Nonparametric Estimation of Risk-Neutral Densities

Maria Grith[2], Wolfgang Karl Härdle[1] and Melanie Schienle[3]

[1] Professor at Humboldt-Universität zu Berlin, Ladislaus von Bortkiewicz Chair of Statistics and Director of CASE - Center for Applied Statistics and Economics, Spandauer Straße 1, 10178 Berlin, Germany and National Central University, Graduate Institute of Statistics, CDA - Centre for Complex Data Analysis, No. 300, Jhongda Rd., Jhongli City, Taoyuan County 32001,Taiwan, (R.O.C.), Email: haerdle@wiwi.hu-berlin.de

[2] Research Associate at Humboldt-Universität zu Berlin, Ladislaus von Bortkiewicz Chair of Statistics, Spandauer Straße 1, 10178 Berlin, Germany, Email: gritmari@wiwi.hu-berlin.de

[3] Professor at Humboldt-Universität zu Berlin, Chair of Econometrics and CASE - Center for Applied Statistics and Economics, Spandauer Straße 1, 10178 Berlin Germany, Email: melanie.schienle@wiwi.hu-berlin.de

Summary. This chapter deals with nonparametric estimation of the risk neutral density. We present three different approaches which do not require parametric functional assumptions on the underlying asset price dynamics nor on the distributional form of the risk neutral density. The first estimator is a kernel smoother of the second derivative of call prices, while the second procedure applies kernel type smoothing in the implied volatility domain. In the conceptually different third approach we assume the existence of a stochastic discount factor (pricing kernel) which establishes the risk neutral density conditional on the physical measure of the underlying asset. Via direct series type estimation of the pricing kernel we can derive an estimate of the risk neutral density by solving a constrained optimization problem. The methods are compared using European call option prices. The focus of the presentation is on practical aspects such as appropriate choice of smoothing parameters in order to facilitate the application of the techniques.

1 Introduction

Most of our economic understanding of investment under uncertainty is based on pure Arrow-Debreu securities (Arrow 1964, Debreu, 1959), which pay one unit of currency at the end of a period if a state of nature is realized and zero otherwise. Their theoretical state-contingent prices are the starting point for pricing any security in an economic equilibrium under uncertainty. In a continuum of states, the prices of the Arrow-Debreu securities can be characterized

by the state-price density, which yields one dollar if the final state is in the interval $[x, x + \mathrm{d}x]$ when starting from any point x. The existence and form of a state-price density can be justified by preference-based equilibrium models (Lucas, 1978) or reasoning from arbitrage-based models (Merton, 1973). We focus on the latter, where the state-price density is denoted as risk neutral density (RND). It exists if the underlying market is dynamically complete, i.e. any position can be replicated by a cash-flow neutral (self-financing) trading strategy over subsequent trades. We assume this for the rest of the chapter. Then the RND also uniquely characterizes the equivalent martingale measure under which all asset prices discounted at the risk-free rate are martingales.

In standard option pricing models such as Merton (1976), Heston (1993) or Bates (1996), estimation of the risk neutral density crucially depends on underlying model assumptions such as the underlying asset price dynamics and the statistical family of distributions that the risk neutral density is assumed to belong to. Consumption based asset pricing models prespecify preferences of the representative agent and condition therefore the shape of the pricing kernel (Lucas (1978) and Rubinstein (1976)). Recent empirical findings, however, question the validity of these popular specifications which drive the overall result (Campbell, Lo and McKinlay (1997)). Nonparametric estimation offers an alternative by avoiding possibly biased parametric restrictions and therefore reducing the respective misspecification risk. Since nonparametric estimation techniques require larger sample sizes for the same accuracy as a parametric estimation procedure, increasing availability of large data sets as intraday traded option prices have raised their feasibility. On the other hand, due to their flexibility, many existing nonparametric risk neutral density estimation techniques are afflicted by irregularities such as data sparsity in the tails, negative probabilities and failure of integration to unity. We will address these problems by appropriate choices of smoothing parameters, by employing semiparametric techniques or imposing relevant constraints.

We present a thorough picture of nonparametric estimation strategies for the RND q: We study direct standard kernel based approaches (local polynomial regression) which are flexible and yield point estimates as opposed to series expansion, sieve methods or splines. Though shape constraints such as convexity or monotonicity of the call price are hard to incorporate directly in the estimation step. Therefore, in particular in small samples, they are not satisfied leading to problems with economic theory. Thus we also propose an indirect way of estimation by employing series methods for directly controlling constraints in the estimation. In the following, we will briefly outline the main ideas for direct or indirect estimation of q.

In a dynamically complete market, the price of a European call is obtained by discounting the expected payoff, where the expectation is taken with respect to the risk neutral measure

$$C(X, \tau, r_{t,\tau}, \delta_{t,\tau}, S_t) = e^{-r_{t,\tau}\tau} \int_0^\infty (S_T - X)^+ q(S_T|\tau, r_{t,\tau}, \delta_{t,\tau}, S_t)\, dS_T. \quad (1)$$

Here S_t is the underlying asset price at time t, X the strike price, τ the time to maturity, $T = t + \tau$ the expiration date, $r_{t,\tau}$ the deterministic risk free interest rate at t until maturity T, $\delta_{t,\tau}$ the corresponding dividend yield of the asset, and $q(S_T|\tau, r_{t,\tau}, \delta_{t,\tau}, S_t)$ is the conditional risk neutral density. We assume that these state variables contain all essential information needed for estimation of C and q while quantities such as stochastic market volatility, trading volumes, bid-ask spreads are negligible. We write $q(S_T)$ instead of $q(S_T|\cdot)$ to keep notation simple. The risk neutral density can be derived from (1) as

$$q(S_T) = e^{r_{t,\tau}\tau} \left\{ \frac{\partial^2 C}{\partial X^2} \right\}_{X=S_T}, \tag{2}$$

see Breeden and Litzenberger (1978). It has been exploited to derive two standard nonparametric kernel estimation strategies for q: Either obtain an estimate of the RND from estimating a continuous twice-differentiable call function in all its arguments from traded options by smoothing in the call price, or alternatively, by smoothing in the implied volatility space.

In addition to these standard approaches, here we also introduce a third indirect way via series estimation of the empirical pricing kernel. Assuming that all the variables other than X are fixed, the price of the European call option with strike price X expiring in τ years under the historical measure p is given by

$$\begin{aligned} C(X) &= e^{-r_{t,\tau}\tau} \int_0^\infty (S_T - X)^+ \frac{q(S_T)}{p(S_T)} p(S_T) dS_T \\ &= e^{-r_{t,\tau}\tau} \int_0^\infty (S_T - X)^+ m(S_T) p(S_T) dS_T, \end{aligned} \tag{3}$$

where p is the subjective density of the stock price at the expiration of the option, at time T and m is the so called pricing kernel characterizing the change of measure from q to p.

The rest of this chapter is organized as follows: Section 2 describes kernel based regression methods for direct estimation of the RND from the call price function, Section 3 introduces the pricing kernel concept and explains the indirect method of estimating RND, Section 4 concludes. Throughout the chapter, empirical studies using EUREX DAX Index based European option data illustrate the methods and compare their performance.

2 Estimation of RND based on the Second Derivative

The standard approach for a nonparametric estimator of the risk neutral density is by estimating the second derivative of the call price with respect to the strike X. Then an estimate for q is obtained by discounting according to

(2). Therefore in the following we focus on estimation of the second derivative of a regression function.

Call the d-dimensional vector of covariates $\mathbf{Z}$ which comprises all estimation relevant variables of $(X, \tau, r_{t,\tau}, \delta_{t,\tau}, S_t)$ from (1) and denote call prices as response Y. From paired observations Y_i and $\mathbf{Z}_i = (Z_{ik})_{k=1}^d$, for $i = 1, ..., n$ we want to estimate the following general, possibly nonlinear relationship

$$Y_i = C(\mathbf{Z}_i) + \varepsilon_i, \quad i = 1, ..., n \tag{4}$$

where $C(\cdot) : \mathbb{R}^d \to \mathbb{R}$ is a smooth function in all d dimensions and ε is i.i.d. with $\mathsf{E}[\varepsilon|\mathbf{Z}] = 0$.

Kernel based methods are local techniques for estimating the function C at any point $\mathbf{z}$ in its domain; they use a weighted average of the Y_i's to yield fitted values via

$$\widehat{C}(\mathbf{z}) = \sum_{i=1}^{n} w_i(\mathbf{z}) Y_i \tag{5}$$

where the weights $w_i(\mathbf{z})$ assigned to each point of fit $\mathbf{z}$ decline with the distance $|\mathbf{Z}_i - \mathbf{z}|$ and satisfy $\frac{1}{n}\sum_{i=1}^n w_i(\mathbf{z}) = 1$. Kernel regression methods use kernel functions to construct weights. A univariate kernel is a smooth, symmetric real-valued squared integrable function $K(u) : \mathbb{R} \to \mathbb{R}$ which integrates to one. We can think of a standard kernel function as a probability density with potentially compact support. Examples of such K are presented in Table 1.

Table 1. Kernel functions $K(u)$

Uniform	$\frac{1}{2}\,\mathrm{I}(\lvert u\rvert \le 1)$
Triangle	$(1 - \lvert u\rvert)\,\mathrm{I}(\lvert u\rvert \le 1)$
Epanechnikov	$\frac{3}{4}(1 - u^2)\,\mathrm{I}(\lvert u\rvert \le 1)$
Quartic (Biweight)	$\frac{15}{16}(1 - u^2)^2\,\mathrm{I}(\lvert u\rvert \le 1)$
Triweight	$\frac{35}{32}(1 - u^2)^3\,\mathrm{I}(\lvert u\rvert \le 1)$
Gaussian	$\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}u^2)$
Cosine	$\frac{\pi}{4}\cos(\frac{\pi}{2}u)\,\mathrm{I}(\lvert u\rvert \le 1)$

Furthermore, there exist more general types of kernel functions, so called higher order kernels, which can be used for bias refinements in the estimation, see Subsection 2.1. The order of a kernel $\upsilon > 0$ is defined as the first nonzero moment of the kernel, that is

$$\begin{aligned} &\int u^l K(u)du = 0, \quad l = 1, ..., \upsilon - 1 \\ &\int u^\upsilon K(u)du = \kappa_\upsilon \neq 0 \end{aligned} \tag{6}$$

and $\kappa_\upsilon < \infty$. Solving the system of equations (6) for kernel functions integrating to unity for a fixed υ, yields a υ-th order kernel. The larger υ, however, the more "wiggly" the resulting kernel becomes - covering more and more negative areas. Here we mostly consider standard second order kernels, which are nonnegative functions.

