Pricing Kernel Modeling

Denis Belomestny
Universität Duisburg-Essen
denis.belomestny@uni-due.de

Shujie Ma
Department of Statistics
University of California at Riverside, Riverside
shujie.ma@ucr.edu

Wolfgang Karl Härdle
Ladislaus von Bortkiewicz Chair of Statistics
CASE-Center for Applied Statistics and Economics
Humboldt-Universität zu Berlin
D-10178 Berlin, Germany
Sim Kee Boon Institute for Financial
Economics, Singapore Management University, Singapore
haerdle@wiwi.hu-berlin.de

December 29, 2014

Abstract We propose a new method to estimate the empirical pricing kernel based on option data. We estimate the pricing kernel nonparametrically by using the ratio of the risk-neutral density estimator and the subjective density estimator. The risk-neutral density is approximated by a weighted kernel density estimator with varying unknown weights for different observations, and the subjective density is approximated by a kernel density estimator with equal weights. We represent the European call option price function by the second order integration of the risk-neutral density, so that the unknown weights are obtained through one-step penalized least squares estimation with the Kullback-Leibler divergence as the penalty function. Asymptotic results of the resulting estimators are established. The performance of the proposed method is illustrated empirically by simulation and real data application studies.

Key words: Empirical Pricing Kernel; Kernel; Kernel Density Estimation; Nonparametric Fitting; Kullback-Leibler Divergence

JEL classification: C00, C14, G12

Acknowledgement: This research was financially supported by the Deutsche Forschungsgemeinschaft (DFG) via SFB 649 "Ökonomisches Risiko" and IRTG 1792 "High-Dimensional Non-Stationary Times Series" is gratefully acknowledged.

1 Introduction

The pricing kernel (PK) is an important link between economics and finance and it plays a pivotal role in assessing the risk aversion over equity returns. The method introduced by Aït-Sahalia and Lo (2000) to construct empirical pricing kernels has been used frequently in the literature. In their paper, a pricing kernel K is defined as a ratio of the economic risk which contains the preferences of investors and statistical risk which provides information on the dynamics of the data generating process (DGP). The economic risk is well assessed by Arrow-Debreu prices and can be estimated by the risk neutral density q obtained from the derivative market. Thus, obtaining an accurate estimator of q is a crucial step for pricing kernel estimation. Aït-Sahalia and Lo (2000) present several methods to estimate q by using different models or nonparametric estimators, e.g., a smooth local volatility part of the Black and Scholes (1973) method. Breeden and Litzenberger (1987) have demonstrated that for a continuum of strikes the risk neutral density is proportional with the quotient of the European call options with respect to the strike price. Härdle et al. (2014) developed uniform confidence bands that proved to be helpful for testing parametric specifications of pricing kernels and permitted extension to estimating risk aversion patterns. Golubev et al. (2014) and Beare and Schmidt (2014) proposed statistical tests of pricing kernel monotonicity. Beare (2011) has shown how the theory of monotone rearrangements may be used to derive an explicit solution for the cost minimizing measure preserving derivative written on some underlying asset. Beare and Schmidt (2015) refer to the phenomenon of the nonmonotone shape of pricing kernel estimates as the first order of stochastic dominance, in the sense that the returns of a portfolio of contingent claims written on the index are significantly higher than the index return. Moreover, Grith et al. (2013) proposed a systematic modelling approach to describing the evolution of the empirical pricing kernels, and investigated the relationship of the model and the shape of the EPK to the economic conditions. Grith et al. (2015) retained the expected utility framework in a one period model and illustrated the case when the state is defined with respect to a reference point. They further investigated how the model relates the shape of the EPK to the economic conditions.

In this paper, we take a fresh look at this problem and propose to estimate the pricing kernel nonparametrically by controlling the ratio of the risk-neutral density and the subjective density. The risk-neutral density is approximated by a weighted kernel density estimator with varying unknown weights for different observations. By observing stock prices or returns that investors expect to obtain at time to maturity, the subjective density can be approximated by the kernel density estimate of historical stock prices with equal weights. We represent the European call option price function by the second order integration of the risk-neutral density, so that the unknown weights are obtained through one-step penalized least squares estimation with the Kullback-Leibler divergence as the penalty function. Statistical risk provides an overview over statistical properties of the DGP and is given by the distribution p of future prices conditional on current prices also known as historical density. The historical density p can be estimated by using the past of the stock time series S_t. Due to the large number of observations in the derivative option market, the risk neutral density q can be well estimated with large-sample asymptotic properties established.

The paper is organized as follows. In Section 2 and 3 we formulate the problem and construct the estimators. In Section 4 and 5 we present the main and auxiliary theoretical results. In Section 6, we present the special case of KL-divergence of log-normals. Section 7 provides empirical results of a simulation study. The proposed estimation procedure is illustrated by analyzing a Strike-Call price dataset in Section 8. As in earlier studies, the PK is non-monotone and allows interpretation of time varying risk preferences.

2 Problem formulation

Let $X_1, ..., X_n$ be i.i.d. random variables distributed with a density $p(x)$. Let further $(Z_1, Y_1), \dots, (Z_m, Y_m)$ be a sample of pairs of explanatory and response variables satisfying

$$Y_i = f(Z_i) + \sigma\varepsilon_i, \quad i = 1, \dots, m, \tag{2.1}$$

where ε_i are $\mathrm{N}(0,1)$ i.i.d. random variables. The additive errors scheme (2.1) applies to European call prices in an intraday context. In fact, for statistical analysis Renault (1997) interprets the error as mispricings which could be exploited by arbitrage strategies. Denote $\|\varphi\|_m^2 = \frac{1}{m}\sum_{i=1}^m |\varphi(Z_m)|^2$ for any function φ on $\mathbb{R}$. The problem is to find a probability density q minimizing

$$\mathcal{Q}(q) \stackrel{\text{def}}{=} \|f - Aq\|_m^2 + \lambda\,\mathrm{KL}(q||p), \quad \lambda > 0 \tag{2.2}$$

where A is the operator of second order integration and $\mathrm{KL}(\bullet||\bullet)$ is Kullback-Leibler divergence

$$\mathrm{KL}(q||p) \stackrel{\text{def}}{=} \int_{\mathbb{R}} q(x) \log \frac{q(x)}{p(x)}\, dx.$$

