-r_{t,T}(T-t)} \geq S_te^{-d_{t,T}(T-t)}.$$

**Remark 2.2** Aït-Sahalia and Duarte (2003) derive a different representation of (7). By using relationship (2) they show that these constraints can be expressed by

$$\int_0^{+\infty} C''(K) dK = e^{-r_{t,T}(T-t)} \quad (10)$$

$$\int_0^{+\infty} KC''(K) dK = e^{-r_{t,T}(T-t)} F(t, T), \quad (11)$$

where  $F(t, T)$  is the forward with maturity  $T$ . Since we want to keep the opportunity to estimate the option price function without using the second derivative  $C''$  we have chosen the representations (8) and (9).

The construction of our estimator is accomplished in the next section by the following steps. First an estimator  $\hat{C}'(K)$  will be defined that satisfies conditions (5) and (6). From this estimator either the SPD can be estimated by further differentiation or the call price function can be derived by integration. In both cases the estimator will be modified with respect to the remaining condition (7), either by using (10) and (11), or by using (8) and (9) in a way specified in the following section.

### 3 Constrained Estimation of Price Functions and State Price Densities

In this section we introduce a completely kernel based estimator of the call price function which fulfills the restrictions stated in section 2.2. The method is a modification of the convex kernel estimator of a regression function  $C$  which has recently been proposed by Birke and Dette (2007). The authors use the fact that a differentiable function  $C$  is convex if and only if its first derivative is increasing. Therefore they apply a monotonizing procedure given in Dette, Neumeyer and Pilz (2006) to the first derivative of an unconstrained kernel estimator of the regression function. The primitive of this isotonization is the convex estimator of the regression function.

### 3.1 Monoton rearrangements of estimators

First we shortly introduce the monotonizing procedure mentioned above. If  $U$  is a uniformly distributed random variable on the interval  $[0, 1]$  and  $g$  is a strictly increasing differentiable function, then  $g(U)$  has the density

$$(g^{-1})'(u)I_{[g(0), g(1)]}(u) \quad (12)$$

and its distribution function is given by

$$\int_{-\infty}^t (g^{-1})'(u)I_{[g(0), g(1)]}(u)du = \int_0^1 I\{g(z) \leq t\}dz = g^{-1}(t)$$

for  $t \in [g(0), g(1)]$ .

For a positive kernel  $K_d$  of order 2 and a bandwidth  $h_d$

$$\frac{1}{h_d} \int_0^1 K_d\left(\frac{g(v) - u}{h_d}\right)dv$$

is a smoothed version of the density in (12) and the corresponding distribution function is given by

$$\frac{1}{h_d} \int_0^1 \int_{-\infty}^t K_d\left(\frac{g(v) - u}{h_d}\right)dudv \quad (13)$$

which is, by the argumentation above, also an approximation of the inverse  $g^{-1}$  of the function  $g$ . If the function  $g$  is not increasing, the functions

$$\phi(g)(t) = \int_0^1 I\{g(z) \leq t\}dz$$

and

$$\phi_{h_d}(g)(t) = \frac{1}{h_d} \int_0^1 \int_{-\infty}^t K_d\left(\frac{g(v) - u}{h_d}\right)dudv$$

are still increasing because they are distribution functions and their generalized inverses are increasing approximations of the function  $g$ . This method of obtaining an increasing function from an arbitrary one is also called increasing rearrangement (see e.g. Bennett and Sharpley, 1988).

In the context of convex function estimation we need in a first step an increasing rearrangement of the derivative of a regression estimate. If  $\hat{C}$  is a consistent and differentiable estimator of the regression function  $C$  and  $\hat{C}'(x) = (\partial/\partial x)\hat{C}(x)$  denotes its first derivative, then

$$\phi_{h_d}(\hat{C}')(t) = \frac{1}{h_d} \int_0^1 \int_{-\infty}^t K_d\left(\frac{\hat{C}''(v) - u}{h_d}\right)dudv \quad (14)$$

is an increasing estimator of  $C'^{-1}$ . The increasing estimator  $\phi_{h_d}(\hat{C}')^{-1}(t)$  of the derivative  $C'$  is then obtained by inversion. Because  $\hat{C}$  is a consistent estimator of the regression function



which is convex, the unconstrained estimator  $\hat{C}'$  of the derivative should be nearly increasing if the sample size  $n$  is large enough and should therefore only differ slightly from the increasing estimator  $\phi_{h_d}(\hat{C}')^{-1}(z)$ . A convex estimator of the regression function  $C$  is then given by

$$\tilde{C}_{h_d}(x, a) = \int_a^x \phi_{h_d}(\hat{C}')^{-1}(z) dz + \hat{C}(a) \quad (15)$$

for any  $a \in [0, 1]$ . Birke and Dette (2007) show that the so defined convex regression estimator has asymptotically the same and in finite samples a very similar behavior as the unconstrained estimator  $\hat{C}$  one starts with. For the setting of estimating a call price function it is important that, in addition to the convexity, the derivative of the estimator fulfills the boundary conditions described in section 2.2. This is not guaranteed by using the estimator  $\tilde{C}_{h_d}(x, a)$ . By cutting the derivative of the regression estimate or its isotone rearrangement at the bounds it can be forced to fulfill these conditions. So, if  $c_l$  and  $c_u$  denote the lower and upper bound of  $C'$  in (5),

$$\check{C}_{h_d}(x, a) = \int_a^x \phi_{h_d}(\check{C}')^{-1}(z) dz + \hat{C}(a) \quad (16)$$

with

$$\check{C}'(x) = \hat{C}'(x)I\{c_l \leq \hat{C}'(x) \leq c_u\} + c_l I\{\hat{C}'(x) < c_l\} + c_u I\{c_u < \hat{C}'(x)\}$$

and

$$\hat{C}_{h_d}(x, a) = \int_a^x \tilde{\phi}_{h_d}(\hat{C}')^{-1}(z) dz + \hat{C}(a) \quad (17)$$

