Nonparametric and Semiparametric Modeling and Estimation of Risk-Neutral Densities

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

[1] Ladislaus von Bortkiewicz Chair of Statistics, School of Business and Economics, Humboldt-Universität zu Berlin, Spandauer Straße 1, 10178 Berlin, Germany gritmari@wiwi.hu-berlin.de,
[2] Ladislaus von Bortkiewicz Chair of Statistics and CASE - Center for Applied Statistics and Economics, School of Business and Economics, Humboldt-Universität zu Berlin, Spandauer Straße 1, 10178 Berlin, Germany stat@wiwi.hu-berlin.de
[3] Chair of Econometrics, Institute for Statistics and Econometrics, Humboldt-Universität zu Berlin, Spandauer Straße 1, 10178 Berlin, Germany 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 neither on the underlying asset price dynamics nor on the distributional form of the risk neutral density. The first estimator is obtained as 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 performance of the presented methods is 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 are characterized by the state-price density, which yields one dollar if the final state is in the

interval $[x, \mathrm{d}x]$ when starting from any point x. While one way to justify existence and form of a state-price density is from preference-based equilibrium models (Lucas (1978)), we focus here on the reasoning from arbitrage-based models (Merton (1973)). In these models the state-price density is called risk neutral density (RND) under the assumption that the underlying market is dynamically complete, which we will adopt in the following. The RND also uniquely characterizes the 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 or the statistical family of distributions that the risk neutral density is assumed to belong to. Recent empirical findings, however, question the validity of these popular parametric specifications which drive the overall result (See e.g. Campbell, Lo, McKinlay (1997)). Nonparametric estimation therefore pose an important alternative by avoiding such 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 of intraday traded option prices have raised the feasibility of nonparametric methods. On the other side, due to their flexibility, however, many existing nonparametric risk neutral density estimation techniques are afflicted by irregularities such as data sparsity in the tails, negative probabilities, integrability to one. We will address these problems by appropriate choices of bandwidths and kernels, suggesting semiparametric techniques or imposing relevant constraints.

The rest of the chapter is organized as follows: section 2 describes kernel based regression methods for direct estimation of RND from the call prices function, section 3 introduces the concept of pricing kernel 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, comparing their performance.

2 Estimation of RND based on the second derivative

The price of a European call is obtained by discounting the conditionally expected payoff, where the expectation is taken with respect to the risk neutral measure:

$$C(K, \tau, r_{t,\tau}, \delta_{t,\tau}, S_t) = e^{-r_{t,\tau}\tau} \int_0^\infty \max(S_T - K, 0) q(S_T | \tau, r_{t,\tau}, \delta_{t,\tau}, S_t)\, dS_T \quad (1)$$

where S_t is the underlying asset price at time t, K the strike price, τ the time to maturity, $T = t + \tau$ the expiration date, $r_{t,\tau}$ the deterministic risk free interest rate for that maturity T, $\delta_{t,\tau}$ corresponding dividend yield of the

asset, $q(S_T|\cdot)$ is the conditional risk neutral density. We assume that these state variables contain all essential information contained for the estimation of C and q and quantities such as stochastic market volatility, trading volumes, ask-bid spreads are negligible. Then (**??**) poses the basis for estimation of RND.

For ease of notation, from now on we write $q(S_T)$ instead of $q(S_T|\tau, r_{t,\tau}, \delta_{t,\tau}, S_t)$. For fixed τ, assuming $r_{t,\tau} = r$ $\delta_{t,\tau} = \delta$, the risk neutral density can be derived from (**??**) as

$$q(K) = e^{r\tau}\frac{\partial^2 C}{\partial K^2}$$

The relation is due to Breeden and Litzenberger (1978) and serves as the basis of many current semi-parametric and nonparametric approaches to the estimation of q. We focus our attention on two possible estimation strategies: first, estimate a continuous twice-differentiable call function in all its arguments from traded options by smoothing in the call price or secondly, by smoothing in the implied volatility space. The second method is called "semi-parametric" because it uses a parametric function to translate call prices into implied volatility and vice versa. 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 use kernel based regression methods in each of the two cases because they are locally flexible a nd yield point estimates as opposed to series expansion and or sieve methods.