The penalized version of the least squares problem (2.2) leads to estimating q:

$$\begin{aligned} &\text{minimize} - 2\int_{\mathbb{R}} \omega(x)\, q(x)\, dP_n(x) + \int_{\mathbb{R}} |q(x)|^2\, dx + h^2 R(q) \\ &\text{subject to } q \text{ is a continuous density,} \end{aligned} \tag{2.3}$$

where $P_n(x)$ is the cdf of $\{X_i\}_{i=1}^n$, ω stands for the Radon-Nikodym derivative dQ/dP, R is a roughness penalization term and h is the smoothing parameter. Under the choice

$$R(q) = \int_{\mathbb{R}} |q'|^{\,2} dx$$

the solution q of (2.3) satisfies the boundary value problem

$$\begin{aligned} -h^2 q''(x) + q(x) = \omega\,(x) dP_n(x), &\quad -\infty < x < \infty \\ q(x) \to 0, &\quad |x| \to \infty \end{aligned} \tag{2.4}$$

and is given by

$$q_n(x) = \frac{1}{nh} \sum_{i=1}^{n} \omega(X_i) K\left(\frac{x - X_i}{h}\right)$$

provided that $n^{-1} \sum_{i=1}^{n} \omega(X_i) = 1$ and $K(\bullet)$ is a two-sided exponential kernel given as

$$K(u) = \frac{1}{2} \exp(-|u|), \tag{2.5}$$

for $u \in \mathbb{R}$, see Vapnik (1995) and Vapnik (1998). For the fixed design $X_1, \ldots, X_n$ we approximate the solution of the minimization problem (2.3) by

$$q_{n,m} = \operatorname{argmin}_{q \in \mathcal{C}_{n,X}} \mathcal{Q}_{n,m}(q),$$

where

$$\mathcal{C}_{n,X} \stackrel{\text{def}}{=} \left\{ \sum_{i=1}^{n} w_i K_h(x - X_i), \quad n^{-1} \sum_{i=1}^{n} w_i = 1 \right\},$$

with $w_i = \omega(X_i)$, and

$$K_h(x) \stackrel{\text{def}}{=} \frac{1}{h} K\left(\frac{x}{h}\right), \quad x \in \mathbb{R}$$

and

$$\mathcal{Q}_{n,m}(q) \stackrel{\text{def}}{=} \|f - Aq\|_m^2 - \frac{\lambda}{n} \sum_{i=1}^{n} w_i \log w_i. \tag{2.6}$$

Further, instead of (2.6) we consider its empirical version:

$$\tilde{\mathcal{Q}}_{n,m}(q) \stackrel{\text{def}}{=} \frac{1}{m} \sum_{i=1}^{m} (\{Y_i - (Aq)(Z_i)\}^2 - \frac{\lambda}{n} \sum_{i=1}^{n} w_i \log w_i.$$

The form of the penalty can be motivated by the fact that

$$n^{-1} \sum_{i=1}^{n} w_i \log w_i = \int_{\mathbb{R}} \omega(x) \log \omega(x) \, dP_n \xrightarrow{P} \int_{\mathbb{R}} \omega(x) \log \omega(x) \, dP = KL(q||p).$$

2.1 Finitely many constraints

The problem is

$$\begin{aligned} &\text{minimize } L(f) \stackrel{\text{def}}{=} \int_{\Omega} m\{x, f(x)\} \, dx + \frac{1}{2} h^2 \|\nabla f\|^2 \\ &\text{subject to } f \in W^{1,2}(\Omega), \quad Af = c, \end{aligned} \tag{2.7}$$

where $\Omega = domain(X)$ and $m(x, u)$ satisfies the assumptions

1. $m\{\bullet, f(\bullet)\} \in L^1(\Omega)$ for all $f \in W^{1,2}(\Omega)$

2. $m(x, \bullet)$ is differentiable on $\mathbb{R}$ for all $x \in \Omega$

3. $m_u\{\bullet, f(\bullet)\} \in L^2(\Omega)$ for all $f \in W^{1,2}(\Omega)$

and $A : L^2(\Omega) \to \mathbb{R}^p$ is bounded linear operator.

Theorem 2.1. *The function $f_0 \in W^{1,2}(\Omega)$ solves (2.7) if and only if $f_0 \in W^{2,2}(\Omega)$ and there exists $\lambda_0 \in \mathbb{R}^p$ such that (f_0, λ_0) solves the boundary value problem with constraints*

$$\begin{aligned} -h^2 \nabla f(x) + m_u(x, f(x)) + A^* \lambda(x) = 0,\ x \in \Omega \\ \frac{\partial f}{\partial n} = 0,\ x \in \partial\Omega \\ Af = c, \end{aligned} \tag{2.8}$$

where n is the outward normal on $\partial\Omega$.

3 Main results

Theorem 3.1. *Let $q_{n,m}$ be the minimizer of $\mathcal{Q}_{n,m}(q)$, $\tilde{q}_{n,m}$ be the minimizer of $\tilde{\mathcal{Q}}_{n,m}(q)$ and $\Delta \stackrel{\mathrm{def}}{=} \tilde{q}_{n,m} - q_{n,m}$. If*

$$\|f\|_\infty + \|AK_h\|_\infty < \infty \tag{A1}$$

and

$$\inf_{q \in \mathcal{C}_{n,X}} Q_{n,m}(q) < \lambda/4 \tag{A2}$$

then

$$P_Y\left\{ \|A\Delta\|_m^2 \geq C\sqrt{\frac{U}{m}} + \mathcal{O}(\lambda) \right\} \leq n \exp(-U^2/B^2) \tag{3.1}$$

for some constants $C > 0$ and $B = B(f, K_h, \sigma)$.

PROOF. We have

$$\mathcal{Q}_{n,m}(q) - \tilde{\mathcal{Q}}_{n,m}(q) = T_1 + \lambda T_2,$$

where

$$T_1 \stackrel{\mathrm{def}}{=} m^{-1} \sum_{i=1}^{m} \left\{ |f(Z_i) - (Aq)(Z_i)|^2 - |Y_i - (Aq)(Z_i)|^2 \right\}$$

and

$$T_2 \stackrel{\mathrm{def}}{=} n^{-1} \sum_{i=1}^{n} \log \frac{p(X_i)}{q(X_i)} + n^{-1} \sum_{i=1}^{n} \log(nw_i).$$

i Bounds for T_1

We have

$$T_1 = -\frac{1}{m}\sum_{i=1}^{m}\varepsilon_i^2 + \frac{2}{m}\sum_{i=1}^{m} f(Z_i)\varepsilon_i + \frac{2}{m}\sum_{i=1}^{m}(Aq)(Z_i)\varepsilon_i = T_{11} + T_{12} + T_{13}$$

where $\varepsilon_i \sim \mathrm{N}(0,\sigma^2)$. Since we consider q being in the class of convex combinations $\mathcal{C}_{n,X}$ and a linear functional of convex combinations achieves its maximum value at the vertices we have:

$$\sup_{q\in\mathcal{C}_{n,X}} |T_{13}(q)| \le 2 \sup_{q\in\mathcal{C}_{n,X}} \left|\frac{1}{m}\sum_{i=1}^{m}(Aq)(Z_i)\varepsilon_i\right| = 2 \max_{j=1,\ldots,n}\left|\frac{1}{m}\sum_{i=1}^{m}(AK_{jh})(Z_i)\varepsilon_i\right|,$$

where $K_{jh} = K_h(x - X_j)$. Hence and due to lemma 4.1

$$P_Y(\sup_{q\in\mathcal{C}_{n,X}} |T_{13}(q)| > U/\sqrt{m}) \le 2n\exp\left(-\frac{U^2}{8\sigma^2 \max_j \|AK_{jh}\|_m^2}\right).$$