Set $K_h(u) = \frac{1}{h}K\left(\frac{u}{h}\right)$ for all $u \in \mathbb{R}$ where h is the bandwidth, the smoothing parameter. In a d-dimensional space, for each pair $\mathbf{z}$ and $\mathbf{Z}_i$ the multivariate kernel function $\mathcal{K}(\mathbf{z} - \mathbf{Z}_i) : \mathbb{R}^d \to \mathbb{R}$ must analogously fulfil

$$\mathcal{K}_H(\mathbf{z} - \mathbf{Z}_i) = \frac{1}{|H|}\mathcal{K}\{H^{-1}(\mathbf{z} - \mathbf{Z}_i)\}$$

where $H = \text{diag}(\tilde{h})$ is the diagonal matrix of bandwidths $\tilde{h} = [h_1, ..., h_d]$. The matrix H can in general also contain off-diagonal elements - but in practice such generality is not needed. Define the multidimensional kernel $\mathcal{K}_H(\mathbf{z} - \mathbf{Z}_i)$ as a product of univariate kernels

$$\mathcal{K}_H(\mathbf{z} - \mathbf{Z}_i) = \prod_{k=1}^{d} K\left(\frac{z_k - Z_{ik}}{h_k}\right).$$

For expositional simplicity we let $h_1 = \ldots = h_d = h$. Details on how to choose the optimal bandwidths are addressed in the next section.

The simplest case of choosing w_i in (5) is to use Nadaraya-Watson weights

$$w_i(\mathbf{z}) = \frac{\mathcal{K}_h(\mathbf{z} - \mathbf{Z}_i)}{\sum_{i=1}^{n}\mathcal{K}_h(\mathbf{z} - \mathbf{Z}_i)}.$$

These are a special constant case of general local polynomial weights derived below. Besides, other choices of weights such as in the k-nearest neighbour or the Gasser-Müller estimator are possible.

Estimators of the second derivative of a function are constructed by twice differentiating the estimator of the function. Such estimators, however, have inferior statistical properties and are therefore not included here (see e.g. Härdle at al. (2004) and Fan and Gijbels (1996 for details). We focus on local polynomial estimation which directly yields estimates of derivatives. The idea of local polynomial regression is based on Taylor expansion approximating an unknown function C at a point $\mathbf{z}$. In order to keep notation simple, we first illustrate the method for the univariate case. The multivariate case is systematically the same and will be sketched afterwards.

Locally, any sufficiently smooth function C can be approximated by a polynomial of degree p

$$\begin{aligned} C(Z_i) &= \sum_{j=0}^{p}\frac{C^{(j)}(z)}{j!}(Z_i - z)^j + \mathcal{O}\{(Z_i - z)^{p+1}\} \qquad (7) \\ &= \sum_{j=0}^{p}\beta_j(Z_i - z)^j + \mathcal{O}\{(Z_i - z)^{p+1}\} \end{aligned}$$

with $i = 1, ..., n$. Therefore by minimizing a locally weighted least squares regression

$$\min_{\beta} \sum_{i=1}^{n} \left\{ Y_i - \sum_{j=0}^{p} \beta_j(z)(z - Z_i)^j \right\}^2 \mathcal{K}_h(z - Z_i) \tag{8}$$

the solution $\hat{\beta}_0(z)$ provides an estimator of C at point z, while $j!\hat{\beta}_j(z)$, with $j = 1, ..., p$ are the estimated derivatives at that point. Closed forms for $\hat{\beta}(z) = (\hat{\beta}_0(z), \ldots, \hat{\beta}_p(z))$ can be obtained by solving (8) via equating the corresponding system of first order conditions to zero. As we are interested in an estimator for the second derivative of a function, we should choose $p \geq 2$. As will be outlined in the subsection below, for good statistical properties without requiring too much smoothness $p = 3$ will be a suitable choice.

In d-dimensional case, expansion (7) will include mixed terms which must be appropriately ordered. Then the interpretation of the coefficients is similar: $\widehat{\beta}_0(\mathbf{z})$ is the estimator of C at point $\mathbf{z}$, while $j!\widehat{\beta}_j(\mathbf{z}) = j! \left[\beta_{j1}(\mathbf{z}), \cdots, \beta_{j\mu}(\mathbf{z})\right]$ with $\mu = 1, ..., N_j$ is N_j-dimensional vector of j-th order derivatives of C evaluated at point $\mathbf{z}$. It is obvious that $N_0 = 1$ (β_0 is the local constant) and $N_1 = d$ (β_1 is the vector of partial derivatives) but for $j \geq 2$ the expansion contains cross order derivatives and the general formula for N_j is

$$N_j = \binom{d + j - 1}{j - 1} = \frac{(d + j - 1)!}{d!(j - 1)!}.$$

For example, when $j = 2$ we have $N_2 = d(d + 1)/2$ distinct derivatives and, $\nabla^{(2)}\hat{C}(\mathbf{z}) = 2\hat{\beta}_2(\mathbf{z})$ is the estimate of

$$\nabla^{(2)} C(\mathbf{z}) = \begin{pmatrix} \frac{\partial^2 C(\mathbf{z})}{\partial z_1^2} \\ \frac{\partial^2 C(\mathbf{z})}{\partial z_1 \partial z_2} \\ \vdots \\ \frac{\partial^2 C(\mathbf{z})}{\partial z_d^2} \end{pmatrix}.$$

For the estimation of the RND we are interested in the second derivative of the call price with respect to the strike price X. In our notation with $\mathbf{Z} = (X, \tau, r_{t,\tau}, \delta_{t,\tau}, S_t)$, this is $2\beta_{21}$. Thus

$$\hat{q}(S_T) = 2e^{r_{t,\tau}\tau}\hat{\beta}_{21}(S_T, \mathbf{z}_{-1}) = e^{r_{t,\tau}\tau} \left\{ \frac{\widehat{\partial^2 C}(\mathbf{z})}{\partial z_1^2} \right\}_{X = S_T}$$

with $\mathbf{z}_{-1} = (\tau, r_{t,\tau}, \delta_{t,\tau}, S_t)$.

2.1 Statistical Properties

Assume for simplicity that C is univariate and has continuous derivatives of total order $(p + 1)$. The probability density function f of Z is continuous, it

is $f \geq 0$, and f is $(p+1)$ times continuously differentiable. The kernel K is a bounded second order kernel with compact support and the $\mathsf{E}[\varepsilon^2|Z=z]$ exists and is continuous in z. Let $\widehat{C}^{(j)}$ denote the estimator of $C^{(j)}$ based on a p-th order local polynomial fit ($j \leq p$). The results below are standard and can be found for instance in Li and Racine (2007).

Theorem 1. *When $p-j$ is odd, the bias is*

$$\mathsf{E}\left[\widehat{C}^{(j)}(z)\right] - C^{(j)}(z) = h^{p-j+1} c_{1,j,p} \left\{ \frac{\omega^{(p+1)}(z)}{(p+1)!} \right\} + o(h^{p-j+1}). \tag{9}$$

When $p-j$ is even, the bias is

$$\begin{aligned} \mathsf{E}\left[\widehat{C}^{(j)}(z)\right] - C^{(j)}(z) = {} & h^{p-j+2} c_{2,j,p} \left\{ \frac{\omega^{(p+2)}(z)}{(p+2)!} \right\} \int u^{p+2} K(u) du \\ & + h^{p-j+2} c_{3,j,p} \left\{ \frac{\omega^{(p+1)}(z) f^{(1)}(z)}{f(z)(p+1)!} \right\} \end{aligned} \tag{10}$$

where $\omega(z) = \left\{\widehat{C}(z) - C(z)\right\} f(z)$. In either case

$$\mathsf{Var}\left(\widehat{C}^{(j)}(z)\right) = \left\{ \frac{c_{4,j,p}\sigma^2(z)}{nh^{2j+1}} \right\} + o\left\{(nh^{2j+1})^{-1}\right\} , \tag{11}$$

where $\sigma^2(z) = \mathsf{E}[\varepsilon^2|Z=z]$ is the residual variance. The exact form of the constants $c_{a,j,p}$ for $a = 1,2,3,4$ can be found in Ruppert and Wand (1994).

Theorem 1 provides asymptotic bias and variance expressions of local polynomial estimators of degree p for a general j-th derivative. For illustration consider the special case $p = 0$ and $j = 0$ - local constant estimation of a function. The bias is

$$\frac{h^2}{2} \left\{ C^{(2)}(z) + 2\frac{C^{(1)}(z) f^{(1)}(z)}{f(z)} \right\} \mu_2(K) , \tag{12}$$

with $\mu_2(K) = \int u^2 K(u)\, du$. For $p = 1$ and $j = 0$ - local linear estimation of a function - the bias becomes

$$\frac{h^2}{2} \left\{ C^{(2)}(z) \right\} \mu_2(K). \tag{13}$$

Observe in general from (9) and (10) that the bias for $p-j$ even contains an additional design dependent term with factor $\frac{f^{(1)}(z)}{f(z)}$ as opposed to the odd case. Sign and size of this quantity, however, depend on the shape of the underlying estimated function and the shape of f_Z. In particular at the boundary of the support of Z, small values of f inflate the entire term. Therefore odd values of $p-j$ are preferable avoiding such boundary bias problems

and pertaining the same variance. In our case, we are interested in the second derivative. We therefore choose the polynomial order $p = 3$ and not $p = 2$ according to Theorem 1.