with

$$\begin{aligned} \tilde{\phi}_{h_d}(\hat{C}')^{-1}(z) &= \phi_{h_d}(\hat{C}')^{-1}(z)I\{c_l \leq \phi_{h_d}(\hat{C}')^{-1}(z) \leq c_u\} \\ &\quad + c_l I\{\phi_{h_d}(\hat{C}')^{-1}(z) < c_l\} + c_u I\{c_u \leq \phi_{h_d}(\hat{C}')^{-1}(z)\} \end{aligned}$$

for all  $x \in [0, 1]$  and any fixed lower integration bound  $a \in [0, 1]$  are two reasonable estimates of the call price function. It has proven to be more manageable for the asymptotic consideration to cut after the isotonization but both methods result in nearly the same estimate. As described in Birke and Dette (2007) the choice of the lower integral bound  $a$  is irrelevant for the asymptotic behavior of the estimator but plays an important role in finite sample settings if the unconstrained estimator is not convex. Therefore we now define  $L^2$ -optimal estimators of the call price function, that means, convex estimators that minimize the  $L^2$ -distance to the unconstrained estimator  $\hat{C}$ .

**Theorem 3.1** *Let  $\Phi_{h_d}$  denote a function with  $(\Phi_{h_d})' = \phi_{h_d}(\check{C}')^{-1}$  or  $(\Phi_{h_d})' = \tilde{\phi}_{h_d}(\hat{C}')^{-1}$ , respectively. The point  $a^* \in [0, 1]$  minimizes the  $L^2$ -distance*

$$\int_0^1 (\hat{C}(x) - \check{C}_{h_d}(x, a))^2 dx$$

or

$$\int_0^1 (\hat{C}(x) - \hat{C}_{h_d}(x, a))^2 dx$$

if and only if it fulfills

$$\hat{C}(a^*) - \check{\Phi}_{h_d}(a^*) = \int_0^1 (\hat{C}(a) - \check{\Phi}_{h_d}(a)) da.$$

Then we have

$$\check{C}_{h_d}(x, a^*) = \int_0^1 \check{C}_{h_d}(x, a) da.$$

and

$$\hat{C}_{h_d}(x, a^*) = \int_0^1 \hat{C}_{h_d}(x, a) da$$

**Remark 3.1** A reasonable estimator of the SPD can be obtained by differentiating  $\phi_{h_d}(\check{C}')^{-1}$ , that is

$$\begin{aligned} e^{-r_{t,T}(T-t)} \hat{p}_Q(x) &= \frac{\partial}{\partial x} \phi_{h_d}(\check{C}')^{-1}(x) \\ &= \frac{1}{\phi_{h_d}(\check{C}')'(\phi_{h_d}(\check{C}')^{-1}(x))} \end{aligned}$$

and

$$\hat{p}_Q(x) = \frac{h_d e^{r_{t,T}(T-t)}}{\int_0^1 K_d\left(\frac{\check{C}'(v) - \phi_{h_d}(\check{C}')^{-1}(x)}{h_d}\right) dv}.$$

Because  $\check{\phi}_{h_d}(\hat{C}')^{-1}$  is not differentiable we cannot use this for estimating the SPD.

### 3.2 Asymptotic behavior

Birke and Dette (2007) show that the convex estimator defined there has the same asymptotic behavior as the unconstrained one, that is, both are consistent estimates for  $C$  and have asymptotically the same normal distribution. The aim of this section is to proof an analog result for the estimate  $\hat{C}_{h_d}(x, a)$ ,  $a \in (0, 1)$  by tracing it back to the situation in Birke and Dette (2007). For the sake of completeness the theorem given there is again stated in this section. During the whole section we restrict ourselves to the case of a local constant estimate as unconstrained estimator of the regression function, that is

$$\hat{C}(x) = \frac{\sum_{i=1}^n K_r\left(\frac{x-X_i}{h_r}\right) Y_i}{\sum_{i=1}^n K_r\left(\frac{x-X_i}{h_r}\right)} \quad (18)$$

with kernel  $K_r$  and bandwidth  $h_r$ .

**Theorem 3.2 (Birke and Dette, 2007)** Assume that  $K_d$  is a two times continuously differentiable kernel of order 2 and  $K_r$  is a three times continuously differentiable kernel of order 3. Let the bandwidths fulfill

$$\begin{aligned} h_d, h_r &\rightarrow 0, & nh_d, nh_r &\rightarrow \infty, & h_d/h_r^{3/2} &\rightarrow 0, \\ nh_r^7 &= O(1) & \text{and} & (\log h_r^{-1})^{3/2}/nh_r^5 h_d &= o(1). \end{aligned}$$

If the regression function  $C$  is strictly convex and three times continuously differentiable, then we have for any  $x \in (0, 1)$  with  $C''(x) > 0$  and any  $a \in (0, 1)$

$$\tilde{C}_{h_d}(x, a) - C(x) = \hat{C}(x) - C(x) + o_P\left(\frac{1}{\sqrt{nh_r}}\right)$$

The proof of this theorem strongly relies on the fact that the first derivative of the unconstrained estimator  $\hat{C}$  is still two times continuously differentiable. This is not the case for the convex estimate  $\check{C}_{h_d}(x, a)$  where we start with  $\check{C}'$  as an estimator of the derivative of  $C$ . Therefore we cannot use Theorem 3.2 to show its asymptotic normality. The estimate  $\hat{C}_{h_d}(x, a)$  has the advantage that its unconstrained estimator is still three times continuously differentiable and hence, we state the following theorem for  $\hat{C}_{h_d}(x, a)$ .