Similarly, lemma 4.1 implies

$$P_Y(|T_{12}| > U/\sqrt{m}) \le 2\exp\left(-\frac{U^2}{8\sigma^2\|f\|_m^2}\right)$$

and

$$P_Y(|T_{11} + \sigma^2| > U/\sqrt{m}) \le \exp\left(-\frac{U^2}{4\sigma^4}\right) + \exp\left(-\frac{U\sqrt{m}}{3\sigma^2}\right).$$

For $U \ll \sqrt{m}$

$$P_Y\left(|T_1| > \sqrt{\frac{U}{m}}\right) < n\exp(-u^2/B^2)$$

where $B^2 = \max\{8\sigma^2 \max_j \|AK_{jh}\|_m^2, 8\sigma^2\|f\|_m^2, 4\sigma^4\}$.

ii Bounds for T_2

First, let us define

$$T_{21} \stackrel{\text{def}}{=} n^{-1}\sum_{i=1}^{n}\log\frac{q_n(X_i;W)}{p_n(X_i)} - n^{-1}\sum_{i=1}^{n}\log nw_i = n^{-1}\sum_{i=1}^{n}(\{\log(\xi_i) - \log w_i\},$$

where

$$\xi_i = \left\{\frac{\sum_{j=1}^{n} w_j K_{ih}(X_j)}{\sum_{j=1}^{n} K_{ih}(X_j)}\right\}$$

If function $\log w(\bullet)$ can be assumed Lipschitz then

$$T_{21} = \mathcal{O}_p(h).$$

Further,

$$T_{22} = n^{-1}\sum_{i=1}^{n} \log \frac{q_n(X_i;W)}{p_n(X_i)} - n^{-1}\sum_{i=1}^{n} \log \frac{q_n(X_i;W)}{p(X_i)}$$
$$= n^{-1}\sum_{i=1}^{n} \log p(X_i) - n^{-1}\sum_{i=1}^{n} \log p_n(X_i)$$

and it is well known that $T_{22} = \mathcal{o}(n^{-1/2})$ almost surely, provided

$$\mathrm{E}\,|X|^L < \infty,\ L > \kappa > 2, \quad h \asymp n^{-\kappa/(3\kappa+2)}, \quad p^{1/2} \in W^{2,2}(\mathbb{R}). \tag{A3}$$

Let us now note that $T_2 = T_{21} + T_{22}$. Thus we have proved that with probability greater than $1 - n\exp(-u^2/B^2)$

$$\sup_{q\in\mathcal{C}_{n,X}} |\mathcal{Q}_{n,m}(q) - \tilde{\mathcal{Q}}_{n,m}(q)| \le \sqrt{\frac{U}{m}} + \mathcal{O}(\lambda h)$$

Hence,

$$0 \le \mathcal{Q}_{n,m}(\tilde{q}_{n,m}) - \mathcal{Q}_{n,m}(q_{n,m}) \le \mathcal{Q}_{n,m}(\tilde{q}_{n,m}) - \tilde{\mathcal{Q}}_{n,m}(q_{n,m}) + \sqrt{\frac{U}{m}} + \mathcal{O}(\lambda h)$$
$$\le \mathcal{Q}_{n,m}(\tilde{q}_{n,m}) - \tilde{\mathcal{Q}}_{n,m}(\tilde{q}_{n,m}) + \sqrt{\frac{U}{m}} + \mathcal{O}(\lambda h)$$
$$\le 2\sqrt{\frac{U}{m}} + \mathcal{O}(\lambda h)$$

with probability greater than $1 - n\exp(-U^2/B^2)$. On the other hand

$$0 \le \mathcal{Q}_{n,m}(\tilde{q}_{n,m}) - \mathcal{Q}_{n,m}(q_{n,m}) = \frac{2}{m}\sum_{i=1}^{m}(A\Delta)(Z_i)(Aq_{n,m} - f)(Z_i) +$$
$$+ \frac{1}{m}\sum_{i=1}^{m}(A\Delta)^2(Z_i) + \frac{\lambda}{n}\sum_{i=1}^{n}\log\frac{q_{n,m}(X_i)}{\tilde{q}_{n,m}(X_i)}.$$

Without loss of generality we can assume that $\frac{1}{n}\sum_{i=1}^{n}\log\frac{p(X_i)}{q_{n,m}(X_i)} \ge 0$ and due to (A2)

$$\|f - Aq_{n,m}\|_m^2 < \lambda/4$$

If $\|A\Delta\|_m^2 \ge 4\lambda$ then

$$\left|\frac{1}{m}\sum_{i=1}^{m}(A\Delta)(Z_i)(Aq_{n,m} - f)(Z_i)\right| \le \|A\Delta\|_m \|Aq_{n,m} - f\|_m \le \lambda^{1/2}\|A\Delta\|_m/2 \le \|A\Delta\|_m^2/4$$

and

$$\frac{2}{m}\sum_{i=1}^{m}(A\Delta)(Z_i)(Aq_{n,m}-f)(Z_i)+\|A\Delta\|_m^2 \geq \|A\Delta\|_m^2/2$$

Thus, either $\|A\Delta\|_m^2 \leq 4\lambda$ or

$$\frac{1}{2m}\sum_{i=1}^{m}(A\Delta)^2(Z_i)+\frac{\lambda}{n}\sum_{i=1}^{n}\log\frac{q_{n,m}(X_i)}{\tilde{q}_{n,m}(X_i)} \leq \mathcal{O}(\lambda h)+2\sqrt{\frac{U}{m}}.$$

It remains to prove that

$$n^{-1}\sum\nolimits_{i=1}^{n}\log\frac{q_{n,m}(X_i)}{\tilde{q}_{n,m}(X_i)}=\mathcal{O}_p(1).$$