With higher order kernels (6) of order υ and corresponding higher smoothness assumptions the bias in Theorem 1 can be further reduced to be of order h^{υ} for fixed p and j with $\upsilon > p - j + 2$ without changing the rate in the variance. In practice the order υ, however, cannot be chosen too large as with increasing υ the estimates have robustness problems in finite samples due to negative weights associated with the kernels (Müller (1988)).

Observe from Theorem 1 that kernel estimation of a derivative is harder than of the function itself. While the variance in the function estimation decreases with $\mathcal{O}(1/(nh))$ the corresponding rate in the second derivative is only $\mathcal{O}_P(1/(nh^5))$ which is much slower. Therefore the finite sample performance of second derivatives lacks the precision of the fit achieved for the function.

Rates of convergence can be judged according to the mean squared error (MSE). Assuming that $p - j$ is odd, it is

$$\begin{aligned} \mathrm{MSE}(z,h,j) &= \mathsf{E}\left[\hat{C}^{(j)}(z) - C^{(j)}(z)\right]^2 \\ &= \mathcal{O}\{\underbrace{h^{2(p-j+1)}}_{bias^2} + \underbrace{(nh^{2j+1})^{-1}}_{var}\}. \end{aligned} \tag{14}$$

For constructing confidence intervals of the nonparametric estimates use the following normality result

Theorem 2. *Under some additional moment assumptions it is for $p - j$ odd*

$$\sqrt{nh^{2j+1}}\{\hat{C}^{(j)}(z) - C^{(j)}(z)\} \to \mathsf{N}(0, V_j) \tag{15}$$

with V_j as in Theorem 1.

For a precise statement of the standard moment conditions see Li and Racine (2007). Analogous results to Theorem 1 and 2 hold for d-dimensional functions. The only remarkable systematic difference is that the dimension of the regressors enters in the rate of the variance which is then $\mathcal{O}_P\{(nh^{2j+d})^{-1}\}$. Likewise the rate of convergence to the asymptotic distribution also deteriorates with d and is nh^{2j+d}. This phenomenon is known in the literature as the curse of dimensionality capturing the fact that finite sample performance of nonparametric estimators decreases with an increasing number of regressors. Therefore in practice, appropriate semiparametric dimension reduction techniques are used. They keep high modeling flexibility but yield better finite sample properties in regression settings with more than three regressors. See Subsection 2.3 for details.

2.2 Selection of the Smoothing Parameter

In practice, most important for good nonparametric estimation results is an appropriate choice of bandwidth. Other parameters like the selection of a

suitable kernel K only have little influence on the final result in practice. Asymptotically the choice of K has no effect, and in finite samples its impact is negligible (see Marron and Nolan (1988)). For the choice of order p of the employed local polynomial estimator it is sufficient to follow the logic outlined above.

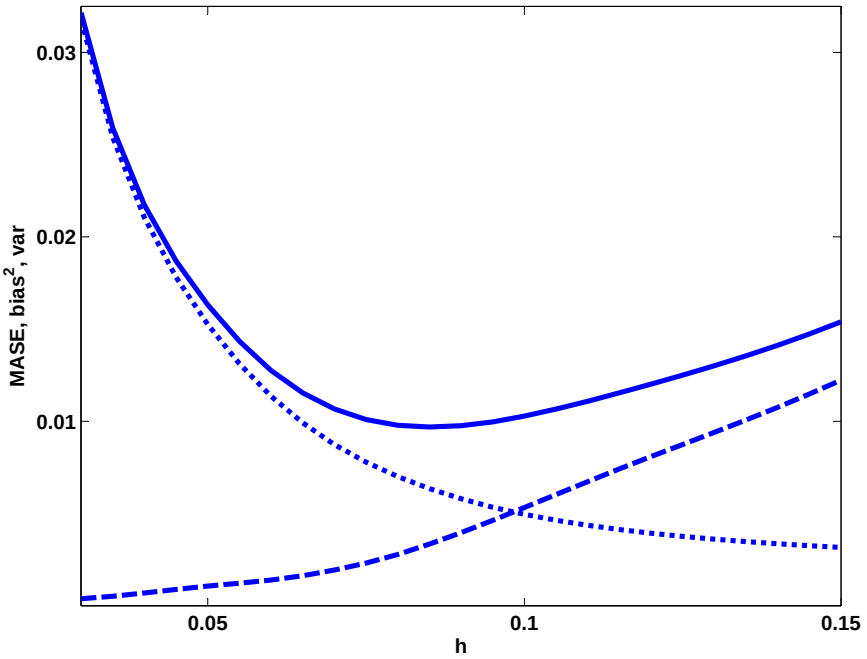

Fig. 1. MASE (solid line), squared bias (dashed line) and variance part (dotted line) for simulated data, weights $w(x) = \mathsf{I}(x \in [0.05, 0.95])$

An optimal bandwidth should minimize both bias and variance of the estimator. Though according to Theorem 1 there is a tradeoff between these quantities as smaller bandwidths would reduce the bias but inflate the variance. Therefore selecting h by minimizing $\mathrm{MSE}(\mathbf{z}, h, j)$ (multivariate analogue to (14)) balances bias and variance (see Figure 1 for an illustration in averages). However, such a choice depends on the location $\mathbf{z}$. For a global choice, use an integrated criterion like the weighted integrated mean square error (WIMSE)

$$\text{WIMSE}(h,j) = \int \text{MSE}(\mathbf{z},h,j)\psi(\mathbf{z})d\mathbf{z} = \int \mathsf{E}[\hat{C}^{(j)}(\mathbf{z}) - C^{(j)}(\mathbf{z})]^2\psi(\mathbf{z})d\mathbf{z} \tag{16}$$

where $\psi(\mathbf{z})$ is a nonnegative weight function which ensures that WIMSE is well defined. Instead of an integrated criterion also an averaged criterion like the mean average squared error (MASE) can be used which replaces integration with summation in (16). When using a second order kernel straightforward calculations yield

$$h^\star = \begin{cases} cn^{-1/(2p+d+2)} & \text{for } p-j \text{ odd} \\ c'n^{-1/(2p+d+4)} & \text{for } p-j \text{ even} \end{cases} \tag{17}$$

for the optimal bandwidth $h^\star$ in a multivariate setting with constants $c, c' > 0$ depending on kernel constants and higher derivatives of C and the density f of regressors at $\mathbf{z}$ as we can see from (9),(10) and (11). In our case of interest for $p = 3, j = 2$ and $d = 1$, it is $h^\star = n^{-1/9}$. As for larger j also p must be enlarged, the optimal bandwidth for estimating the j-th derivative decreases in j. Note, however, that $h^\star$ is not feasible in practice because the constants c, c' in (17) contain unknown quantities such as higher derivatives of C and the density f of regressors.

A way to operationalize these are plug-in methods. They replace unknown quantities by pilot estimates and then calculate $h^\star$ via (17). The rule-of-thumb additionally uses normality assumptions in the distribution of the regressors and for the kernel to calculate exact constants. For $p = j = 0$ it is $h_k \approx s_k n^{-1/(4+d)}$ for $h = (h_1, \ldots, h_d)$ with s_k the standard deviation of observations of covariate Z_k. It is an easy and fast way to obtain a rough estimate and can be used for pilot estimates in plug-in procedures. Nevertheless, a bandwidth choice based on these procedures yields only an asymptotically optimal selection as the employed criteria are asymptotic ones.

In small samples, however, there are better choices which can be made by data driven cross-validation (CV) methods. In general, these procedures yield valid finite sample bandwidth choices, but do not have closed form solutions. Therefore computation intensive numerical methods must be used in order to obtain such an automatic bandwidth h_{CV}. This can amount to a feasibility issue in particular for time series. We present a least squares cross-validation for local cubic estimation as our interest is in estimating the second derivative of C. Here, we select h_{CV} as minimizer of the sum of squared errors between obtained local cubic fit and observed response used as cross-validation criterion.

$$\text{CV}(\tilde{h}) = \sum_{i=1}^{n}\sum_{j\neq i}^{n}\Big\{Y_i - \widehat{C}_{\tilde{h},-i}(\mathbf{Z}_i) - \widehat{C}^{(1)}_{\tilde{h},-i}(\mathbf{Z}_i)(\mathbf{Z}_j - \mathbf{Z}_i) \\ -\frac{1}{2}\widehat{C}^{(2)}_{\tilde{h},-i}(\mathbf{Z}_i)(\mathbf{Z}_j - \mathbf{Z}_i)^2\Big\}^2 M(\mathbf{Z}_i) \tag{18}$$

where $0 \leq M(\mathbf{Z}_i) \leq 1$ is a weight function that ensures existence of the limit for n large. and $(\widehat{C}_{\tilde{h},-i}, \widehat{C}^{(1)}_{\tilde{h},-i}, \widehat{C}^{(2)}_{\tilde{h},-i})$ denote the local cubic regression estimate obtained without using the i-th observation $(\mathbf{Z}_i, C_i)$. This way we ensure that the observations used for calculating $\widehat{C}_{\tilde{h},-i}(\cdot)$ are independent of $\mathbf{Z}_i$. It can be shown that asymptotically h_{CV} converges to the corresponding theoretical bandwidth obtained from (17). This design driven choice of bandwidth is completely free of functional form assumptions and therefore most appealing in finite samples at the expense of potentially long computation time.

2.3 Dimension Reduction Techniques

While flexible, high-dimensional kernel regression requires large data samples for precise results in terms of tight pointwise confidence intervals. Ait-Sahalia and Lo (1998), for example, use one year option data to empirically derive the call function based on five-dimensional kernel regression. Asymptotically, rates of convergence of nonparametric estimators decrease the more regressors are included in the model. This is referred to as the "curse of dimensionality" (see Section 2.1. for theoretical details). Hence, there is a need to keep the dimension or equivalently the number of regressors low.