Assuming a set of paired observation C_i and $\mathbf{Z}_i = [S_{ti}, K_i, \tau_i r_{t_i,\tau_i}, \delta_{t_i,\tau_i}]$ for $i = 1, ..., n$, we wish to estimate the relationship between the two variables C_i and $\mathbf{Z}_i$ that satisfy the following general, possibly nonlinear relationship:

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

where $C(\bullet) : \Re^5 \to \Re$ is a smooth function in all directions and ε_i is i.i.d. with $\mathrm{E}[\varepsilon_i|\mathbf{Z}_i] = 0$.

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

$$\hat{C}(\mathbf{z}) = \sum_{i=1}^{n} w_i(\mathbf{z}) C_i$$

where the weights $w_i(\mathbf{z})$ assigned to each point of fit $\mathbf{z}$ decline as the $\mathbf{Z}_i$-s get further away from the estimation point. A kernel regression method uses kernel functions to construct weights. A univariate kernel is a non-negative real-valued integrable function $k(u) : \Re \to \Re$ which integrates to one and is symmetrical.

Table 1. Kernel functions $k(u)$

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

We can think of k as a probability distribution with potentially compact support. Possible kernel functions are presented in Table **??**.

Set $k_h(u) = \frac{1}{h}k\left(\frac{u}{h}\right)$ for all $u \in \Re$ with h the bandwidth, a smoothing parameter. In our five-dimensional space, for each pair $\mathbf{z}$ and $\mathbf{Z}_i$ the multivariate kernel function $\mathcal{K}(z - \mathbf{Z}_i) : \Re^5 \to \Re$ can be rewritten as

$$\mathcal{K}_h(\mathbf{z} - \mathbf{Z}_i) = \frac{1}{h^5}\mathcal{K}\left(\frac{\mathbf{z} - \mathbf{Z}_i}{h}\right)$$

where $\mathbf{u} = (\mathbf{z} - \mathbf{Z}_i)/h$ is now a rescaled vector and we can write the multidimensional kernel $\mathcal{K}_h(\mathbf{z} - \mathbf{Z}_i)$ as a product of univariate kernels:

$$k_S\left(\frac{S_t - S_{ti}}{h}\right)k_K\left(\frac{K - K_i}{h}\right)k_\tau\left(\frac{\tau - \tau_i}{h}\right)k_r\left(\frac{r_{t,\tau} - r_{t_i,\tau_i}}{h}\right)k_\delta\left(\frac{\delta_{t,\tau} - \delta_{t_i,\tau_i}}{h}\right)$$

In practice, for stationary or mixing data, it is not overly restrictive to use this particular form of multidimensional kernel with the same bandwidth h in all directions. Details on how to choose h are subject to a the next section.

For the moment we will return to the choice of the weight functions. The simplest case is to use the Nadaraya-Watson estimator but it is well known that this estimator has boundary problems. The choice of Nadaraya-Watson type smoothers can be improved upon by making use of local polynomial kernel smoothing. We will present this method in detail here. The basic idea of local polynomial regression rests upon the use of locally weighted least squares regression (Fan and Gijbels (1996)):

$$\min \sum_{i=1}^{n} \{C_i - C(\mathbf{Z}_i)\}^2 \mathcal{K}_{\mathbf{h}}(\mathbf{z} - \mathbf{Z}_i) \tag{2}$$

Since C is a continuous and differentiable function we can approximate it for a point $\mathbf{Z}_i$ near $\mathbf{z}$ by means of Taylor's expansion where we abstract for the mixed effects and consider them insignificant for the final results:

$$C(\mathbf{Z}_i) \approx C(\mathbf{z}) + \sum_{j=1}^{p} \frac{C^{(j)}(\mathbf{z})}{j!} T_i^j(\mathbf{z}) \equiv \beta_0(\mathbf{z}) + \sum_{j=1}^{p} \beta_j(\mathbf{z}) T_i^j(\mathbf{z})$$