Note that

$$n^{-1}\sum\nolimits_{i=1}^{n}\log\frac{p(X_i)}{q_{n,m}(X_i)}<1/4 \tag{3.2}$$

and

$$\begin{aligned}\frac{\lambda}{n}\sum_{i=1}^{n}\log\frac{p(X_i)}{\tilde{q}_{n,m}(X_i)} &\leq -\frac{\lambda}{n}\sum_{i=1}^{n}\log(n\tilde{w}_i)+\mathcal{O}(\lambda h) \\ &\leq \tilde{Q}_{n,m}(\tilde{q}_{n,m})+\mathcal{O}(\lambda h) \\ &\leq \tilde{Q}_{n,m}(q_{n,m})+\mathcal{O}(\lambda h) \\ &\leq \lambda/4+\sqrt{\frac{u}{m}}+\mathcal{O}(\lambda h)\end{aligned} \tag{3.3}$$

with probability greater than $1-n\exp\left(-u^2/B^2\right)$. Combining (3.2) and (3.3) one gets (3.1) $\square$

Let us define for any $q\in\mathcal{C}_{n,X}$

$$T(q)\stackrel{\text{def}}{=}\mathcal{Q}_{n,m}(q)-\mathcal{Q}(q)=\frac{1}{n}\sum_{i=1}^{n}\log\frac{q(X_i)}{p(X_i)}-\mathrm{KL}(p||q)$$

Clearly

$$\begin{aligned}\sup_{q\in\mathcal{C}_n}|T(q)| &= \sup_{q\in\mathcal{C}_n}\left|\frac{1}{n}\sum_{i=1}^{n}\log\frac{q(X_i)}{p(X_i)}-\mathrm{KL}(q||p)\right| \\ &\leq \sup_{q\in\tilde{\mathcal{C}}_n}\left|\frac{1}{n}\sum_{i=1}^{n}\log\frac{q(X_i)}{p(X_i)}-\mathrm{KL}(q||p)\right|,\end{aligned}$$

where

$$\tilde{\mathcal{C}}_n\stackrel{\text{def}}{=}\mathrm{conv}_n(\mathcal{H})=\left\{\sum_{i=1}^{n}w_iK_h(x-a_i),\quad \sum_{i=1}^{n}w_i=1,\quad a\in\mathbb{R}^n\right\}.$$

and

$$\mathcal{H} \stackrel{\text{def}}{=} \{K_h(x-a), \quad a \in \mathbb{R}\}.$$

The following lemma holds

Lemma 3.1. *If density p is such that $0 < a \le p(x) \le b$ for all x then with probability at least $1-\exp(-t)$*

$$\sup_{q \in \tilde{\mathcal{C}}_n} \left| \frac{1}{n} \sum_{i=1}^{n} \log \frac{q(X_i)}{p(X_i)} - \mathrm{KL}(q||p) \right| \le \mathrm{E}_X \left[\frac{c_1}{\sqrt{n}} \int_0^b \log^{1/2} \mathcal{D}(\mathcal{H}, \varepsilon, d_n)\, d\varepsilon \right] + c_2 \sqrt{\frac{t}{n}},$$

where c_1 and c_2 are constants that depend on a and b, $\mathcal{D}(\mathcal{H}, \varepsilon, d_n)$ is the covering number of $\mathcal{H}$ at scale ε with respect to empirical distance d_n

$$d_n^2(\varphi_1, \varphi_2) = n^{-1} \sum_{i=1}^{n} \{\varphi_1(X_i) - \varphi_2(X_i)\}.$$

We refer to Eggermont and LaRiccia (2001) for the proofs of Lemma 3.1.

4 Auxiliary results

Lemma 4.1. *Let $A = \{a_{ij}, \quad i,j = 1,\dots,N\}$ be a $N \times N$ matrix. Denote the values S_A and λ_A by*

$$S_A^2 = 2\,\mathrm{tr}(A^\top A)^2, \quad \lambda_A = \|AA^\top\|_\infty.$$

If $\varepsilon_1, \dots, \varepsilon_N$ are i.i.d. $N(0,\sigma^2)$ random variables and $b = (b_1, \dots, b_N)^\top$ is a deterministic vector then

$$P(2|b^\top A\varepsilon| > z\sigma\|b\|(2\lambda_A)^{1/2}) \le \exp\left(-z^2/2\right)$$

and

$$P(|\varepsilon^\top A^\top A\varepsilon - \mathrm{tr}(A^\top A)| > zS_A) \le \exp\left(-z^2/4\right) + \exp\left(-zS_A/6\lambda_A\right).$$

We refer to Spokoiny and Zhilova (2013) for the proofs of Lemma 4.1.

4.1 KL-divergence of log-normals

Let S_t be a stochastic process following a geometric Brownian motion (GBM) with drift μ and volatility σ. S_t follows the stochastic differential equation

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t,$$

where W_t is a Wiener process or Brownian motion. Let $p(x,\mu,\sigma)$ be the probability density function of a log-normal distribution with parameters μ and σ, then

$$p(x,\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left\{-\left(\log(x) - \mu\right)^2 / \left(2\sigma^2\right)\right\}.$$

Similarly define $q(x,r,\sigma)$ as the probability density function of a log-normal distribution with parameters r and σ, then

$$q(x,r,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left\{-(\log(x)-r)^2 / \left(2\sigma^2\right)\right\}.$$

Thus the log-likelihood ratio is

$$\begin{aligned}
\log\frac{q(x,\mu,\sigma)}{p(x,r,\sigma)} &= \log\{q(x,r,\sigma)\} - \log\{p(x,\mu,\sigma)\} \\
&= -(\log(x)-r)^2 / \left(2\sigma^2\right) + (\log(x)-\mu)^2 / \left(2\sigma^2\right) \\
&= (r-\mu)\{2(\log(x)-r) + (r-\mu)\} / \left(2\sigma^2\right).
\end{aligned}$$

The Kullback-Leibler divergence between p and q is

$$\begin{aligned}
\mathrm{KL}(q\,\|p) &= \int_{\mathbb{R}} q(x,\mu,\sigma)\log\frac{q(x,\mu,\sigma)}{p(x,r,\sigma)}dx \\
&= \left(\frac{r-\mu}{2\sigma^2}\right)\int_0^\infty \frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-(\log(x)-r)^2/\left(2\sigma^2\right)\right\}\{2(\log(x)-r)+(r-\mu)\}\,d\log(x) \\
&= \left(\frac{r-\mu}{2\sigma^2}\right)\int_{\mathbb{R}} \frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-(y-r)^2/\left(2\sigma^2\right)\right\}\{2(y-r)+(r-\mu)\}\,dy \\
&= \left(\frac{r-\mu}{2\sigma^2}\right)\left\{2\sigma^2+(r-\mu)\right\} = r-\mu+(r-\mu)^2/\left(2\sigma^2\right).
\end{aligned}$$