There exists a vast literature on methods which reduce the complexity of high dimensional regression problems resulting in better feasibility. In particular, the reduction of dimensionality is achieved by putting some structure on the model by e.g. imposing a parametric model or an additive or partially linear structure. The resulting models are so-called semiparametric models, among which the additive models are the most flexible kind requiring the least structural assumptions. In additive models, the regression function additively separates the influence of each univariate regressor. Thus estimation is restricted to a surface of the full-dimensional space of regressors $\mathbf{Z}$, which allows to construct estimators with univariate nonparametric rates of convergence and thus substantially improved finite sample properties. We refer to Mammen, Linton and Nielsen (1999) and Linton and Nielsen (1995) for detailed methods in this case. Here, however, we will focus on suitable parametric assumptions tailored to financial modeling.

One way is to use no-arbitrage arguments and collapse S_t, $r_{t,\tau}$ and $\delta_{t,\tau}$ into the forward price $F_t = S_t e^{(r_{t,\tau}-\delta_{t,\tau})\tau}$ in order to express the call pricing function as

$$C(S_t, X, \tau, r_{t,\tau}, \delta_{t,\tau}) = C(F_{t,\tau}, X, \tau, r_{t,\tau})$$

Alternatively use the non-arbitrage relation to estimate dividends and express the function in terms of the discounted stock price, that is either by $S_t^0 = S_t e^{-\delta_{t,\tau}} = S_t - D_{t,\tau}$ where $D_{t,\tau}$ is the present value of the dividends to be paid before the expiration. Thus it is

$$C(S_t, X, \tau, r_{t,\tau}, \delta_{t,\tau}) = C(S_t^0, X, \tau, r_{t,\tau}) \; .$$

A further reduction of the number of regressors is achieved by assuming that the call option function is homogeneous of degree one in S_t and X so that

$$C(S_t, X, \tau, r_{t,\tau}, \delta_{t,\tau}) = XC(S_t/X, \tau, r_{t,\tau}, \delta_{t,\tau}).$$

Combining the assumptions of the last two equations, the call price function can be further reduced to a function of three variables: moneyness $M_t = \frac{S_t^0}{K}$, maturity τ and risk free interest rate $r_{t,\tau}$. Notice that by smoothing with respect to moneyness, rather than with respect to the dividend adjusted index level we implicitly assume the theoretical option function is homogeneous of degree one with respect to the index and strike price. The basic Black-Scholes formula is an example of such a function, and as shown by Merton (1973) and discussed in Ingersoll (1987), a call price is homogeneous of degree one in the asset price and strike price if the asset's return distribution is independent of the level of the underlying index. We use these dimension reduction techniques in in the empirical study in both settings, direct estimation of the RND from the call prices and but also in indirect estimation via implied volatility.

2.4 Application

We use tick data on the DAX index based European options prices maturing in one month (21 trading days), provided by EUREX for 20040121. The transformed data according to a methodology by Fengler (2005) contain date stamp, implied volatility, type of the option, maturity, strike price, option price, interest rate, intraday future price, average dividend rate.

The index stock price varies within one day and one needs to identify the price at which a certain transaction has taken place. Intraday DAX index prices are available on EUREX. Several authors report that the change of the index price is stale and for every pair option/strike we use instead the prices of futures contracts closest to the time of the registered trade, see e.g. Jackwerth 2000.

Original strike prices are given on an equidistant grid and in order to account for movements in the intraday price we use the following transformation $\frac{X_i}{F_i} S_t e^{r_{t,\tau} - \delta_{t,\tau}}$, where X_i and F_i are paired observations and S_t is the median intraday stock price, $r_{t,\tau}$ is the one month interest rate (linearly interpolated EURIBOR rates, for the desired maturity) and $\delta_{t,\tau}$ the average dividend (see Fengler, 2005). Conditional on these values we estimate q and interpret it as an average curve for the estimation date.

We use only at-the-money and out-of-the-money call options and in-the-money puts translated in call prices by using the put call parity

$$C_t - P_t = S_t e^{-\delta_{t,\tau}\tau} - Xe^{-r_{t,\tau}\tau}$$

This guarantees that unreliable observations (high volatility) will be removed from our estimation samples. Since, as mentioned before, the intraday stock price varies, we use its median to compute the risk neutral density. For this price, we verify if our observations satisfy the arbitrage condition and delete for our sample those who do not satisfy it

$$S_t \geq C_i \geq \max(S_t - X_i e^{-r_{t,\tau}\tau}, 0).$$

Finally, if we have different call price observations observations for the same strike price we take their median at that point. In this way we ensure that we have a one to one relationship between every call and strike price.

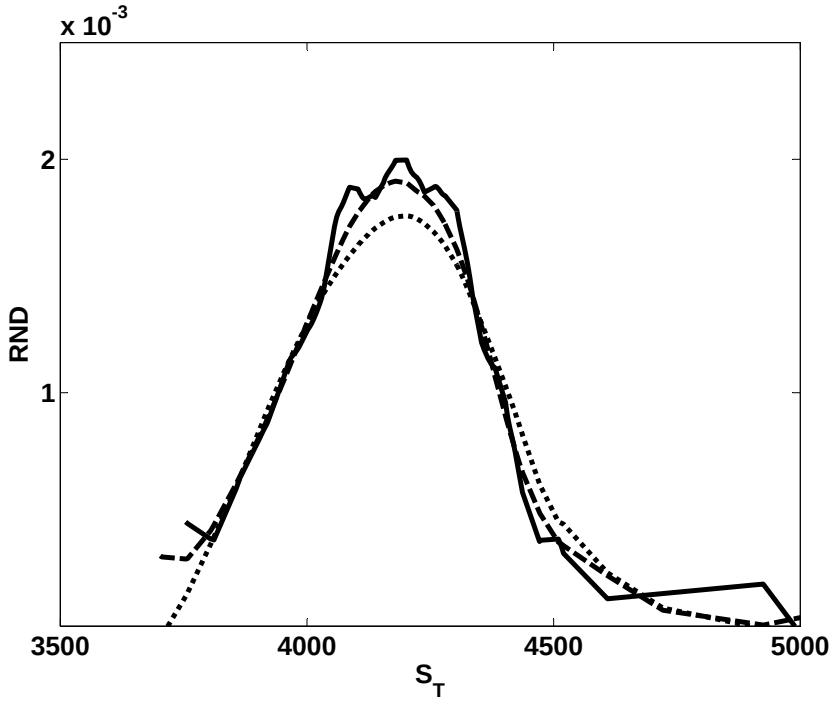

Fig. 2. $\hat{q}(S_T)$ by local polynomial smoother for the optimally chosen bandwidth $h = 114.34$ by cross-validation (solid line) and oversmoothing bandwidths $h = 227.59$ (dashed line) and $h = 434.49$ (dotted line)

Smoothing in call option space

As described in Section 2.1 local polynomial estimation allows to compute the second derivative of the call price directly, in a single step. We use local

polynomial smoothing of degree three and a quartic kernel to reduce finite sample bias. In the first step, we rescale the call price by dividing it by S_t and we smooth in this direction. We use cross-validation to choose the optimal bandwidth; however, this bandwidth appears to undersmooth in extreme areas around the mode and in the tails yielding a wiggly estimator in these regions (see in Figure 2). Therefore we decide to gradually increase the bandwidth to "stabilize" the estimate in the extremes. However, Figure 2 also illustrates that this should be done with care, as too much of oversmoothing can easily cause a huge bias at the mode.

Smoothing in Implied Volatility Space

In practice, the smoothing is mainly done in the implied volatility span because call prices respond asymmetrically to changes in the strike prices. In the present context, implied volatility is the volatility that yields a theoretical value for the option equal to the observed market price of that option, when using the Black-Scholes pricing model. We then estimate a smooth function $\hat{\sigma}$ and recover the call price by a bijective function evaluated at some fixed values of the regressors and variable σ

$$\begin{aligned}\hat{C}(S_t, X, \tau, r_{t,\tau}, \delta_{t,\tau}) &= C_{BS}(\cdot\,; \hat{\sigma}(S_t, X, \tau, r_{t,\tau}, \delta_{t,\tau})) \\ &= e^{-\delta_{t,\tau}\tau} S_t \Phi(y + \sigma\sqrt{\tau}) - e^{-r_{t,\tau}\tau} X \Phi(y)\end{aligned}$$

where Φ is the distribution function of the standard normal distribution and

$$y = \frac{\log(\frac{S_t}{K}) + (r_{t,\tau} - \frac{1}{2}\sigma^2)\tau}{\sigma\sqrt{\tau}}.$$

In this chapter we use a method based on Rookley (1997) who shows how to improve the efficiency of the estimator by estimating σ and its first two derivatives by local polynomial regression and plugging them into a modified version of the Black-Scholes formula. Below we describe the method for fixed maturity of one month.