$T_i^j(\mathbf{z}) = \left[(S_t - S_{ti})^j, (K - K_i)^j, (\tau - \tau_i)^j, (r_{t,\tau} - r_{t_i,\tau_i})^j, (\delta_{t,\tau} - \delta_{t_i,\tau_i})^j\right]$ with $j = 1, ..., p$. Here p is the degree of the polynomial. Usually, if we are interested in the ν-th derivative the degree of the polynomial should be $p = \nu + 1$. Efficient estimators are obtained for $p + 1$ odd. Substituting this expression in (**??**) we get:

$$\min_{\beta_0,\beta} \sum_{i=1}^{n} \left\{ C_i - \beta_0(\mathbf{z}) - \sum_{j=1}^{p} \beta_j(\mathbf{z}) T_i^j(\mathbf{z}) \right\}^2 \mathcal{K}_{\mathbf{h}}(\mathbf{z} - \mathbf{Z}_i)$$

$\hat{\beta}_0(\mathbf{z})$ is the estimator of C at point $\mathbf{z}$, while $j!\hat{\beta}_j(\mathbf{z}) = j!(\beta_{j1}(\mathbf{z})...\beta_{j5}(\mathbf{z}))$, is five-dimensional vector of partial j-th derivatives of C w.r.t. each of five variables evaluated at point $\mathbf{z}$. For the estimation of RND we are interested in the second derivative of the call price with respect to K. In our notation this is $2\beta_{22}$:

$$\hat{q}(S_T|S_t, K, \tau, r_{t,\tau}, \delta_{t,\tau})) = 2\beta_{22} = e^{r_{r_{t,\tau}\tau}} \left\{ \hat{C}''(K|S_t, \tau, r_{t,\tau}, \delta_{t,\tau}) \right\}_{K=S_T}$$

2.1 Selection of smoothing parameter

Kernel density estimation requires two parameters: the kernel function K and the bandwidth parameter h. The choice of univariate kernel has in practice little influence on the final result because kernel functions can be rescaled such that the difference between two kernel density estimates using two different kernels is almost negligible (Marron and Nolan (1988)). The determinant factor is the choice of the bandwidths. Traditional methods for choosing the smoothing parameter are based on finding a tradeoff between bias and variance. If a bandwidth is chosen to be too low, the estimated function tends have low bias but high variance. If the function is over-smoothed, variance can be reduced, but only at the cost of inducing higher bias.

There are two basic approaches to addressing the smoothing question: theoretical and data driven methods. Theoretical bandwidth in multidimensional case $h_{opt} \approx n^{\frac{-1}{d+4}}$, where d stands for dimension, are based on minimizing MISE or its asymptotic counterpart AMISE but they are not feasible in practice because they rely on the computation of some unknown quantities such as the second derivative of the population support. The first practical approach -eyeballing - relies on interactive graphical techniques where we start with a low bandwidth and gradually increases the bandwidth until the function is sufficiently smooth to the appearance of the eye. The approach itself is inherently subjective but often valid depending upon the end purpose of the research in question. The other general approach is to rely upon automatic bandwidth selectors. These methods attempt to minimize a global criteria such as asymptotic mean integrated squared error (AMISE) or a cross-validation function. Generally speaking, plug-in methods use the first criteria and derive

their name from the underlying principle: if you have an expression involving an unknown parameter, replace the unknown parameter with an estimate. A theoretically optimal bandwidth based on the "plug-in" procedure yields asymptotically efficient rate of convergence, but their feasibility in practice will depend on how far the real quantities are from the theoretical ones.

A data-driven bandwidth h can be computed using a cross-validation function, which attempts to minimize the sum of squared errors between smoothed and actual values of the dependent variable:

$$\mathrm{CV}(h) = \sum_{i=1}^{n} \left\{ C_i - \widehat{C}_{h,-i}(\mathbf{Z}_i) \right\}^2$$

where $\widehat{C}_{h,-i}(\bullet)$ denotes the kernel regression estimate which is obtained without using the i-th observation $(\mathbf{Z}_i, C_i)$. This way we ensure that the observations used for calculating $\widehat{C}_{h,-i}(\bullet)$ are independent of $\mathbf{Z}_i$, and rules out an artificial undersmoothing.