5 Simulation studies

In this section, we use a simulated example to illustrate the proposed nonparametric estimation procedure. We first generate $(X_t, 1 \le t \le n)$ with $n = 500$ i.i.d observations from the log-normal distribution with density function $p(x,\mu,\sigma)$, where $\mu = \log(2)$ and $\sigma = 0.5$, and then generate $(Z_i, 1 \le i \le m)$ i.i.d. with $m = 500$ independently from Uniform(0,5). The nonparametric function $f(\bullet)$ is simulated from the second order integration of log-normal probability density function $q(x,r,\sigma)$ with $r = 0$ and $\sigma = 0.5$. Then $f(z) = \int_z^M \int_u^\infty q(s)\,ds du = \int_z^M [1 - \Phi\{(\log u - r)/\sigma\}]\,du$, where $M = x_{\max}$. The error term $\varepsilon_i \sim N(0,1)$ and $\vartheta = 0.05$. We let the responses $(Y_i, 1 \le i \le m)$ be generated from the following model:

$$Y_i = f(Z_i) + \vartheta\varepsilon_i, i = 1,\ldots,m. \tag{5.1}$$

Figure 1 shows the plot of the true mean function $f(\bullet)$ (solid line) and the scatter plot of the simulated data points. Clearly, function $f(\bullet)$ has a decreasing pattern. The data set is simulated according to the real Strike-Call price data applications, in which the call price monotonically decreases as the strike price increases.

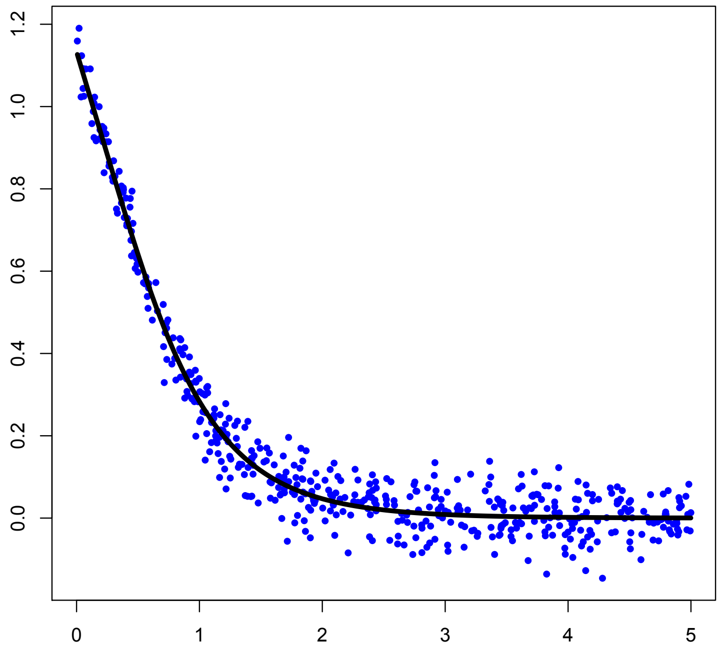

Figure 1: The plots of $f(z)$ (solid line) and $\{(Z_i, Y_i)\}_{i=1}^m$ (points) against z.

Then the estimated weights $\widehat{W} = \{\widehat{\omega}(X_t)\}_{t=1}^n$ are obtained by minimizing

$$\widetilde{Q}_{n,m}(q) = m^{-1}\sum_{i=1}^{m}\{Y_i - (Aq_n)(Z_i)\}^2 - \lambda n^{-1}\sum_{t=1}^{n}\log(n\omega_t), \tag{5.2}$$

subject to $\sum_{t=1}^n \omega_t = 1$, where $q_n(z; W) = \sum_{t=1}^n \omega_t K_h(z - X_t)$ and $\omega_t = \omega(X_t)$. Then $Aq_n(z; W) = \frac{1}{h}\sum_{t=1}^n w_t \int_z^M \left(\int_s^\infty K\left(\frac{u-X_t}{h}\right) du\right) ds$. Here we use the Gaussian kernel function $K(u) = \varphi(u)$ and let $\lambda = 0.01, 0.1, 0.5, 1$.

Figure 2 shows the plots of $q_n\left(x; \widehat{W}\right)$ (dashed thick line), $q(x)$ (solid thick line), the log-normal density function $p(x)$ (solid thin line) and its kernel estimate $p_n(x)$ (dashed thin line) against x with $\lambda = 0.01, 0.1, 0.5, 1$. The density estimate curves $p_n(x)$ are close to the true density curves $p(x)$ for all cases, so it demonstrates that $p(x)$ is well estimated by its kernel density estimator $p_n(x)$. When λ is small ($\lambda = 0.01$), the control over the KL-divergence is loose and therefore the weighted kernel density estimate curve $q_n\left(x; \widehat{W}\right)$ is close to the corresponding true density curve $q(x)$. As λ increases from 0.01 to 1, $q_n\left(x; \widehat{W}\right)$, the closer to $p(x)$ is more enforced and consequently is getting closer to $p(x)$ and moving further from $q(x)$.

Figure 3 shows the plots of the estimated nonparametric function $\widehat{f}(z) = (Aq_n)\left(z; \widehat{W}\right)$ (dashed line) and the true function $f(z)$ (solid line) against z with $\lambda = 0.01, 0.1, 0.5, 1$. The estimated function $\widehat{f}(z)$ is closer to the true function $f(z)$ for smaller value of λ. Figure 4 shows the empirical pricing kernel defined as EPK=$q_n\left(x; \widehat{W}\right)/p_n(x)$ for $\lambda = 0.01, 0.1, 0.5, 1$. We observe that the EPK function is a decreasing function for all λ.

6 Real data analysis

We use a Strike-Call dataset on November 16, 2011 to estimate model (5.1). The data are from RDC of CRC649 Berlin. There are $m = 1621$ observations on this day. Let $(Z_i, Y_i, i = 1, \ldots, 1621)$ be the strike and call prices. Figure 5 shows the scatter plot of the call prices against the strike prices. Clearly, the option call price has a monotone decreasing pattern. We use the realizations of historic stock price (DAX index) $(X_t, t = 1, \ldots, n)$ from March 12, 2009 to November 16, 2011, so that $n = 500$. Figure 6 shows the plot of DAX index during this time period. The density function of price X_t is estimated by $p_n(x) = n^{-1}\sum_{t=1}^{n} K_h(x - X_t)$. The risk-neutral density function is defined as in (2.5) $q_n(z) = n^{-1}\sum_{t=1}^{n} w(X_t)K_h(z - X_t)$. The estimated weights $\widehat{W} = \{\widehat{w}(X_t)\}_{t=1}^{n}$ are obtained by minimizing (5.2). Define the "moneyness" at time t as $M_t = X_t/Z$, i.e., the stock price are scaled against the observed strike prices so that e.g. "at the money" correspond to $M_t = 1$. Figure 7 contains the plots of $q_n\left(x; \widehat{W}\right)$ (dashed line) and $p_n(x)$ (solid line) against moneyness with tunning parameter $\lambda = 0.1, 0.5$. The graphs in the right and left panels look similar, so the estimation of the risk-neutral density function is not sensitive to the choice of different values for λ in this example.