For each pair $\{C_i, X_i\}_{i=1}^n$ we define the rescaled call option $c_i = C_i/S_t$ in terms of moneyness $M_i = S_t/X_i$ so that starting from the Black-Scholes formula for the call price we can write

$$\begin{aligned}C_i &= C\{M_i; \sigma(M_i)\} = \Phi(d_1) - \frac{e^{-r\tau}\Phi(d_2)}{M_i} \\ d_1 &= \frac{\log(M_i) + \left\{r_{t,\tau} + \frac{1}{2}\sigma(M_i)^2\right\}\tau}{\sigma(M_i)\sqrt{\tau}} \\ d_2 &= d_1 - \sigma(M_i)\sqrt{\tau}.\end{aligned}$$

For simplification we drop the indices. The risk neutral density can be expressed in terms of rescaled call price

$$q(\cdot) = e^{r\tau}\frac{\partial^2 C}{\partial X^2} = e^{r\tau}S\frac{\partial^2 c}{\partial X^2}$$

with

$$\frac{\partial^2 C}{\partial X^2} = \frac{\mathrm{d}^2 c}{\mathrm{d}M^2}\left(\frac{M}{X}\right)^2 + 2\frac{\mathrm{d}c}{\mathrm{d}M}\frac{M}{X^2}$$

and

$$\begin{aligned}\frac{\mathrm{d}^2 C}{\mathrm{d}M^2} = {} & \varphi(d_1)\left\{\frac{\mathrm{d}^2 d_1}{\mathrm{d}M^2} - d_1\left(\frac{\mathrm{d}d_1}{\mathrm{d}M}\right)^2\right\} \\ & - \frac{e^{-r\tau}\varphi(d_2)}{M}\left\{\frac{\mathrm{d}^2 d_2}{\mathrm{d}M^2} - \frac{2}{M}\frac{\mathrm{d}d_2}{\mathrm{d}M} - d_2\left(\frac{\mathrm{d}d_2}{\mathrm{d}M}\right)^2\right\} \\ & - \frac{2e^{-r\tau}\Phi(d_2)}{M^3}\end{aligned}$$

where φ is the probability density function of the standard normal distribution. The results depend further on the following quantities, where $\sigma(M)$, $\sigma'(M)$, $\sigma''(M)$ are smooth functions in moneyness direction

$$\begin{aligned}\frac{\mathrm{d}^2 d_1}{\mathrm{d}M^2} = {} & -\frac{1}{M\sigma(M)\sqrt{\tau}}\left\{\frac{1}{M} + \frac{\sigma'(M)}{\sigma(M)}\right\} \\ & + \sigma''(M)\left\{\frac{\sqrt{\tau}}{2} - \frac{\log(M) + r\tau}{\sigma(M)^2\sqrt{\tau}}\right\} \\ & + \sigma'(M)\left\{2\sigma'(M)\frac{\log(M) + r\tau}{\sigma(M)^3\sqrt{\tau}} - \frac{1}{M\sigma(M)^2\sqrt{\tau}}\right\}\end{aligned}$$

$$\begin{aligned}\frac{\mathrm{d}^2 d_2}{\mathrm{d}M^2} = {} & -\frac{1}{M\sigma(M)\sqrt{\tau}}\left\{\frac{1}{M} + \frac{\sigma'(M)}{\sigma(M)}\right\} \\ & - \sigma''(M)\left\{\frac{\sqrt{\tau}}{2} + \frac{\log(M) + r\tau}{\sigma(M)^2\sqrt{\tau}}\right\} \\ & + \sigma'(M)\left\{2\sigma'(M)\frac{\log(M) + r\tau}{\sigma(M)^3\sqrt{\tau}} - \frac{1}{M\sigma(M)^2\sqrt{\tau}}\right\}.\end{aligned}$$

In order to estimate $\sigma(M)$ and its associated first and second derivatives with respect to moneyness we use univariate local polynomial kernel regression of degree three and a quartic kernel. The optimal bandwidth has been

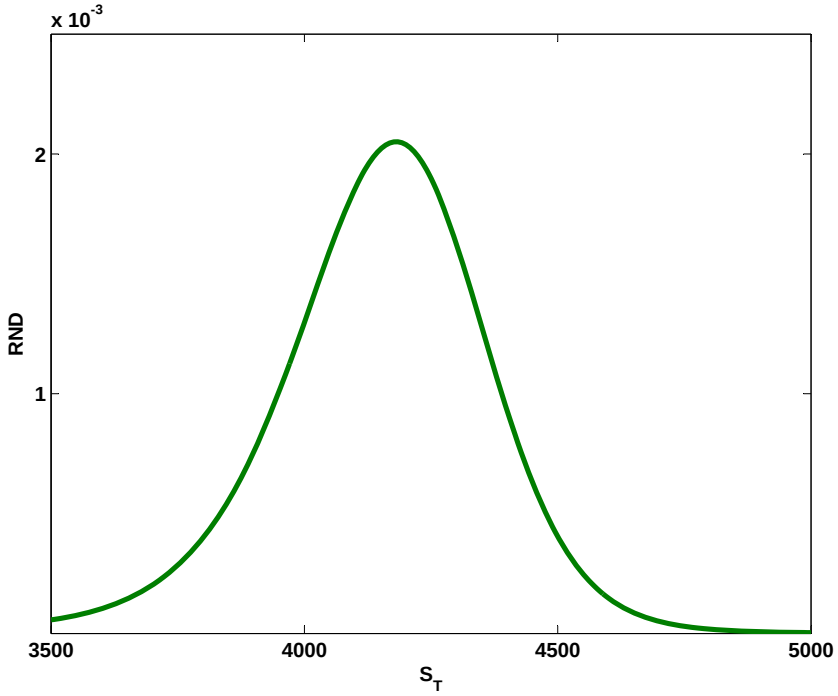

Fig. 3. $\hat{q}(S_T)$ by Rookley method with oversmoothing bandwidth $h = 372.42$

computed using cross-validation criteria (18) for the implied volatility. Oversmoothing bandwidths, see Figure 3, improve the fit in the tails because they allow for more observations to be included in the estimators while having little effects on the values of $\hat{q}$ situated in the middle of the distribution, where the estimates by different bandwidths overlap almost perfectly. It follows that smoothing in implied volatility yields a more stable estimator in terms of shape in finite sample, for varying bandwidths. This can be well seen in Figures 2 and 3. It is because the implied volatility responds with a fairly constant magnitude to the changes in the strike price over the estimation domain.

Problems and Refinements

In applications the support of strike prices is mostly compact and thus bounded. As shown in Section 2.1. the quality of estimates in regions close to the boundary might be low due to small values of the regressors' density when using even order polynomials. By using a polynomial of order 3, estimation is design adaptive for the second derivative avoiding this problem.

Furthermore, associated with the boundary, option data is characterized by scarce observations close to the bounds. In general, nonparametric techniques do not perform well in regions with sparse data and other methods are required. Parametrization of the tails using Pareto type distributions might be advantageous leaving however the question of how to join the two regions in order to assure that the resulting distribution integrates to one. Alternatively, Rookley (1997) proposes to further parametrize these distributions by matching them with an Edgeworth expansion type density

$$q(S_T) = \frac{1}{S_T\sigma}\Phi(Z)\{1 + \beta(Z^3 - 3Z) + \gamma(Z^4 - 6Z^2 + 3)\}$$

for $Z = \frac{\log(S_T)-\tilde{\mu}}{\tilde{\sigma}}$, where $\tilde{\mu}$ and $\tilde{\sigma}$ are the conditional mean and standard deviation of $\log(S_T)$ implied by the risk neutral measure, and β and γ are coefficients related to the higher moments of $\log(S_T)$.

In order for the risk neutral density to be well defined, an estimate of the call price function C must satisfy certain high-level conditions (see e.g. Aït-Sahalia and Duarte (2003): It should be (1) positive, (2) decreasing in X, (3) convex, and (4) its second derivative should exist, be nonnegative and integrable. Given that the first derivative of C with respect to X is the (negative) discounted cumulative density function of q conditions (2) and condition (3) can be summarized by the following inequality

$$-e^{r_{t,\tau}} \leq \frac{\partial C(S_t, X, \tau, r_{t,\tau}, \delta_{t,\tau})}{\partial X} \leq 0.$$

Convexity requires

$$\frac{\partial^2 C(S_t, X, \tau, r_{t,\tau}, \delta_{t,\tau})}{\partial^2 X} \geq 0.$$

Nonparametric kernel estimates may violate these constraints, unless we deal with large samples of observations. Imposing constraints like monotonicity or convexity directly in the estimation leads to nontrivial optimization problems in topological cones. If it is crucial for the outcome to fulfill the shape restrictions in small samples, it is recommended to use series type estimation methods which easily allow to incorporate them directly in the estimation. In general, these constrains must be applied directly to the call price, because theoretical properties of the implied volatility are not well known. For further references see Ait-Sahalia (2003). This will be illustrated in the next section.

3 Estimation of the RND via Empirical Pricing Kernel

In the previous section, we studied nonparametric kernel methods for estimating q as the discounted second derivative of the call price function and

discussed the problems associated with kernel type estimators in this setting. Now, we propose a new approach, based on series expansion of the pricing kernel.

In financial mathematics the relationship between the physical measure p and RND q of a financial asset can be represented via the pricing kernel m. Also called stochastic discount factor, the pricing kernel is the quotient of the Arrow security prices and the objective probability measure and summarizes information related to asset pricing. Thus it is

$$q(S_T) = m(S_T)p(S_T). \tag{19}$$

From a behavioral economics perspective m describes risk preferences of a representative agent in an exchange economy. In many applications, the empirical pricing kernel is the object of interest. In most of the studies Aït-Sahalia and Lo (2000), Brown and Jackwerth (2004), Grith et al. (2009) it has been estimated as a ratio of two estimated densities: $\hat{q}$ computed as the second derivative of a smooth call function (as described in Section 2) and $\hat{p}$ based on historical returns. This approach leads to difficulties in deriving the statistical properties of the estimator. In particular, the sample sizes for estimating p and q may differ substantially: p uses daily observations, whereas q is based on intraday high-frequency observations. On the other hand, methods for estimating p are in general much simpler and more stable compared to those for q for which typically nonparametric kernel estimation of a second derivative is required. Direct estimation of the pricing kernel can be seen as an improvement in this sense.

For estimating q, however, a series approach is additionally appealing, as high-level shape constraints are straightforward to incorporate in finite samples. Recall that for kernel type estimators this is not the case, see the end of Subsection 2.4.

We introduce the series expansion for the pricing kernel in (3). With an estimate of the physical measure from historical data and the pricing kernel m from option prices, these indirectly imply an estimate of q via (19). In statistical theory and also in practice, this indirect way of estimating q has a faster rate of convergence than using series methods directly for q in (1) which require the choice of an additional regularization parameter to guarantee invertibility of an ill-posed statistical problem. In particular, in (19) large values of S_T are downweighted by integrating over the physical measure, while they enter undamped in (1) leading to unreliable results.