**Theorem 3.3** *Let  $K_d$  be a two times continuously differentiable kernel of order 2,  $K_r$  be a three times continuously differentiable kernel of order 3 and let the bandwidths fulfill*

$$h_d, h_r \rightarrow 0, \quad nh_d, nh_r \rightarrow \infty, \quad h_d/h_r^{3/2} \rightarrow 0, \quad h_r^2/h_d \rightarrow 0, \\ nh_r^7 = O(1) \quad \text{and} \quad (\log h_r^{-1})^{3/2}/nh_r^5 h_d = o(1).$$

If the regression function  $C$  is strictly convex and three times continuously differentiable, then we have for any  $x \in (0, 1)$  with  $C'''(x) > 0$  and any  $a \in (0, 1)$

$$\hat{C}_{h_d}(x, a) - C(x) = \hat{C}(x) - C(x) + o_P\left(\frac{1}{\sqrt{nh_r}}\right).$$

Theorem 3.3 yields that the constrained estimator  $\hat{C}_{h_d}$  of the call pricing function is consistent if the unconstrained estimator is consistent. Moreover, we obtain the following corollary about its asymptotic distribution.

**Corollary 3.1**

$$\sqrt{nh_r} \left( \hat{C}_{h_d}(x, a) - C(x) - b_n(x) \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \gamma(x))$$

for any  $x, a \in (0, 1)$ , where the bias and asymptotic variance are given by

$$b_n(x) = h_r^3 \kappa_3(K_r) \frac{(Cf)^{(3)} - Cf^{(3)}}{f}(x), \\ \gamma(x) = \int_{-1}^1 K_r^2(y) dy \left( \frac{\sigma^2}{f} \right)(x),$$

respectively and  $\kappa_3(K_r) = 1/3! \int_{-1}^1 u^3 K_r(u) du$ .

Because Theorem 3.3 and Corollary 3.1 hold for every  $a \in (0, 1)$  their assertions are also true for the  $L^2$ -optimal estimator of the call price function.

## 4 Finite Sample Behavior

### 4.1 The estimator of Aït-Sahalia and Duarte

We will give a brief review of the estimator proposed by Aït-Sahalia and Duarte (2003) which we will compare in the following simulation study to the estimator defined in the previous section. The method is a modification of the classical constrained least squares regression problem (see e.g. Brunk, 1970, Wright, 1982 or Dykstra, 1983) combined with a smoothing step. Assume that the sampled data of strikes and option prices  $\{X_i, Y_i\}_{1 \leq i \leq n}$  is already ordered with respect to increasing strike prices, i.e.  $X_i \leq X_j$  for  $1 \leq i < j \leq n$ . In the first step the option prices  $Y_i$  of the sampled data are substituted by values  $m_i$  which are calculated by the constrained least squares regression due to Dykstra (1983) (see also Appendix A in Aït-Sahalia and Duarte (2003) for the application of Dykstra's algorithm in this context). The algorithm consists in finding values  $\{m_i\}_{1 \leq i \leq n}$  by minimizing

$$\sum_{i=1}^n (m_i - Y_i)^2,$$

over all tuples  $(m_1, \dots, m_n) \in \mathbb{R}^n$  subject to the constraints motivated by conditions (5) and (6),

$$-e^{-r_{t,T}(T-t)} \leq \frac{m_{i+1} - m_i}{X_{i+1} - X_i} \leq 0 \quad \text{for all } 1 \leq i \leq n-1,$$

$$\frac{m_{i+2} - m_{i+1}}{X_{i+2} - X_{i+1}} \geq \frac{m_{i+1} - m_i}{X_{i+1} - X_i} \quad \text{for all } 1 \leq i \leq n-2.$$

In the second step, a local linear estimator is applied to the transformed data  $\{X_i, m_i\}_{1 \leq i \leq n}$ , i.e.  $\hat{C}^{(1)} = \hat{C}_{1,1}$  is calculated as in (4), but for the estimation of the SPD the true derivative  $\hat{C}^{(2)} = \hat{C}'_{1,1}$  is taken (instead of  $\hat{C}_{2,p}$  with some  $p \geq 2$  for example). This construction enables us to state, that not only the transformed data  $m_i$  but also the estimates satisfy the constraints  $-e^{-r_{t,T}(T-t)} \leq \hat{C}_{1,1}(x) \leq 0$  and  $\hat{C}'_{1,1}(x) \geq 0$  (see Proposition 1 in Aït-Sahalia and Duarte, 2003). Finally, the conditions given in (7) are satisfied by modifications of the estimators subject to (10) and (11). An estimator of the call price function is obtained from  $\hat{C}'_{1,1}$  by using equations (1) and (2), that is

$$\hat{C}_{AD}(K) = \int_K^\infty (S_T - K) \hat{C}'_{1,1}(S_T) dS_T.$$

### 4.2 Simulation Study

We consider a log-normal SPD,

$$\varphi_{\log}(x) = \frac{1}{\sqrt{2\pi\nu x}} \exp\left\{-\frac{1}{2}\left(\frac{\ln(x) - \mu}{\nu}\right)^2\right\}.$$

It is well known that in the Black, Scholes and Merton framework (see Black and Scholes, 1973 and Merton, 1973), the price of a call option,

$$C(K) \equiv C(t, T, S_t, K, r_{t,T}, d_{t,T}) = e^{-d_{t,T}(T-t)} S_t N(d_1) - K e^{-r_{t,T}(T-t)} N(d_2),$$

where  $N(\cdot)$  denotes the standard normal distribution function,  $s$  the (annualized) volatility of the log-returns of the underlying asset price and

$$d_1 = \frac{\ln(S_t/K) + (r_{t,T} - d_{t,T} + \frac{1}{2}s^2)(T-t)}{s\sqrt{T-t}}, \quad d_2 = d_1 - s\sqrt{T-t},$$

induces a lognormal SPD for the random variable  $\ln(S_T/S_t)$  with mean  $\mu = (r_{t,T} - d_{t,T} - \frac{1}{2}s^2)(T-t)$  and variance  $\nu^2 = s^2(T-t)$ . For simplicity we have set the interest rate  $r_{t,T}$  and the dividend yield  $d_{t,T}$  equal to zero and the other parameters to common values (except  $\mu$  which calculates from the relation above),

$$S_t = 100.0, \quad \mu = 4.5851, \quad \nu = 0.2 \quad \text{and} \quad T - t = 1.$$