It can be shown that the bandwidth selected by minimizing CV fulfills an optimality property. Denote the bandwidth selected by the cross-validation criterion by $\widehat{h}_{cv}$, then this bandwidth is asymptotically optimal in the following sense

$$\frac{ISE(\hat{h}_{cv})}{\min_h ISE(h)} \xrightarrow{a.s.} 1$$

where $a.s.$ indicates convergence with probability 1. In other words, this means that the for $\widehat{h}_{cv}$ asymptotically coincides with h_{opt}

Bandwidth choice for the derivative estimation is similar to the choice of the bandwidth for the function itself. The theoretically optimal bandwidth for the derivatives is $n^{-}1/(s+2l+4)$ for kernel of order $l+1$, where s stands for derivative. Using cross-validation, we will minimize instead:

$$\mathrm{CV}(h) = \sum_{i=1}^{n} \left\{ C_i^{(2)} - \widehat{C}''_{h,-i}(\mathbf{Z}_i) \right\}^2$$

where $C_i^{(2)}$ is the second derivative based on scaled differences of the observations.

2.2 Dimension reduction techniques

While flexible, a high-dimension kernel regression requires large data samples for precise results in terms of confidence intervals. Ait-Sahalia and Lo (1998), for example, use a year's worth of option data to empirically derive the call function based on five-dimensional kernel regression. Quite generally, larger samples make for smaller variances. Conversely, small samples make the bias-variance tradeoff even more accute. For a given bias, the variance of the model

response is larger than for a model built from a larger sample. This is referred to as the 'curse of dimensionality'. Hence, there is a need to keep the dimension or equivalently the number of regressors low.

There has been lot of effort in developing 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. The resulting models are so-called semiparametric models, among which the additive models are the most feasible methods considered in practice. The basic idea of additive models is to take advantage of the fact that a regression surface may be of a simpler structure, that is a function of only certain linear combinations of the coordinates of the predictor variables such that high dimensional surface collapses down to one-dimensional surface.

It is behind the purpose of this chapter to investigate in detail the additive models; instead, we will review some 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

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

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

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

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 K so that:

$$C(S_t, K, \tau, r_{t,\tau}, \delta_{t,\tau}) = KC(S_t/K, \tau, r_{t,\tau}, \delta_{t,\tau})$$

Combining the assumptions of the last two equations, the call pricing 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 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.

These dimension reduction techniques may be used both in direct estimation of RND from the call prices and indirectly via implied volatility. In the empirical study that follows we will use only one regressor.

2.3 Data description

We use tick statistics 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 divident 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 (E.g. Jackwerth 2000) 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.

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 $\frac{K_i}{F_i} S_t e^{r_{t,\tau}-\delta_{t,\tau}}$, where $_i$ 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. 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:

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

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:

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

Finally, if we have different call price observations observationsfor 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.

2.4 Smoothing in call option space

As described in section 2.1 local polynomial method allows us to compute the second derivative of the call price directly, in a single step. We use local polynomial smoothing of degree three and quartic kernel. 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 yields a wiggly estimator and we decide to increase further the bandwidth. In figure (**??**) the black curve depicts $\hat{q}$ for the optimal bandwidth for the second

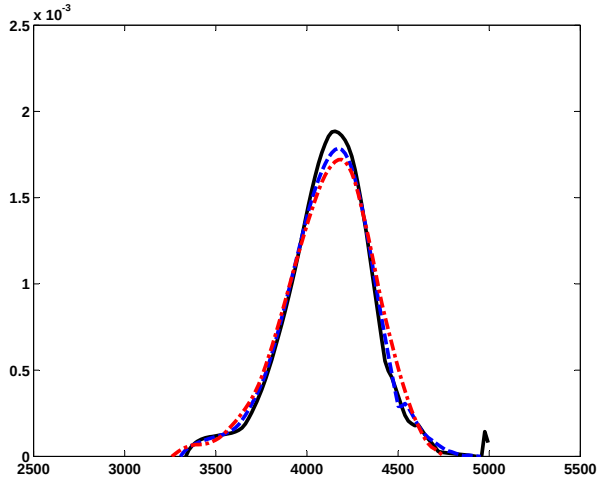

Fig. 1. $\hat{q}(S_T)$ by local ploynomial smoother: $h = 268.64$, $h = 392.63$, $h = 516.62$

derivative by rule-of thumb. This estimator is oversmoothed. If we further increase the bandwidth the estimator becomes even smoother at the expense of the bias.