Figure 8 displays the plots of the estimated nonparametric function $\widehat{f}(z) = (Aq_n)(z)$ (solid line) together with the data points at $\lambda = 0.1, 0.5$. One observes that both PK estimated mean curves fit the data well. Figure 9 finally presents the plot of the empirical PK estimates as the ratio of $q_n\left(x; \widehat{W}\right)$ and $p_n(x)$ against moneyness. Apparently, both of the PKs have a decreasing trend, but prominent peaks are around $M_t = 1$.

7 Conclusions

In this paper, we propose a new method to estimate the pricing kernel nonparametrically. Our new method further confirms the empirical pricing kernel (EPK) phenomenon that the pricing kernel (PK) is non-monotone and allows interpretation of time varying risk preferences. The formulation of the inverse problem opens new insight into statistically fitting EPKs. The proposed method is numerically reasonable. The numerical studies are implemented in the statistical software R and the programming packages will be provided in quantlet. As a future topic, we will study the dynamic patterns of EPKs.

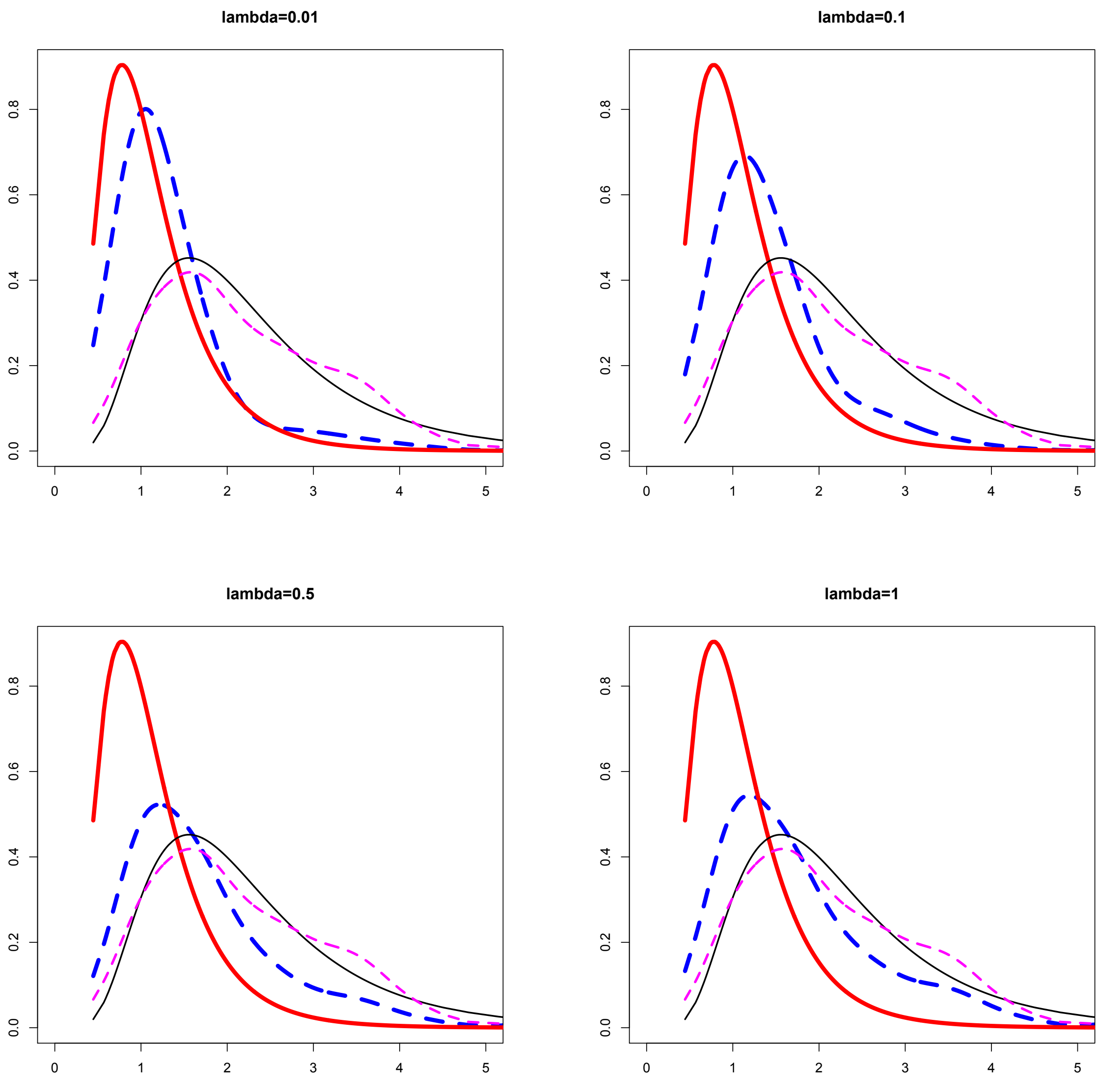

Figure 2: The plots of $q_n\left(x;\widehat{W}\right)$ (dashed thick line), $q\left(x\right)$ (solid thick line), $p\left(x\right)$ (solid thin line) and $p_n\left(x\right)$ (dashed thin line) against x with $\lambda = 0.01, 0.1, 0.5, 1$.