In practice, not for every considerable strike in the interval  $[0, +\infty]$  of the underlying asset options are traded, but typically only for some certain strikes in a range  $[a, b]$  around the actual spot value  $S_t$ . If the underlying asset trades in a bullish regime the actual spot  $S_t$  will be closer to  $b$ , and if it trades in a bearish regime it will be closer to  $a$ . To depict this we assume that we can observe  $n = 51$  tradeable call prices with strikes equally spaced within the interval  $[50, 170]$ . Due to different liquidity of the options, the existence of bid/ask spreads and transactions costs, the observed call prices do not exactly match the theoretical prices of the call price function. Therefore, the model for the observed call prices is

$$Y_i = C(K_i) + \varepsilon_i \quad (i = 1, \dots, 51 = n), \quad (19)$$

with i.i.d.  $\varepsilon_i \sim N(0, \sigma^2)$  and  $\sigma = 1.5$ . Note that this model does not exclude sampled call prices which are negative and thus allow for arbitrage. In our opinion the model is still appropriate, if a heteroscedastic variance function is used, i.e.  $\text{Var}(\varepsilon_i) = \sigma^2(K_i)$  for all strikes. The variance function  $\sigma^2$  must then reflect what can be observed in practice, namely that the variance of the sampled call prices tends to zero, when strikes increase and call prices get close to zero, respectively. Since we want to compare the estimator proposed in the previous section to the unconstrained kernel estimator, we refrain to model the variance homoscedastically to make the differences more clearly, which arise from the two approaches.

Figure 1 shows from left to right the true call price function with one of the simulated samples, the derivative of the call price function and the SPD.

The estimator is now constructed in three steps. First, for 80 equidistant values  $a$  in the interval  $[50, 170]$  estimators  $\hat{C}_{h_d}(x, a)$  are computed as given in (17). According to theorem 3.1 the expression  $\hat{C}_{h_d}(x) \equiv \hat{C}_{h_d}(x, a^*) = \frac{170-50}{80} \sum_{i=1}^{80} \hat{C}_{h_d}(x, a_i)$  is calculated in a second step. Finally, to satisfy the conditions (8) and (9) different corrections can be adopted. Ait-Sahalia and Duarte (2003) propose (see section 3.6 in their paper) to adjust their estimate  $\hat{C}^{(2)}(x)$  of the SPD by

$$e^{-r_{t,T}(T-t)} \hat{C}^{(2)}(x-z) / \int_0^{+\infty} \hat{C}^{(2)}(u) du, \quad (20)$$

for an appropriate translation parameter  $z$ , to fulfill conditions (10) and (11) in Remark 2.2 (or conditions (2.8) and (2.9) in their paper, respectively). This method can also be applied to

our estimator  $\hat{p}_Q(x)$  given in Remark 3.1. This suggests to replace the integral boundaries by  $a$  and  $b$ , respectively, if calls are only observed for strikes in  $[a, b]$ , which in turn implies that values of the SPD outside the interval must be zero. This of course is not true in general.

Since our estimator  $\hat{C}_{h_d}(x)$  is constructed without using the corresponding SPD, we choose to modify it with respect to conditions (8) and (9) by making an affine transformation  $\hat{C}_{\text{mod}}(x) \equiv \alpha + \beta \cdot \hat{C}_{h_d}(x)$ . Here  $\alpha$  and  $\beta$  are determined by minimizing  $\ell(\alpha, \beta) = \alpha^2 + (1 - \beta)^2$  under the constraints  $\hat{C}_{\text{mod}}(a) = S_t e^{-d_{t,T}(T-t)} - a e^{-r_{t,T}(T-t)}$  and  $\hat{C}_{\text{mod}}(b) = 0$ . Again, these constraints imply that the SPD outside the interval  $[a, b]$  is zero.

The bandwidth  $h_r$  is chosen by using the estimator  $\tilde{\sigma}$  for the integrated variance proposed by Rice (1984). Therefore let  $(K_{(i)}, Y_{[i]})$  ( $i = 1, \dots, n$ ) denote the sample of strikes in increasing order with their corresponding call prices. Then the integrated variance is estimated by

$$\tilde{\sigma}^2 = \frac{1}{2(n-1)} \sum_{i=1}^{n-1} (Y_{[i+1]} - Y_{[i]})^2,$$

from which the bandwidth for the regression step

$$h_r = (b - a) \cdot \left( \frac{\tilde{\sigma}^2}{n} \right)^{\gamma_r}$$

and the bandwidth for the density step

$$h_d = h_r^{\gamma_d}.$$

are calculated. In this finite sample simulation study we set  $\gamma_r = 2/3$  and  $\gamma_d = 3.0$ .

In Figure 2 the mean call price function from 500 simulations and its mean derivative are plotted against the true functions (solid line). The estimator proposed in this paper is represented by the dashed line, and the estimator of Aït-Sahalia and Duarte (2003) by the dotted line. It can be seen that both estimators lie very close to the true function.

Bias, variance and mean squared error (MSE) are shown in Figure 3. For comparison we have also plotted these functions for the unconstrained local polynomial kernel estimator applied to the sample data directly. It can be seen that for bias, variance and MSE the constrained estimator performs better than the unconstrained estimator in almost all regions of the interval  $[50, 170]$

## 5 Conclusion

As already pointed out in the introduction, there are several arguments for estimating the call price function nonparametrically. In this paper we propose a completely kernel based estimate of the call price function which fulfills all constraints given by the no-arbitrage principle. One of the major advantages of our method is that we do not have to estimate the state price density first which is essentially the second derivative of the call price function and can only be estimated with a worse rate than the call price function in kernel regression. Our estimator is