2.5 Smoothing in implied volatility space

In practice, the smoothing is done in the implied volatility direction because call prices are a more volatile function of the underlying asset price. We use Black-Scholes formula as a mapping device for translating an option's price to implied volatility. In this case we estimate a smooth function $\hat{\sigma}$ and recover the call price by a bijective transformation according to the Black-Scholes formula evaluated at some fixed values of the regressors and variable σ:

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

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

$$y = \frac{\ln \frac{S_t}{K} + (b - \frac{1}{2}\sigma^2)\tau}{\sigma\sqrt{\tau}}$$

Herein, 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 in a modified version

of the Black-Scholes formula. Below we describe the method for fixed maturity one month.

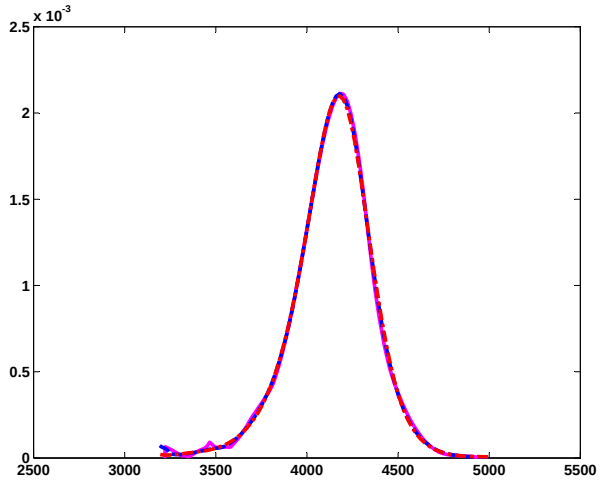

Fig. 2. $\hat{q}(S_T)$ by Rookley method with $h = 0.065$ (magenta), $h = 0.095$ (blue), $h = 0.125$ (red)

For each pair (C_i, K_i) we define the rescaled call option $c_i = C_i/S_t$ in terms of moneyness $M_i = S_t/K_i$ so that starting from the Black-Scholes formula for the call price we can write:

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

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

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

with

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

and

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

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 quartic kernel. The optimal bandwidth has been computed using cross-validation criteria for the implied volatility function and applied to smooth the function itself and its derivatives - figure (**??**), the magenta curve. Rookley shows that the results obtained by using for each derivative the optimal bandwidth yield insignificant changes in the estimation of q. In figure (**??**) we further increase the bandwidth and observe that oversmoothing improves the tails while having little effects on the values of $\hat{q}$ situated in the middle of the distribution. It follows that smoothing in implied volatility yields a more robust estimator to the changes in the bandwidth. It is because the implied volatility is less sensitive to the changes in strike price than the call price.

2.6 Estimation at the boundaries

The support of strike prices is in applications mostly compact, bounded. The occurence of such boundaries affects the quality of the estimates in these regions. By using odd order polynomial smoothing methods, design adaptive estimates of the functions of interest can automatically correct otherwise existing boundary bias. We have already mentioned the superiority of higher order polynomial regression compared to the Nadaraya-Watson estimator in this sense. Though, 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 the question how to join the two regions in order to assure that the resulting distribution integrates to one. Alternatively, Rookley (1997) proposes to further parameterize these distributions by matching them with an edgeworth expansion type density.

2.7 Estimation under constrains

The nonparametric estimate of the function C must satisfy some conditions: (1) it is positive, (2) it is decreasing in K, (3) it is convex, (4) its second derivative exists and it is equal up to a constant to a density (i.e. nonnegative and it integrates to one). Given that the first derivative of C with respect to K is the (negative) discounted cumulative density function of q conditions (2) and condition (3) can be summarized by the following inequality:

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

Convexity requires:

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

Nonparametric regressors may violate these constrains, unless we deal with large samples of observations. In these situations, it is recommended to use constrained weighted least squares regression in order to impose them artificially. 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).