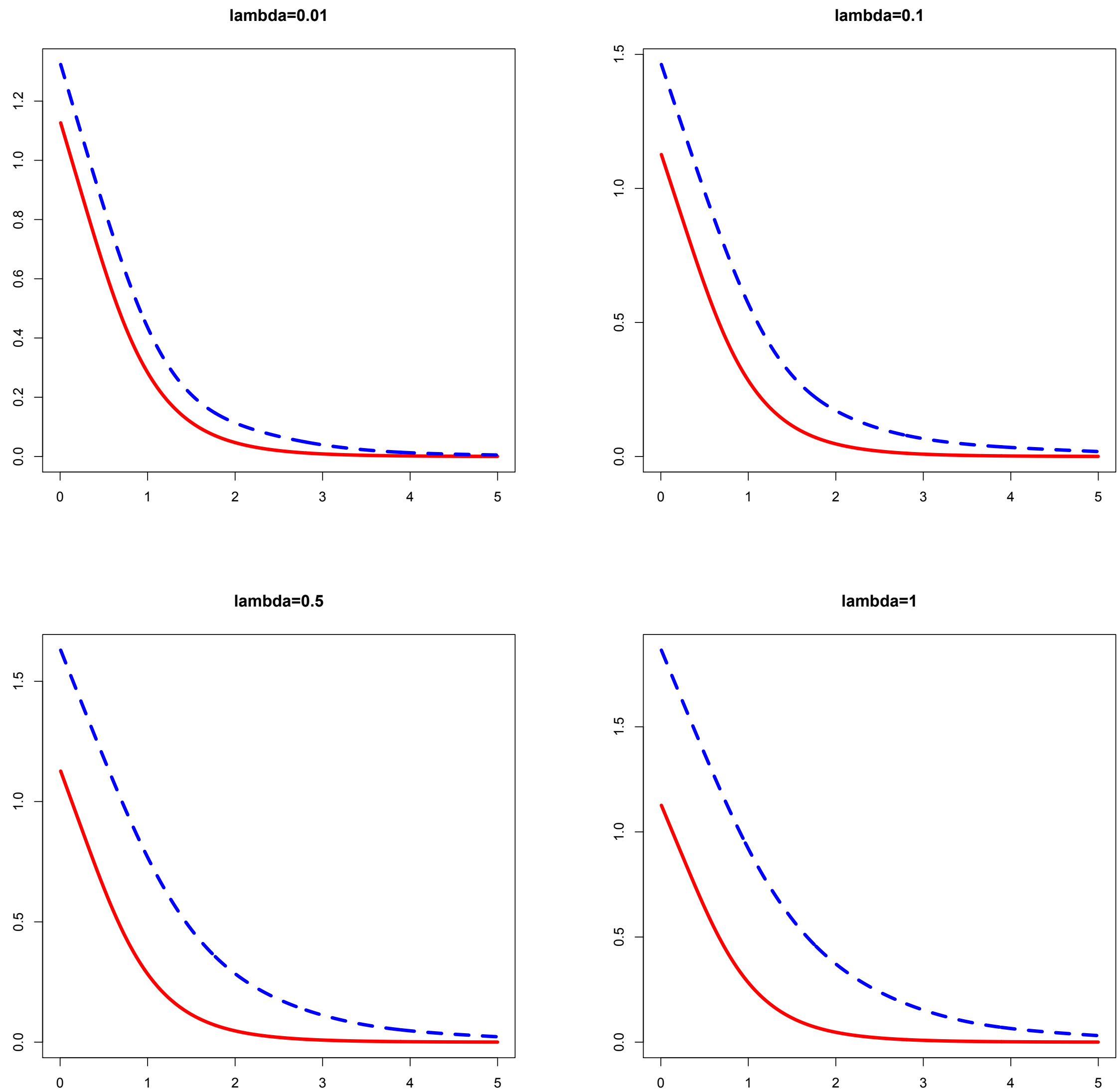

Figure 3: The plots of $\widehat{f}(z)$ (dashed line) and the true nonlinear function $f(z)$ (solid line) against z with $\lambda = 0.01, 0.1, 0.5, 1$.

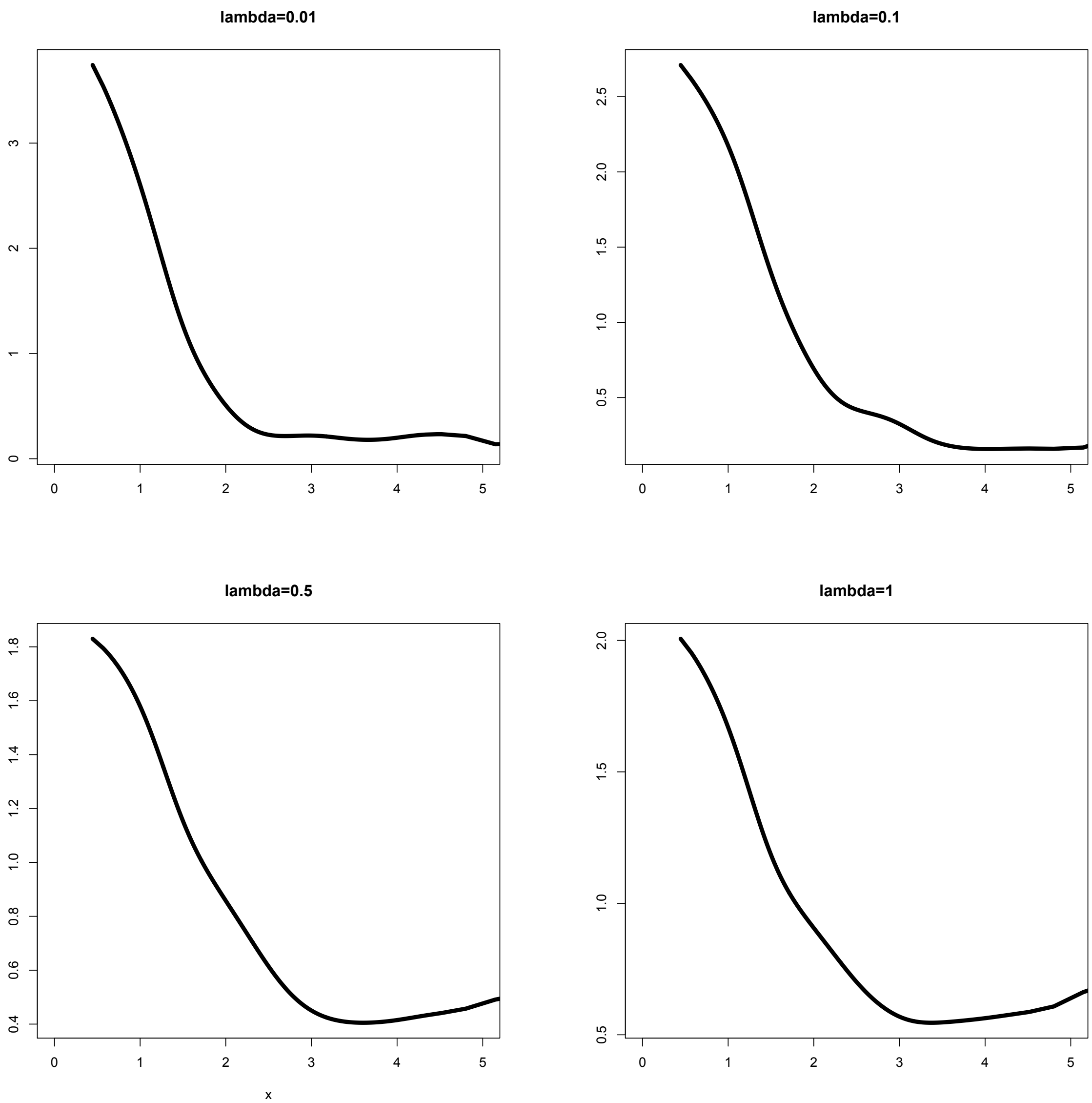

Figure 4: The plots of the EPK against x with $\lambda = 0.01, 0.1, 0.5, 1$.