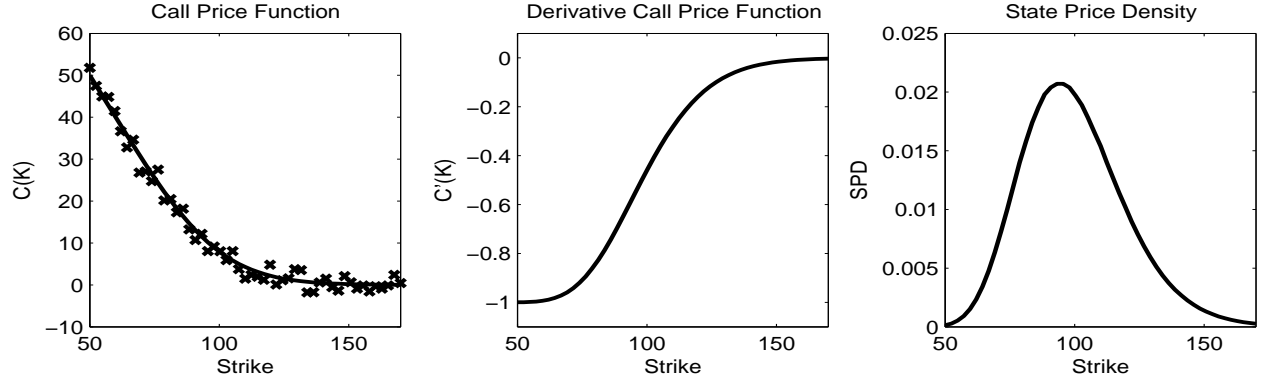


Figure 1: Left: True call price function (solid line) with one of the 500 samples (crosses). Middle: Derivative of the true call price function. Right: State price density.

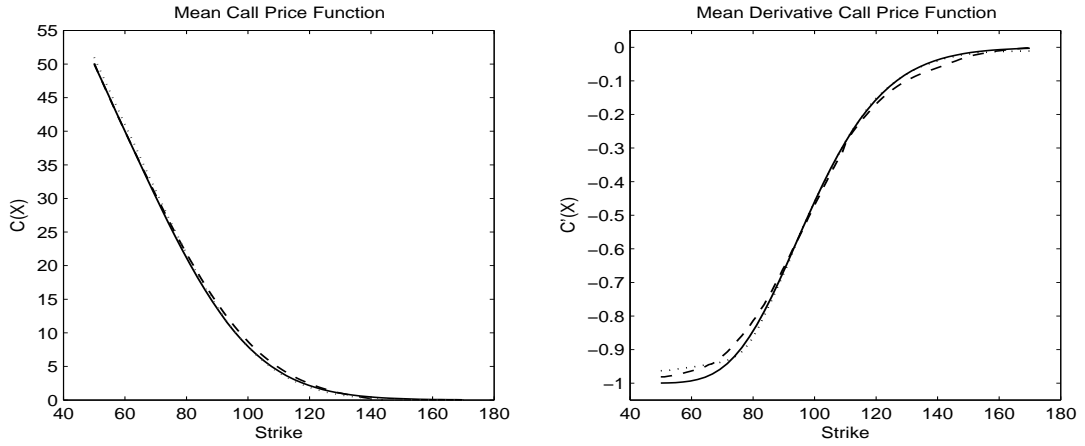


Figure 2: Left: The true call price function (solid line) and the mean of the estimated call price functions (dashed line: the estimator proposed in this paper, dotted line: the estimator of Aït-Sahalia and Duarte, 2003). Right: The mean derivative of our estimated call price function (dashed line) and that of Aït-Sahalia and Duarte (2003) (dotted line) and the true derivative (solid line).