3.1 Direct Estimation of Empirical Pricing Kernel via Series Methods

As for q, there are several factors which drive the form of the pricing kernel. Here, however, we focus on the projection of the pricing kernel on the set of available payoff functions m^*, which allows us to represent m in terms of S_T

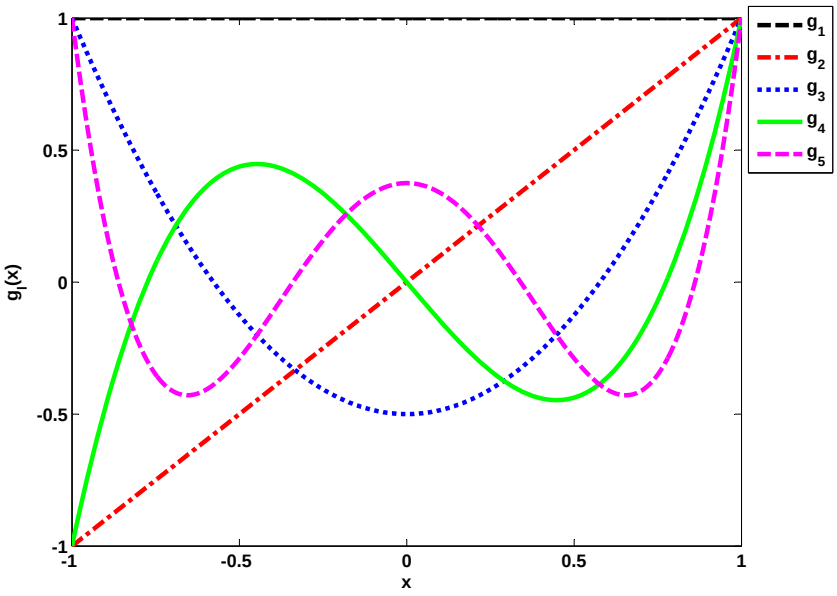

Fig. 4. First five terms of the Legendre polynomials

only. In practice this is a reasonable assumption. Thus we require that m and m^* are close in the following sense

$$||m - m^*||^2 = \int |m(x) - m^*(x)|^2 \mathrm{d}x < \epsilon \tag{20}$$

with ϵ small. Further we assume that m^* has a Fourier series expansion

$$m^*(S_T) = \sum_{l=1}^{\infty} \alpha_l g_l(S_T), \tag{21}$$

where $\{\alpha_l\}_{l=1}^{\infty}$ are Fourier coefficients and $\{g_l\}_{l=1}^{\infty}$ is a fixed collection of basis functions. The functions g_l are chosen as orthonormal with respect to a particular norm. Such a representation is possible if the function is absolutely integrable.

Based on (21), we can construct an estimator for m^* and thus m. If a finite number L of basis functions is sufficient for a good approximation of m then

$$\widehat{m}(S_T) = \sum_{l=1}^{L} \hat{\alpha}_l g_l(S_T). \tag{22}$$

Estimates $\hat{\alpha}_l$ for the coefficients α_l could be obtained by least squares for fixed basis functions g_l if a direct response was observable. Clearly the choice of L

controls the quality of the estimate. The larger L, the better the fit but the higher the computing cost and less robust the result. See Subsection 3.3 for a sophisticated way of selecting the smoothing parameter.

In financial applications the following polynomial basis functions are frequently used: e.g. Laguerre, Legendre, Chebyshev polynomials, see Figure 4 and Subsection 3.5. While asymptotically equivalent, in finite samples their form will influence the size of L. In general, one would prefer to have g_l such that L small is sufficient. For a formal criterion on how to select between different basis options see Li and Racine (2007). They assess different candidate basis functions by comparing a CV-type criterion for fixed L.

Though the form of m is only indirectly determined by relating observable call prices Y_i to strike prices X_i for given T, τ via (3). A response to observed payoffs via the pricing kernel is not directly observable. In sample an estimate of m should fulfill

$$\begin{aligned} Y_i &= e^{-r_{t,\tau}\tau} \int_0^\infty (S_T - X_i)^+ \sum_{l=1}^{L} \hat{\alpha}_l g_l(S_T) p_t(S_T) dS_T + \varepsilon_i \qquad (23) \\ &= \sum_{l=1}^{L} \hat{\alpha}_l \left\{ e^{-r_{t,\tau}\tau} \int_0^\infty (S_T - X_i)^+ g_l(S_T) p_t(S_T) dS_T \right\} + \varepsilon_i \end{aligned}$$

with error ε such that $\mathsf{E}[\varepsilon|X] = 0$. Set

$$\psi_{il} = \psi_l(X_i) = e^{-r_{t,\tau}\tau} \int_0^\infty (S_T - X_i)^+ g_l(S_T) p_t(S_T) dS_T. \qquad (24)$$

Then for known p and fixed basis functions and fixed L, the vector $\hat{\alpha} = (\hat{\alpha}_1, ..., \hat{\alpha}_L)^\top$ is obtained as

$$\arg\min_{\alpha} \sum_{i=1}^{n} \left\{ Y_i - \sum_{l=1}^{L} \alpha_l \psi_l(X_i) \right\}^2 \qquad (25)$$

In practice, however, p is not known and can only be estimated. Therefore instead of ψ_l in (24) we have only estimates $\hat{\psi}_l$ of the basis functions. We consider two possible ways for constructing them. First, regard ψ as an expectation which can be estimated by sample averaging over J different payoffs at time T for fixed τ and given X

$$\hat{\psi}_{il} = e^{-r_{t,\tau}\tau} J^{-1} \sum_{s=1}^{J} (S_T^k - X_i)^+ g_l(S_T^k) \quad \text{with .} \qquad (26)$$

How $(S_T^k)_{k=1}^J$ are obtained is explained in detail in the following subsection. Alternatively, replace p by an estimator, e.g. a kernel density estimator. Then it is

$$\hat{\psi}_{il} = e^{-r_{t,\tau}\tau} \int_0^\infty (S_T - X_i)^+ g_l(S_T)\hat{p}(S_T)dS_T. \tag{27}$$

Here some care is needed in numerical integration to keep discretization errors negligible. Furthermore, for an appropriate choice of bandwidth in $\hat{p}$, both approaches are asymptotically equivalent. In finite samples, however, estimates for q might differ (see Figure 5, for $J = 4500$, and S_T^k simulated based on historical log-returns).

In total we obtain a feasible estimator of α based on a feasible version of (25) as

$$\check{\alpha} = (\hat{\Psi}^\top \hat{\Psi})^- \hat{\Psi}^\top Y. \tag{28}$$

The elements of $\hat{\Psi}_{(n\times L)}$ are given either by (26) or (27) and $Y = (Y_1, \cdots, Y_n)^\top$.

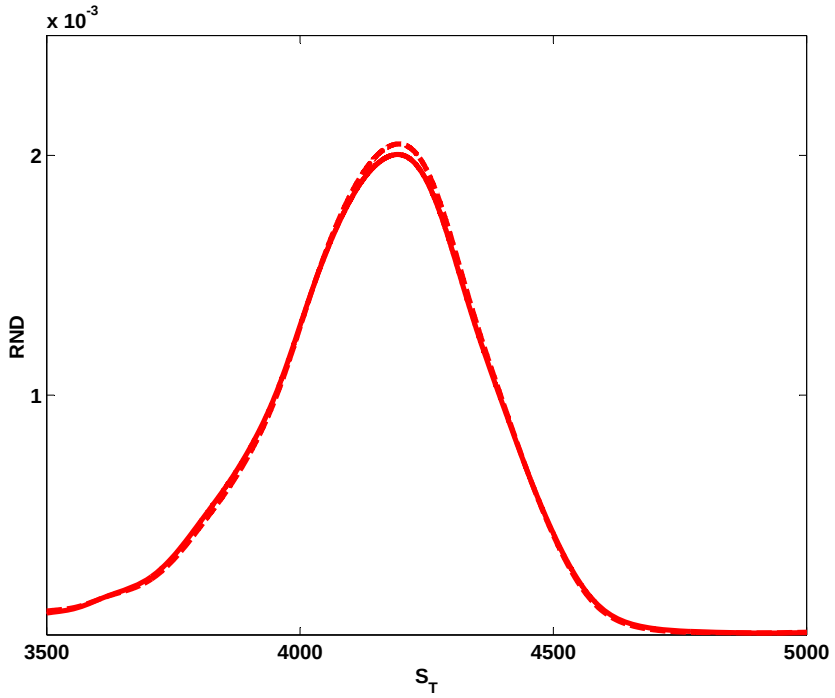

Fig. 5. $\hat{q}(S_T)$ in Legendre basis with $L = 5$ based on approximation (26) (simple line) and (27) of Ψ (dashed line)

Then an estimate of the pricing kernel at observation s of S_T is given by

$$\hat{m}(s) = g^L(s)^\top \check{\alpha}, \tag{29}$$

where $g^L(s) = (g_1(s), \ldots, g_L(s))^\top$. We see in figure (6) that the estimator of m is less stable for different approximations of ψ_{il}. Finally, the risk neutral density is estimated as

$$\hat{q}(s) = \hat{m}(s)\hat{p}(s). \tag{30}$$

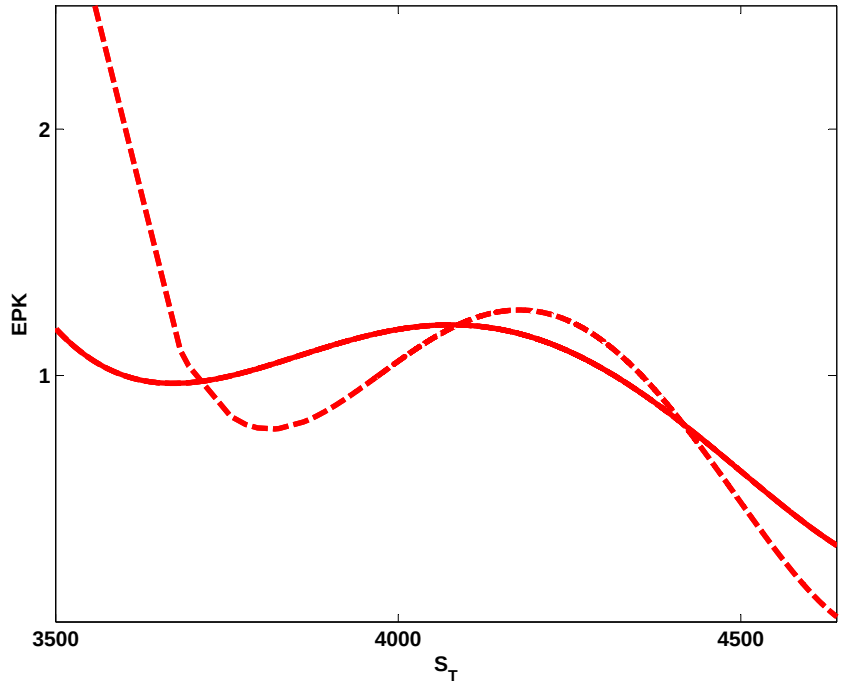

Fig. 6. $\hat{m}^*(S_T)$ by Legendre basis expansion with $L = 5$ based on approximation (26) (solid line) and (27) of Ψ (dashed line)