3 Estimation via direct estimation of the pricing kernel using basis functions

In financial mathematics the relation between the physical p and the risk neutral measure q of a financial asset can be represented via pricing kernel.

We develop the model by assuming that all the variables other than K are fixed. The price of the European call option with strike price K expiring in τ years under the historical measure p is given by:

$$\begin{aligned} C(K) &= e^{-r_{t,\tau}\tau} \int_0^\infty \max(S_T - K, 0) \frac{q(S_T)}{p(S_T)} p(S_T) dS_T \\ &= e^{-r_{t,\tau}\tau} \int_0^\infty \max(S_T - K, 0) m(S_T) p(S_T) dS_T, \end{aligned} \tag{3}$$

where p is the 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 pricing kernels depends on many quantities, in the same fashion q does. However, here we are interested in 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 only. We assume that m and m^* are close in a certain sense (L_2) and that m^* has the following representation: $m^*(S_T) = \sum_{l=1}^{L} \alpha_l g_l(S_T)$, where α_l are coefficients estimated from the data, g_l are known basis functions with $l = \{1, ..., L\}$ and L is fixed.

We assume that the for each paired call-price observation (C_i, K_i) with $i = 1, ..., n$ it holds:

$$C_i(K_i) = C(K_i) + \varepsilon_i, \quad \text{with} \quad \varepsilon_i \quad \text{i.i.d. and} \quad \mathrm{E}[\varepsilon_i | K_i] = 0$$

At every point K we estimate function $C(K)$ based on (**??**):

$$\hat{C}(K) = e^{-r_{t,\tau}\tau} \frac{1}{J} \sum_{j=1}^{J} \max(S_T - K, 0) \sum_{l=1}^{L} \alpha_l g_l(S_T)$$

and find the vector $\alpha = (\alpha_1, ..., \alpha_L)^\top$ based on the following optimization procedure:

$$\min_{\alpha} \sum_{i=1}^{n} \left\{ C_i(K_i) - \hat{C}(K_i) \right\}^2 \tag{4}$$

where the expectation operator has been replaced by the sum.

3.1 Choice of the tuning parameters

When using an orthogonal series, one needs to choose the value of L for the largest term in the series. Here the role of L (or L/n) is similar to that played by the smoothing parameter h for the kernel methods. The following three well-known procedures for selecting L have been studied by Whaba (1985) and Li(1987).

Cross-validation, Mallows(1985): Select $\hat{L}$ to minimize:

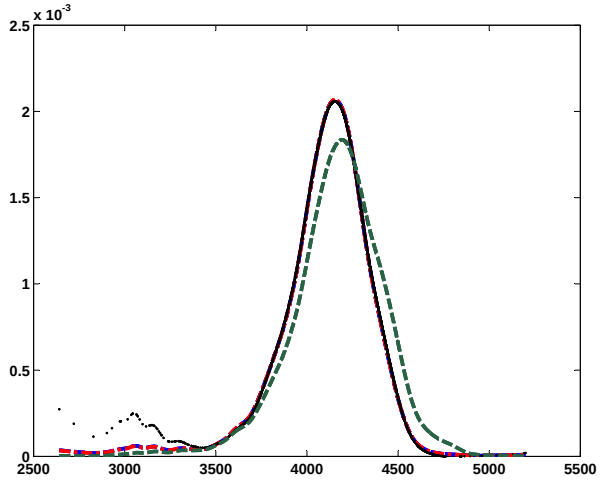

Fig. 3. $\hat{q}(S_T)$ by varying number of functions: 5 (blue), 6 (red), 7 (black)