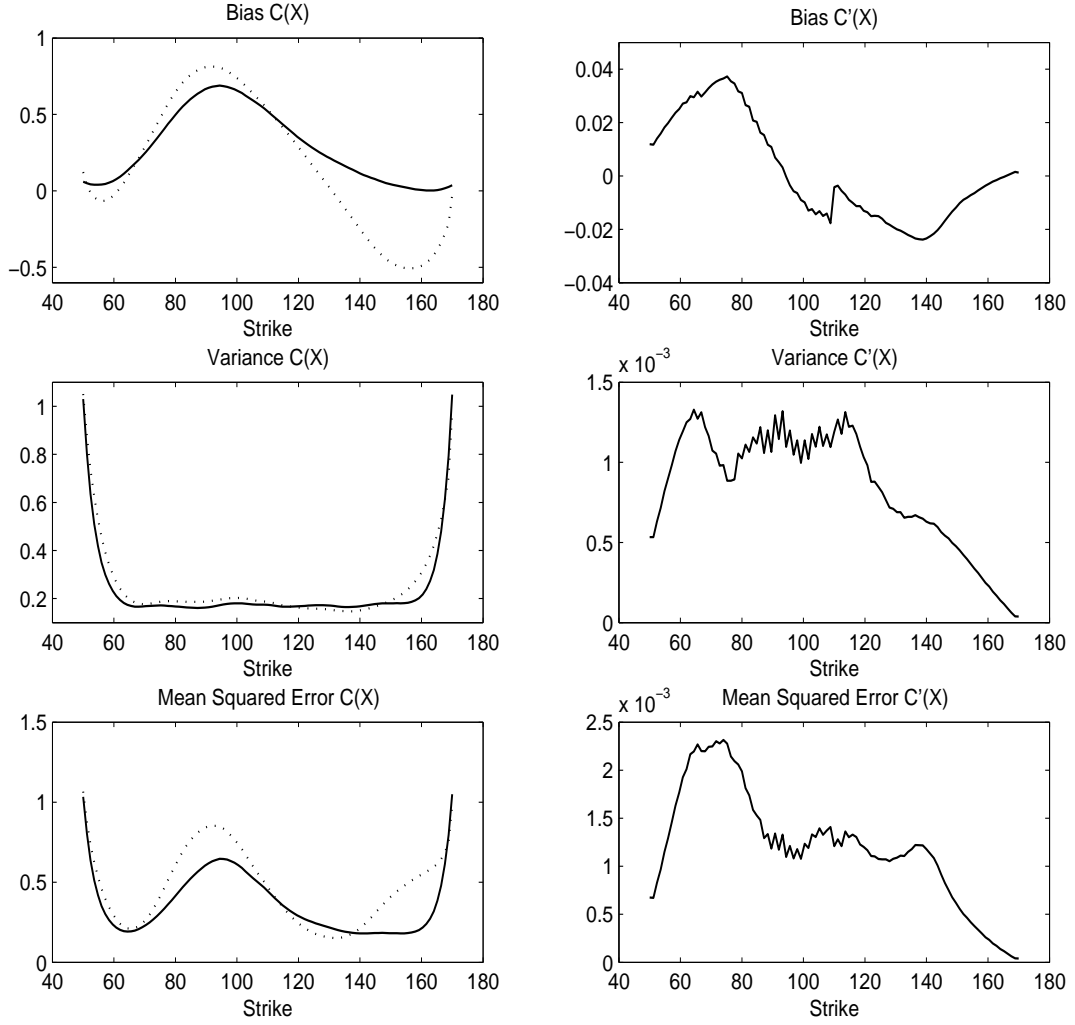


Figure 3: Left Column: Bias, variance and MSE for the estimated call price function with constraints (dashed line) and for the unconstrained local polynomial kernel estimator (dotted line). Right Column: Bias, variance and MSE for the derivative of the call price function with constraints .



asymptotically normal with the same bias and variance as the unconstrained estimate one starts with. Also in finite samples both have a very similar behavior but we see slight advantages for the constrained estimator in the MSE. The results for our estimator are also comparable with those of Aït-Sahalia and Duarte (2003). However, in contrast to the procedure proposed by these authors, the method suggested in this paper does not require constrained optimization techniques which might be computationally extensive.

**Acknowledgements.** The authors are grateful to Prof. Dr. Holger Dette for his helpful comments on this topic. We also would like to thank Y. Aït-Sahalia for some helpful hints concerning their paper and the simulations. The work of one of the authors was supported by the Sonderforschungsbereich 475, Komplexitätsreduktion in multivariaten Datenstrukturen.

## Appendix: Proofs

**Proof of Theorem 3.1.** If  $\Delta(x)$  represents the difference  $\hat{C}(x) - \Phi_{h_d}(x)$  with  $(\Phi_{h_d})' = \phi_{h_d}(\check{C}'')^{-1}$  or  $(\Phi_{h_d})' = \tilde{\phi}_{h_d}(\hat{C}'')^{-1}$  respectively, we have to compute

$$\min_{a \in [0,1]} \int_0^1 (\Delta(a) - \Delta(x))^2 dx = \min_{g \in \Delta([0,1])} \int_0^1 (g - \Delta(x))^2 dx$$

while the last equality holds because of the continuity of  $\Delta$  in both cases. From that it follows that

$$\arg \min_{g \in \Delta([0,1])} \int_0^1 (g - \Delta(x))^2 dx = \int_0^1 \Delta(x) dx.$$

By again using the continuity of  $\Delta$  there has to exist at least one point  $a^* \in [0, 1]$  such that

$$\int_0^1 \Delta(x) dx = \Delta(a^*).$$

Then the difference between the  $L^2$ -optimal convex and the unconstrained estimator has the representation

$$\Delta(a^*) - \Delta(x) = \int_0^1 \Delta(a) da - \Delta(x) = \int_0^1 (\Delta(a) - \Delta(x)) da$$

which is equal to  $\int_0^1 \check{C}_{h_d}(x, a) da$  or  $\int_0^1 \hat{C}_{h_d}(x, a) da$  respectively. □

**Sketch of Proof for Theorem 3.2.** For the sake of completeness and to show the necessity of a two times continuously differentiable unconstrained estimator mentioned in section 3.2 we present the essential steps of the proof here and refer for a detailed proof to Birke and Dette (2007) or Birke (2007).

In a first step we express the increasing rearrangement of the unconstrained estimator by its distribution function using a functional Taylor expansion of the operator which maps a function to its quantile.

This yields for the estimate  $\tilde{C}_{h_d}$  in (15) the representation

$$\begin{aligned}\tilde{C}_{h_d}(x, a) - C(x) &= \int_a^x \left( \phi_{h_d}(\hat{C}')^{-1}(z) - C'(z) \right) dz + \left( \hat{C}(a) - C(a) \right) \\ &= \int_a^x A_n(z) dz + \frac{1}{2} \int_a^x B_{n, \lambda^*}(z) dz + \left( \hat{C}(a) - C(a) \right)\end{aligned}\quad (21)$$

for

$$A_n(t) = -\frac{\phi_{h_d}(\hat{C}') - C'^{-1}}{(C'^{-1})'} \circ C'(t), \quad (22)$$

a remainder  $B_{n, \lambda^*}(z)$  and some  $\lambda^* \in [0, 1]$ .

We now treat the two terms in this expansion separately. Recalling the definition of  $A_n(t)$  we have

$$A_n = \int_a^x A_n(z) dz = A_{n,1} + A_{n,2} + A_{n,3} \quad (23)$$

with

$$\begin{aligned}A_{n,1} &= \frac{1}{h_d} \int_0^1 \int_{C'(a)}^{C'(x)} K_d\left(\frac{C'(v) - t}{h_d}\right) dt (\hat{C}(v) - C(v))' dv \\ A_{n,2} &= \frac{1}{h_d^2} \int_0^1 \int_{C'(a)}^{C'(x)} K_d'\left(\frac{\xi(u, v) - t}{h_d}\right) (\hat{C}'(v) - C'(v))^2 dv dt \\ A_{n,3} &= - \int_{C'(a)}^{C'(x)} (\phi_{h_d}(C')(t) - C'^{-1}(t)) dt\end{aligned}$$

and  $|\xi(u, v) - C'(v)| < |\hat{C}'(v) - C'(v)|$ . For the first term we have (observing that the inequality  $0 < a < x < 1$  implies  $C'(a) < C'(x)$ )