3.2 Estimation of the PDF of S_T

In the empirical study we use two different ways of obtaining $\hat{p}$ from the DAX Index prices at time T. And we look at the sensitivity of $\hat{q}$ w.r.t. $\hat{p}$. First, we extrapolate possible realizations of S_T in the future from historical log-returns. Based on a window of historical DAX Index values of length J we get

$$S_T^k = S_t e^{r_T^k}, \quad \text{for} \quad r_T^k = \log(S_{t-k}/S_{t-(k+1)}). \tag{31}$$

Alternatively, we use a GARCH(1,1) specification for the log-returns to account for slowly decaying autocorrelation in the data. The model is specified as follows

$$\log(S_t/S_{t-1}) = \mu + u_t\,, \qquad u_t \sim f(0, \sigma_t^r). \tag{32}$$

In equation (32), the returns consist of a simple constant, plus an uncorrelated, non-Gaussian disturbance. The conditional variance $(\sigma_t^r)^2$ follows an ARMA(1,1) type specification

$$(\sigma_t^r)^2 = a_1 + a_2 r_{t-1}^2 + a_3(\sigma_{t-1}^r)^2\,. \tag{33}$$

We can estimate the parameters of the model (μ, a_1, a_2, a_3) and retrieve a time series of stochastic volatilities $\{\sigma_{t-k}^r\}_{k=1}^J$. The simulated index prices at time T are obtained as in (31) above for

$$r_T^k = r_{t-k}\frac{\sigma_T^r}{\sigma_{t-k}^r}$$

where we use for the forecasted volatility σ_T^r today's volatility σ_t based on GARCH.

Then the probability density p of S_T is estimated at each point S_T using a kernel density estimator

$$\widehat{p}_h(S_T) = \frac{1}{Jh}\sum_{k=1}^{J} K\left(\frac{S_T^k - S_T}{h}\right) \tag{34}$$

where K is a kernel function and the bandwidth is selected similarly to the criteria introduced in Subsection 2.2. Resulting estimates of the two approaches are illustrated in Figure 7 for $J = 5000$. We observe that they differ significantly in spread and the mode. However, the differences depend on the window length J of returns used to estimate the parameters of the GARCH model, as well as on the choice of the bandwidth used to estimate p, which carries over to q via 30 directly or indirectly.

3.3 Choice of the Tuning Parameters

The quality of the obtained series estimators (29) and (30) depends on a suitable choice of the number $L(n) \to \infty$ for $n \to \infty$ for given basis functions. Note that the role of L (or L/n) is similar to that played by the smoothing parameter h for the kernel methods. There are three well-known procedures for a data-driven optimal selection of L. The first one is Mallows's C_L as proposed in Mallows (1973): Select L_M such that it minimizes

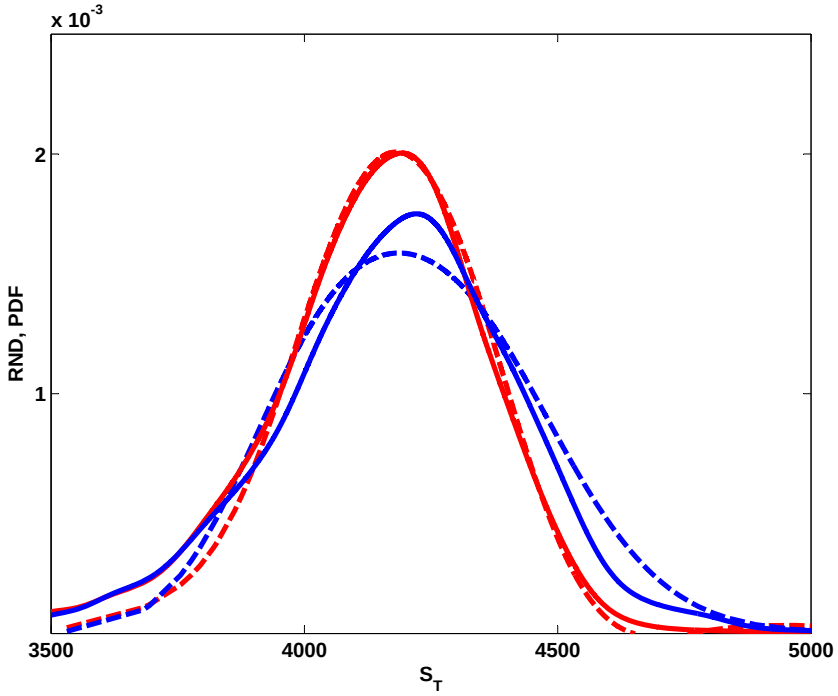

Fig. 7. $\hat{q}$ in Legendre basis with $L = 5$ and $\hat{p}$ based on log-returns (blue) and weighted log-returns (red). Solid and dashed lines correspond to the specifications in the Figure 5

.

$$C_L = n^{-1} \sum_{i=1}^{n} \left\{ Y_i - \sum_{l=1}^{L} \check{\alpha}_l \hat{\psi}_l(X_i) \right\}^2 + 2\sigma^2(L/n)$$

where σ^2 is the variance of ε. One can estimate σ^2 by

$$\hat{\sigma}^2 = n^{-1} \sum_{i=1}^{n} \hat{\varepsilon}_i^2$$

with $\hat{\varepsilon}_i = Y_i - \sum_l \check{\alpha}_l \hat{\psi}_l(X_i)$.

A second way for selecting L is according to generalized cross-validation suggested by Craven and Wahba (1979). Choose L_{GCV} minimizing

$$CV_L^G = \frac{n^{-1} \sum_{i=1}^{n} \left\{ Y_i - \sum_{l=1}^{L} \check{\alpha}_l \hat{\psi}_l(X_i) \right\}^2}{\{1 - (L/n)\}^2}.$$

The last criterion is leave-one-out cross-validation according to Stone (1974): Select L_{CV} minimizing

$$CV_L = \sum_{i=1}^{n} \left\{ Y_i - \sum_{l=1}^{L} \check{\alpha}_l^{-i} \hat{\psi}_l(X_i) \right\}^2$$

where $\check{\alpha}_l^{-i}$ is the leave one estimate of α_l obtained by removing (X_i, Y_i) from the sample.

Li (1987) showed that each of the above three criteria leads to an optimally selected L in the sense that they all minimize the asymptotic weighted integrated squared error (see (16)). In this sense the obtained L are asymptotically equivalent.

3.4 Statistical Properties

Series type estimators are designed to provide good approximations in an L_2 sense, see (20). Therefore asymptotic properties as consistency and rates of convergence should be derived from the asymptotic mean squared error. The rate of convergence for the indirect estimator of q via the pricing kernel depends on the two smoothing parameters h and L.

$$\begin{aligned}
\int_0^\infty \{\hat{q}(S_T) - q(S_T)\}^2 dS_T &= \int_0^\infty \{\hat{m}(S_T)\hat{p}(S_T) - m(S_T)p(S_T)\}^2 dS_T \\
&= \int_0^\infty [\hat{m}(S_T)\{\hat{p}(S_T) - p(S_T)\}]^2 dS_T \\
&\quad + \int_0^\infty [p(S_T)\{\hat{m}(S_T) - m(S_T)\}]^2 dS_T \\
&\quad + \int_0^\infty 2\hat{m}(S_T)\{\hat{p}(S_T) - p(S_T)\}p(S_T)\{\hat{m}(S_T) - m(S_T)\} dS_T
\end{aligned}$$

It easily follows from the law of iterated expectations that the third term equals zero. Consequently, the convergence of $\hat{q}(S_T)$ depends only on the first two terms. Since $\sup \hat{m}(s) = \mathcal{O}_P(1)$ under Assumption 1 given below, the order of convergence for the first term is dominated by $\{\hat{p}(S_T) - p(S_T)\}^2$.

Assumption 1 Suppose that p is twice continuously differentiable, K is a second order kernel and the bandwidth is chosen optimally as $h = cn^{-1/5}$, for a known constant c.

Then the asymptotic mean squared error for the kernel density estimator is

$$||\widehat{p}_h(x) - p(x)||_2^2 = \mathcal{O}_P(n^{-4/5}) \tag{35}$$

This follows along the same logic as the results for local polynomials in Section 2.1. For further details see e.g. Härdle et al. (2004).

The order of convergence for the second term only depends on $\{\hat{m}(S_T) - m(S_T)\}^2$ since $\sup p(s) \leq 1$. The next assumption establishes consistency of $\hat{m}(S_T)$.

Assumption 2 $\{X_i, Y_i\}$ are i.i.d. observations of (X, Y), $\mathsf{Var}(Y|X)$ is bounded on S, the compact connected interval of support of X. Furthermore p is bounded away from zero and m is ν-times continuously differentiable on S. Choose L such that $L^3/n \to 0$ as $n \to \infty$.

Under Assumption 2 it is

$$\int_0^\infty \{\hat{m}(S_T) - m(S_T)\}^2 dS_T = \mathcal{O}_p(L/n + L^{-2\nu}). \tag{36}$$

This result is from Newey (1997) for fixed basis functions ψ_l. With estimated basis $\hat{\psi}_l$ the result still goes through as the convergence of $\hat{\psi}_l$ to the true ψ_l is at parametric rate. The i.i.d. assumption is for simplicity of the exposition only. It can be easily relaxed to mixing type of observations.