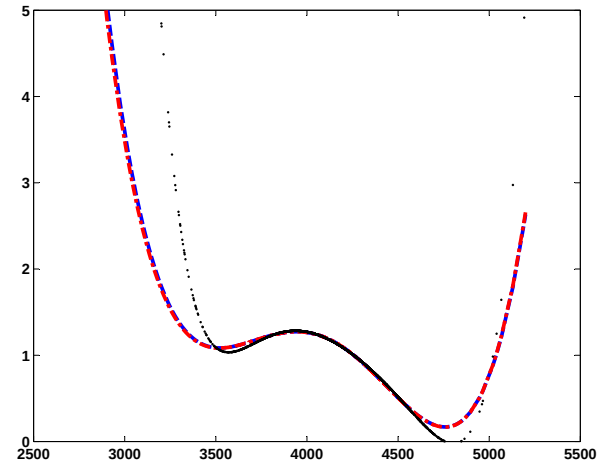

Fig. 4. $\hat{m}^*(S_T)$ by varying number of functions: 5 (blue), 6 (red), 7 (black)

$$\hat{L} = \arg\min_{L} n^{-1} \sum_{i=1}^{n} \left\{ C_i - \hat{C}(K_i) \right\}^2 + 2\sigma^2 (L/n)$$

where σ^2 is the variance of ε_{ti}. In practice one can estimate σ^2 by:

$$\hat{\sigma}^2 = n^{-1} \sum_{i=1}^{n} e_i^2, \quad \text{with} \quad e_i = C_i - \hat{C}(K_i)$$

Generalized cross-validation, Craven and Wahba (1979)): Select $\hat{L}$ to minimize:

$$\hat{L} = \arg\min_{L} \frac{n^{-1} \sum_{i=1}^{n} \left\{ C_i - \hat{C}(K_i) \right\}^2}{(1 - (L/n))^2}$$

Leave on out cross-validation, Stone (1974)): Select $\hat{L}$ to minimize:

$$CV_L \sum_{i=1}^{n} \left\{ C_i - \hat{C}_{-i}(K_i) \right\}^2$$

where $\hat{C}_{-i}(K_i)$ is the leave one estimate of $C(K_i)$ obtained by removing (K_i, C_i)

Li (1987) showed that each of the above three procedures leads to optimally selected L in the sense that the resultion $AWISE$ equals the smallest possible weighted integrated square error.

3.2 Practical aspects related to the implementation

We have chosen the Legendre polynomials, which constitute a system of orthonormal basis on $[-1, 1]$. The Legendre Polynomials may be generated recursively using relation the first two polynomials $P_0(x) = 1, \quad P_1(x) = x$ along with the following relation:

$$(n+1)P_{n+1}(x) = (2n+1)xP_n(x) - nP_{n-1}(x)$$

Other orthogonal basis can be chosen as well: e.g. Chebyshev, Laplace. While asymptotically equivalent, in finite sample the parametric form will influence the choice of L. In order to ensure that every observation is in the specified interval we have replaced every observation S_T by the arguments $x = \frac{2S_T - S_T^{min} - S_T^{max}}{S_T^{max} - S_T^{min}}$ of the basis functions. The physical density of S at T has been estimated using last 5000 daily observation of the DAX Index by kernel density estimator (quartic kernel) and optimal bandwidth by plug-in method.

The same constrains as in the case of kernel methods can be imposed here as well. Additional smoothing can be introduced via the smoothness penalty which adds to (**??**) via an additional smoothing parameter λ:

$$\min_{\alpha,\lambda} \sum_{i=1}^{n} \left\{ C_i(K_i, \tau) - \hat{C}(K_i) \right\}^2 + \lambda \int_0^{\infty} q''(S_T) dS_T \tag{5}$$

There is an easiness derived from the analytical form of the basis functions for the computation of q''. We can choose λ by minimising (**??**) with respect to α and λ jointly. Other methods like cross-validation can be applied in practice.

4 Comparison of the empirical results

Kernel methods for the estimation of q work much better when we estimate in implied volatility space than in strike price space. Figure (**??**) shows estimates by both methods with the same bandwidth: The second alternative (black curve) requires a large bandwidth in order to obtain a smooth function which results in an increasing bias of the estimate. Rookley method yields an estimate (magenta curve) which is more stable. Estimates of risk neutral density based on pricing kernel have similar performance with those resting upon smoothing in implied volatility direction. Varying number of basis function influences the boundary of the density, and not the central part, rich in observations. The shape of the pricing kernel is however, more sensitive to the number of basis functions. This last method can be preferred in some situations that involve dynamic analysis of a bundle of curves, because it summarises the nonparametric curves in possibly easily interpretable parameters.