$$\begin{aligned}A_{n,1}(x) &= \frac{1}{h_d} \int_0^1 \int_{C'(a)}^{C'(x)} K_d\left(\frac{C'(v) - t}{h_d}\right) dt \left( \hat{C}(v) - C(v) \right)' dv \\ &= \int_{C'^{-1}(C'(a)+h_d)}^{C'^{-1}(C'(x)-h_d)} \int_{-1}^1 K_d(t) dt \left( \hat{C}(v) - C(v) \right)' dv \\ &\quad + \int_0^{C'^{-1}(C'(a)+h_d)} \int_{-1}^{\frac{C'(v)-C'(a)}{h_d}} K_d(t) dt \left( \hat{C}(v) - C(v) \right)' dv \\ &\quad + \int_{C'^{-1}(C'(x)-h_d)}^1 \int_{\frac{C'(v)-C'(x)}{h_d}}^1 K_d(t) dt \left( \hat{C}(v) - C(v) \right)' dv \\ &= \Delta_n^{(1.1)}(x) + \Delta_n^{(1.2)}(x) + \Delta_n^{(1.3)}(x),\end{aligned}\quad (24)$$

where the last line defines the terms  $\Delta_n^{(1.j)}$  ( $j = 1, 2, 3$ ),  $\hat{C}(x)$  denotes the Nadaraya-Watson estimate of the regression function defined in (18) and we have used the fact that our construction is based on  $\hat{C}'(x) = \partial \hat{C}(x) / \partial x$  as an estimate for the derivative of the regression function.

The term  $\Delta_n^{(1.1)}$  can be estimated as

$$\Delta_n^{(1.1)}(x) = \hat{m}(x) - m(x) - (\hat{m}(a) - m(a)) + o_P\left(\frac{1}{\sqrt{nh_r}}\right)$$

by using another Taylor expansion of order 2. A straight forward but tedious calculation shows in a similar way that the terms  $\Delta_n^{(1.j)}$ ,  $A_{n,j}$ ,  $j = 2, 3$  and  $\int_a^x B_{n,\lambda^*}(z) dz$  are of order  $o_P(1/\sqrt{nh_r})$ . For details of this calculation see Birke and Dette (2007) or Birke (2007). Therefore, an at least two times continuously differentiable unconstrained estimator is essential to obtain the asymptotic behavior of the constrained estimator.  $\square$

**Proof of Theorem 3.3.** In addition to (17)  $\hat{C}_{h_d}(x, a)$  can be written as

$$\begin{aligned} \hat{C}_{h_d}(x, a) - C(x) &= \int_a^x \phi_{h_d}(\hat{C}')^{-1}(z) dz + \hat{C}(a) - C(x) \\ &\quad + \int_a^x (c_l - \phi_{h_d}(\hat{C}')^{-1}(z)) I\{\phi_{h_d}(\hat{C}')^{-1}(z) \leq c_l\} dz \\ &\quad + \int_a^x (c_u - \phi_{h_d}(\hat{C}')^{-1}(z)) I\{\phi_{h_d}(\hat{C}')^{-1}(z) \geq c_u\} dz \\ &= \tilde{C}_{h_d}(x, a) - C(x) + R_1 + R_2. \end{aligned} \quad (25)$$

On the first part of the right hand side of (25) we can apply Theorem 3.2. The proof is complete if we are able to show that both remainders  $R_1$  and  $R_2$  are of order  $o_P(1/\sqrt{nh_r})$ . Because of the similarity of both remainders we only discuss  $R_2$  here. The estimation of  $R_1$  follows in a similar way. It is

$$\begin{aligned} |R_2| &\leq \int_a^x |c_u - \phi_{h_d}(\hat{C}')^{-1}(z)| I\{\phi_{h_d}(\hat{C}')^{-1}(z) \geq c_u\} dz \\ &= \int_a^x |C'(1) - \phi_{h_d}(\hat{C}')^{-1}(z)| I\{\phi_{h_d}(\hat{C}')^{-1}(z) \geq c_u\} dz \\ &\leq \int_a^x (|C'(1) - C'(z)| + |C'(z) - \phi_{h_d}(\hat{C}')^{-1}(z)|) I\{\phi_{h_d}(\hat{C}')^{-1}(z) \geq c_u\} dz \\ &= \int_a^x (O(1) + |C'(z) - \phi_{h_d}(\hat{C}')^{-1}(z)|) I\{\phi_{h_d}(\hat{C}')^{-1}(z) \geq c_u\} dz. \end{aligned} \quad (26)$$

The estimate  $\hat{C}'$  is uniformly consistent with convergence rate  $O(\log h_r^{-1}/nh_r^3)^{1/2}$  (similar proof as in Mack and Silverman, 1982). With Theorem 3.1 in Neumeyer (2006) and the conditions  $h_r^2/h_d \rightarrow 0$  and  $(\log h_r^{-1})^{3/2}/nh_r^5 h_d \rightarrow 0$  we almost surely obtain

$$|C'(z) - \phi_{h_d}(\hat{C}')^{-1}(z)| = O\left(\left(\frac{\log h_r^{-1}}{nh_r^3}\right)^{1/2} + \frac{1}{h_d} \frac{\log h_r^{-1}}{nh_r^3} + \frac{1}{h_d^3} \left(\frac{\log h_r^{-1}}{nh_r^3}\right)^{3/2} + h_d^2\right) = o(1)$$

uniformly for every  $z \in [a, x]$ , which means that the estimate  $\phi_{h_d}(\hat{C}')^{-1}(z)$  is uniformly consistent for  $C'$ . Therefore the estimation in (26) can be continued as

$$|R_2| \leq \int_a^x (O(1) + |m'(z) - \phi_{h_d}(\hat{C}')^{-1}(z)|) I\{\phi_{h_d}(\hat{C}')^{-1}(z) \geq c_u\} dz$$

$$= O_P(1) \int_a^x I\{\phi_{h_d}(\hat{C}')^{-1}(z) - C'(z) \geq c_u - C'(z)\} dz. \quad (27)$$