The theorem below puts (35) and (36) together for an asymptotic result for q.

Theorem 3. *Assume that Assumptions 1 and 2 hold. Then the integrated square error (*ISE*) converges as*

$$\int_0^\infty \{\hat{q}(S_T) - q(S_T)\}^2 dS_T = \mathcal{O}_p(n^{-4/5} + L/n + L^{-2\nu}) \,. \tag{37}$$

3.5 Implementation

We illustrate the method using the data described in Subsection 2.4. We consider the univariate regression of C on the strike price X for fixed maturity and fixed interest rate. We estimate q using three different systems of orthogonal basis: Laguerre, Legendre and Chebyshev. We found that the fit of the call price is almost identical for fixed L, while $\hat{\alpha}$ varies obviously with the series. There is little sensitivity with respect to the choice of the basis functions that holds also for the empirical pricing kernel and the implied risk neutral density. Based on the selection criteria for L from Section 3.3, we have chosen $L = 5$. We exemplify the method with Legendre polynomials. Estimation results are displayed in Figure (5) to Figure (8).

4 Conclusions

We have studied three nonparametric approaches for estimating the risk neutral density. They are based on fundamentally different techniques: two of

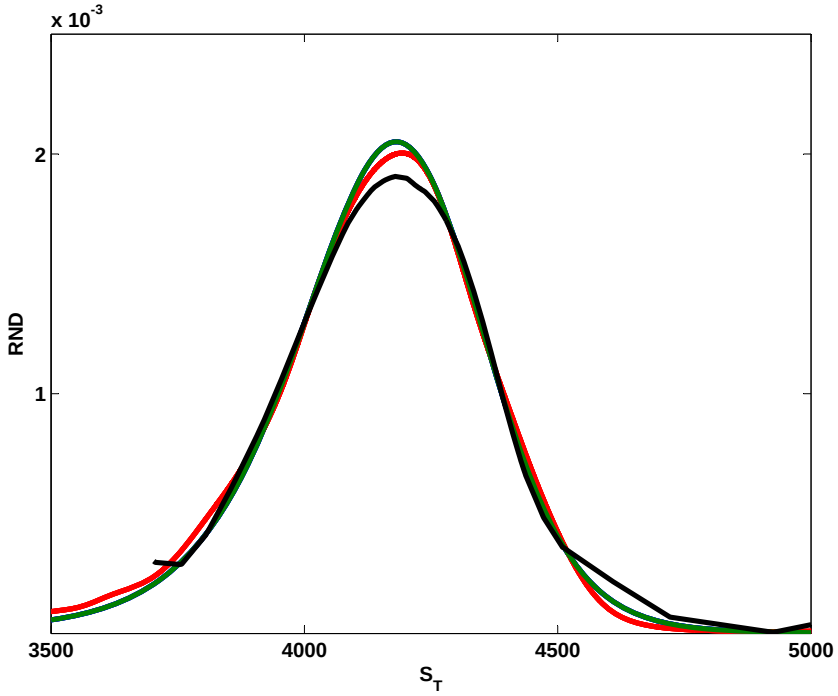

Fig. 8. $\hat{q}(S_T)$ by local polynomial regression with $h = 227.59$ in call space (black), by Rookley method $h = 372.42$ in IV space (green), indirect estimation of the pricing kernel as Legendre basis expansion with $L = 5$ (green)

them use local features and the third one is based on global curve fitting. For these approaches we have described the estimation methodology and their performance in finite sample, in terms of robustness and stability. Statistical properties of all procedures have been derived and illustrated focusing on practically most relevant aspects.

Figure 8 shows estimates of q using the three methods we discussed in this article for suitable choices of tuning parameters. While for the given sample size, all three nonparametric methods yield similar results, there still are some peculiarities. Our empirical results suggest that kernel methods for the estimation of q in implied volatility space work much better than those which smooth in the call price space in our sample. Local polynomial methods applied to call prices yield estimates which are highly sensitive to the choice of the bandwidth: A globally optimal bandwidth might in fact severely undersmooth around the mode or in the tails, resulting in wiggly estimates in this areas. In comparison to this, when we smooth in the implied volatility space, the Rookley method yields a much smoother estimate directly without

additional oversmoothing and performs better in regions of sparse data, see Figures 2 and 3. Estimation of the risk neutral density based on the pricing kernel is not affected by the choice of the basis functions in small samples - differences occur only in the tails due to scarcity of observations at the boundaries in our empirical findings. Generally, series type methods allow for direct incorporation of shape constraints. Thus resulting estimates are consistent with economic theory even in finite samples.

References

1. Aït-Sahalia, Y. and Duarte, J. (2003) Nonparametric option pricing under shape restrictions. Journal of Econometrics 116: 947
2. Aït-Sahalia, Y. and Lo, A. W. (1998) Nonparametric estimation of state-price densities implicit in financial asset prices. Journal of Finance 53: 499-547
3. Aït-Sahalia, Y. and Lo, A. W. (2000) Nonparametric risk management and implied risk aversion. Journal of Econometrics 94: 9-51
4. Arrow, K. J. (1964) The role of securities in the optimal allocation of risk-bearing. Review of Economic Studies 31: 91-96
5. Bates, D. S. (1996) Jumps and stochastic volatility: Exchange rate processes implicit in deutsche mark options. Review of Financial Studies 9(1): 69-107
6. Black, F. and Scholes, M. (1973) The Pricing of Options and Corporate Liabilities. Journal of Political Economy 81: 637-654.
7. Breeden, D. T. and Litzenberger, R. H. (1978) Prices of state-contingent claims implicit in option prices. The Journal of Business 51(4): 621-651
8. Brown, D.P. and Jackwerth, J. C. (2004) The pricing kernel puzzle: Reconciling index option data and economic theory. Manuscript
9. Campbell, J., Lo, A. and McKinlay, A. (1997) The econometrics of financial markets. Princeton University Press, Princeton, New Jersey
10. Debreu, G. (1959) Theory of value: An axiomatic analysis of economic equilibrium. Yale University Press, New Haven
11. Engle, R. F. and Rosenberg, J. V. (2002) Empirical pricing kernels. Journal of Financial Economics 64: 341-372
12. Fan, J., and Gijbels, I. (1996) Local polynomial modelling and its applications. Chapman and Hall, London
13. Fengler, M. R. (2005) Semiparametric modeling of implied volatility, Springer Finance
14. Grith, M., Härdle, W. and Park, J. (2009) Shape invariant modelling pricing kernels and risk aversion. SFB649DP2009-041. Submitted to Journal of Financial Econometrics on 1 June 2009
15. Härdle, W. (1990) Applied nonparametric regression. Econometric Society Monographs No. 19. Cambridge University Press
16. Härdle, W. and Hlavka, Z. (2009) Dynamics of state price densities. Journal of Econometrics 150: 1-15
17. Härdle, W., Müller, M., Sperlich, S. and Werwatz, A. (2004) Nonparametric and Semiparametric Models. Springer Verlag, Heidelberg
18. Härdle, W., Okhrin, Y. and Wang, W. (2009) Uniform confidence for pricing kernels. SFB649DP2010-003. Submitted to Econometric Theory on 1 January 2010
19. Heston, S. (1993) A closed-form solution for options with stochastic volatility with applications to bond and currency options. Review of Financial Studies 6: 327-343
20. Jackwerth, J. C. (1999) Option-implied risk-neutral distributions and implied binomial trees: a literature review. Journal of Derivatives 2: 66-82
21. Jackwerth, J. C.(2000) Recovering risk aversion from option prices and realized returns. Review of Financial Studies 13: 433-451
22. Li, K. C. (1987) Asymptotic optimality for c_p, c_l, cross-validation and generalized cross-validation: Discrete index set. Annals of Statistics 15: 958-975

23. Li, Q. and J. S. Racine (2007) Nonparametric econometrics: Theory and practice. Princeton University Press
24. Linton, O. and Nielsen, J. P. (1995). A kernel method of estimating structured nonparametric regression based on marginal integration. Biometrika 82: 93–100.
25. Longstaff, F. (1995) Option Pricing and the Martingale Restriciton. Review of Financial Studies 8: 1091-1124.
26. Lucas, R.E. (1978) Asset prices in an exchange economy. Econometrica 46: 1429-1445
27. Mallows, C.L. (1973) Some comments on c_p. Technometrics 15: 661-675
28. Mammen, E., Linton, O. and Nielsen, J. (1999). The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. Annals of Statistics 27 (5): 1443–1490.
29. Marron, J. S. and Nolan, D. (1988) Canonical kernels for density estimation. Statistics and Probability Letters 7(3): 195-199
30. Merton, R. C. (1973) Theory of rational option pricing. Bell Journal of Economics, The RAND Corporation 4(1): 141-183, Spring
31. Müller, H. G. (1988) Nonparametric regression analysis of longitudinal data. Lecture Notes in Statistics Vol. 46, New York, Springer-Verlag
32. W.K. Newey (1997) Convergence rates and asymptotic normality for series estimators. Journal of Econometrics 79: 147-168.
33. Renault, E., (1997) Econometric models of option pricing errors. In: Kreps, D.M. and Wallis, K.F. (eds) Advances in Economics and Econometrics: Theory and Applications vol. 3: 223-278. Cambridge University Press, Cambridge
34. Rookley, C. (1997) Fully exploiting the information content of intra day option quotes: Applications in option pricing and risk management, Working paper, University of Arizona
35. Rubinstein (1976) The valuation of uncertain income streams and the pricing of options. Bell Journal of Economics 7(2): 407-425
36. Ruppert, D. and Wand, M. P. (1994) Multivariate locally weighted least squares regression. Annals of Statistics 22(3): 1346-1370
37. Wahba, G. (1985) A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline smoothing problem. Annals of Statistics 4: 1378-1402