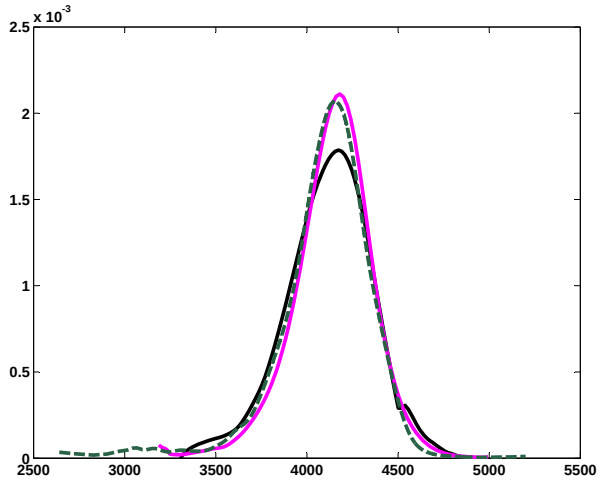

Fig. 5. $\hat{q}(S_T)$ by local polynomial regression of the call prices on the strike price with $h = 392.63$ (black), by Rookley method, smoothing implied volatility as a function of moneyness $h = 0.095$ (magenta), indirect estimation of the pricing kernel as basis expansion with $L = 6$ (green)

5 Conclusion

We have reviewed two methods for the estimation of the risk neutral density based on different techniques: on local features and global curve fitting

procedure respectively. For the two approaches we have described the estimation methodology and compared their performance empirically, showing which method is more robust and stable.

References

1. Ait-Sahalia, Y. and Lo, A. W. (2000) Nonparametric risk management and implied risk aversion, *Journal of Econometrics*, **94**, 9-51.
2. Breeden, D. T. and Litzenberger, R. H. Prices of State-Contingent Claims Implicit, *Option Prices The Journal of Business*, The University of Chicago Press, 1978, 51, 621-651
3. Bondarenko, O. Estimation of risk-neutral densities using positive convolution approximation, *Journal Econometrics*, 2003, 116, 85-112
4. Fengler, M. R. Semiparametric Modeling of Implied Volatility 2005
5. Giacomini, E., Handel, M. and Hrdle, W. Time Dependent Relative Risk Aversion, *Risk Assessment: Decisions in Banking and Finance*, Bol, G., Rachev, S., Wrth, R., Springer, 2008, pp. 15-46
6. Härdle, W. and Hlavka, Z., Dynamics of state price densities, *Journal of Econometrics*, Vol. 150, 2009, pp. 1-15
7. Härdle, W., Krtschmer, V, and Moro, R., A Microeconomic Explanation of the EPK Paradox, *Journal of Financial Econometrics*, submitted on 18.09.09
8. Härdle, W. and Marron, J. S. (1990) Semiparametric comparison of regression curves. *Annals of Statistics*, **18**, 63-89.
9. Härdle, W.; Marron, J. S. and Wand, M. P. Bandwidth choice for density derivatives J. Roy. *Statist. Soc. Ser.* B, 1990, 52, 223-232
10. Hartvig, N. V.; Jensen, J. L. and Pedersen, J. A class of risk neutral densities with heavy tails, *Finance Stoch.*, 2001, 5, 115-128Huynh, K., Kervella, P. and Zheng, J. (2002)
11. Estimating state-price densities with nonparametric regression, *Applied Quantitative Finance*, Edited by W. Härdle, T. Kleinov and G. Stahl, Springer, Berlin, Ch. 8.Jackwerth, J. C.(2000) Recovering risk aversion from option prices and realized returns, *Review of Financial Studies*, **13**, 433-451.
12. Jackwerth, J. C. Option-implied risk-neutral distributions and implied binomial trees: a literature review *Journal of Derivatives*, 1999, 2, 66-82
13. Rookley, C. (1997) Fully exploiting the information content of intra day option quotes: Applications in option pricing and risk management, University of Arizona.
14. Wand, M. P. and Jones, M. C. (1995) Kernel Smoothing. Chapman & Hall