Note that  $\phi_{h_d}(\hat{C}')^{-1}(z) \geq C'(z)$  holds for every  $z \in [a, x]$  with  $\phi_{h_d}(\hat{C}')^{-1}(z) \geq c_u$ . We then obtain by using the Markov inequality

$$\begin{aligned} E \left[ \int_a^x I\{\phi_{h_d}(\hat{C}')^{-1}(z) - C'(z) \geq c_u - C'(z)\} dz \right] &= \int_a^x P \left( |\phi_{h_d}(\hat{C}')^{-1}(z) - C'(z)| \geq c_u - C'(z) \right) dz \\ &\leq \int_a^x P \left( |\phi_{h_d}(\hat{C}')^{-1}(z) - C'(z)| \geq C'(1) - C'(x) \right) dz \\ &\leq \int_a^x \frac{\text{MSE}(\phi_{h_d}(\hat{C}')^{-1}(z))}{(C'(1) - C'(x))^2} dz = O\left(\frac{1}{nh_r^3}\right) \end{aligned}$$

because the first derivative of  $C$  is strictly increasing and therefore the denominator in the Markov inequality is bigger than 0. The estimation of  $R_2$  now results in

$$R_2 = O_P\left(\frac{1}{nh_r^3}\right) = o_P\left(\frac{1}{\sqrt{nh_r}}\right)$$

which concludes the proof of Theorem 3.3. □

## References

- Aït-Sahalia, Y. and Duarte, J. (2003). Nonparametric Option Pricing under Shape Restrictions. *J. Econometrics*, **116**, 9-47
- Aït-Sahalia, Y. and Lo, A.W. (1998). Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices. *J. Finance*, **53** (2), 499-547.
- Banz, R. and Miller, M. (1978). Prices for State-Contingent Claims: Some Estimates and Applications. *J. Business*, **51** (4), 653-672.
- Bennett, C. and Sharpley, R. (1988). *Interpolation of Operators*. Academic Press, N.Y.
- Birke, M. (2007). Schätz- und Testverfahren in der nichtparametrischen Regression unter qualitativen Annahmen. PhD thesis, Ruhr-Universität Bochum (in German).
- Birke, M. and Dette, H. (2007). Estimating a convex function in nonparametric regression. *Scand. J. Statist.* **34**, 384 - 404.
- Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *J. Polit. Econ.*, **81**, 637-659.
- Breedon, D. and Litzenberger, R. (1978). Prices of State-Contingent Claims Implicit in Options Prices. *J. of Business*, **51** (4), 621-651.

- Brunk, H.D. (1970) Estimation of isotonic regression. *Nonparametric Techniques in Statistical Inference* . 177-197, Cambridge Univ. Press, London
- Campa, J., Chang, K. and Reider, R. (1998). Implied Exchange Rate Distributions: Evidence from OTC Option Markets. *J. Intern. Money Finance*, **17** (1), 117-160.
- Dette, H., Neumeyer, N. and Pilz, K. F. (2006). A simple nonparametric estimator of a strictly monotone regression function. *Bernoulli* **12**, 469-490.
- Dykstra, R.L. (1983). An Algorithm for restricted Least Squares. *J. Americ. Stat. Assoc.*, **78**, 837-842.
- Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman & Hall/CRC.
- Fengler, M.R. (2005). *Semiparametric Modeling of Implied Volatility*. Springer.
- Hördahl, P. (1999). Estimating the Implied Distribution of the Future Short Term Interest Rate Using the Longstaff-Schartz Model. Working Paper, European Central Bank.
- Jackwerth, J.C. (2004). *Option-Implied Risk-Neutral Distributions and Risk Aversion*. The Research Foundation of AMIR.
- Jarrow, R. and Rudd, A. (1982). Approximate Valuation for Arbitrary Stochastic Processes. *J. Financial Economics*, **10** (3), 347-369.
- Longstaff, F. (1995). Option Pricing and the Martingale Restriction. *Rew. Fin. Studies*, **8** (4), 1091-1124.
- Mack, Y. P. and Silverman, B. W. (1982). Weak and strong uniform consistency of kernel regression estimates. *Z. Wahrsch. Verw. Gebiete* **61**, 405-415.
- Madan, D. and Milne, F. (1994). Contingent Claim Valued and Hedged by Pricing and Investing in a Basis. *Math. Finance*, **4**, 223-245.
- Malz, A.M. (1996). Using Option Prices to Estimate Realignment Probabilities in the European Monetary System: The Case of Sterling-Mark. *J. Int. Money Finance*, **15**, 717-748.
- Melick, W. and Thomas, C. (1997). Recovering an Asset's Implied PDF from Option Prices: An Application to Crude Oil during the Gulf Crisis. *J. Financial Quantitative Analysis*, **32** (1), 91-115.
- Merton, R.C. (1973). Theory of rational option pricing. *Bell J. Econ. Manage. Sci.*, **4**, 141-183.
- Neumeyer, N. (2007). A note on uniform consistency of monotone function estimators. *Statist. Probab. Lett.* **77**, 693-703. <http://www.ruhr-uni-bochum.de/mathematik3/preprint.htm>

- Potters, M., Cont, R. and Bouchaud, J. (1998). Financial Markets as Adaptive Systems. *Europohysics Letters*, **41** (3), 239-244.
- Rice, J. (1984). Bandwidth Choice for Nonparametric Regression. *Ann. Statist.*, 12, 1215-1230.
- Rubinstein, M (1994). Implied Binomial Trees. *J. Finance*, **49** (3), 771-818.
- Shimko, D. (1993). Bounds on Probability. *Risk*, **6** (4), 33-37.
- Wright, F.T. (1981). The asymptotic behavior of monotone regression estimates. *Ann Statist.* **9**, 443-448.
- Yatchew, A., Härdle, W. (2006). Nonparametric state price density estimation using constrained least squares and the bootstrap. *J. Econometrics* **133**, 579-599.