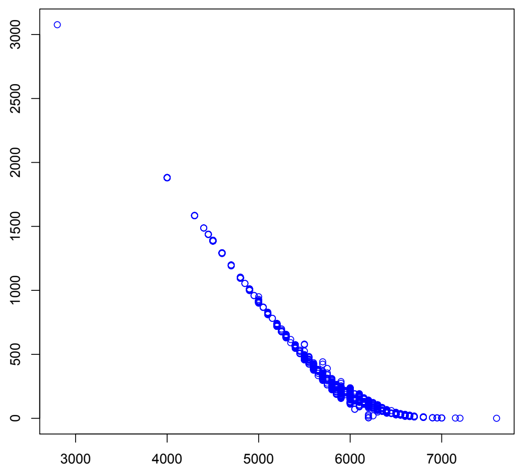

Figure 5: Plot of the call option prices against strike prices

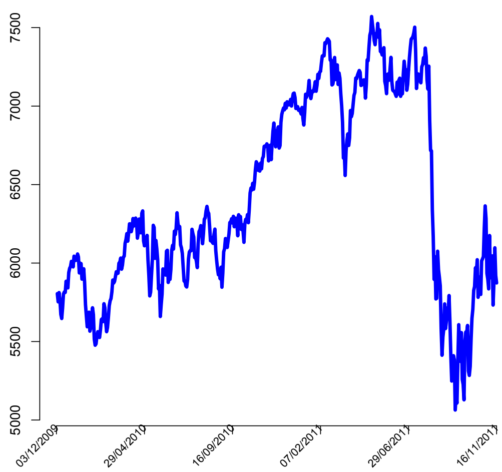

Figure 6: Plot of DAX index.

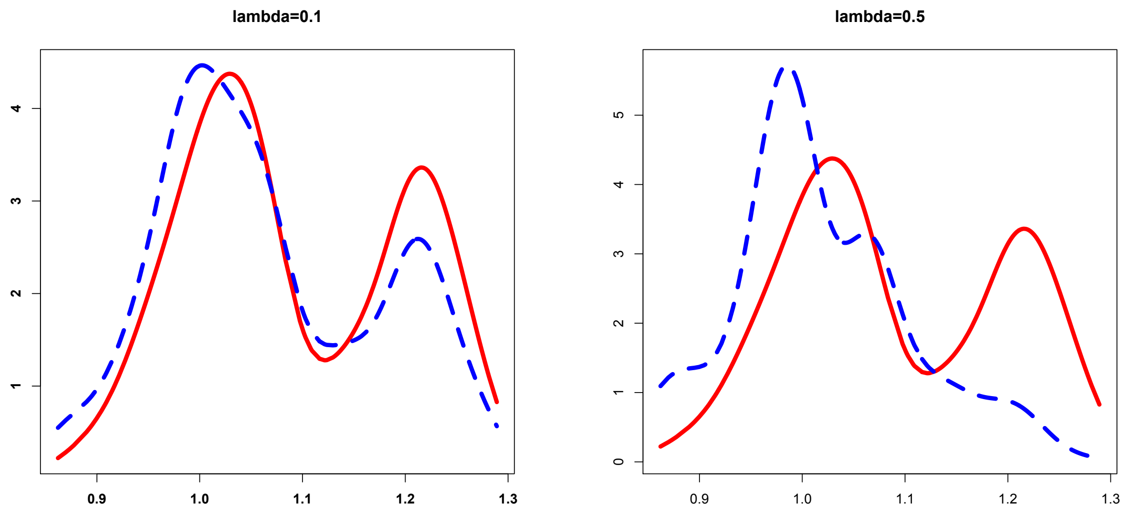

Figure 7: The plots of $q_n\left(x;\widehat{W}\right)$ (dashed line) and $p_n\left(x\right)$ (solid line) against moneyness with $\lambda=0.1,0.5$ for the Strike-Call dataset

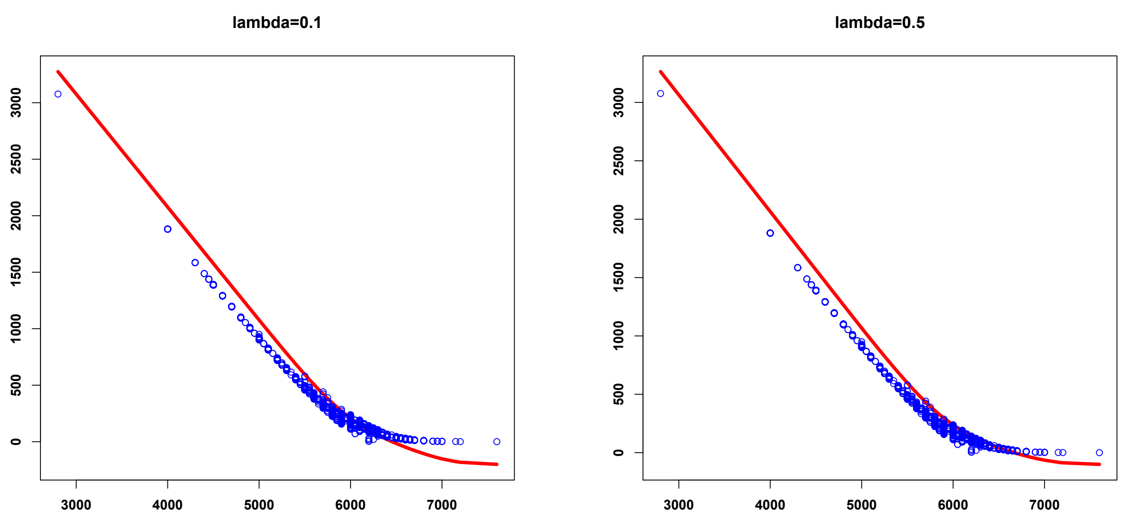

Figure 8: The plots of $\widehat{f}\left(z\right)$ and data points with $\lambda=0.1,0.5$ for the Strike-Call dataset

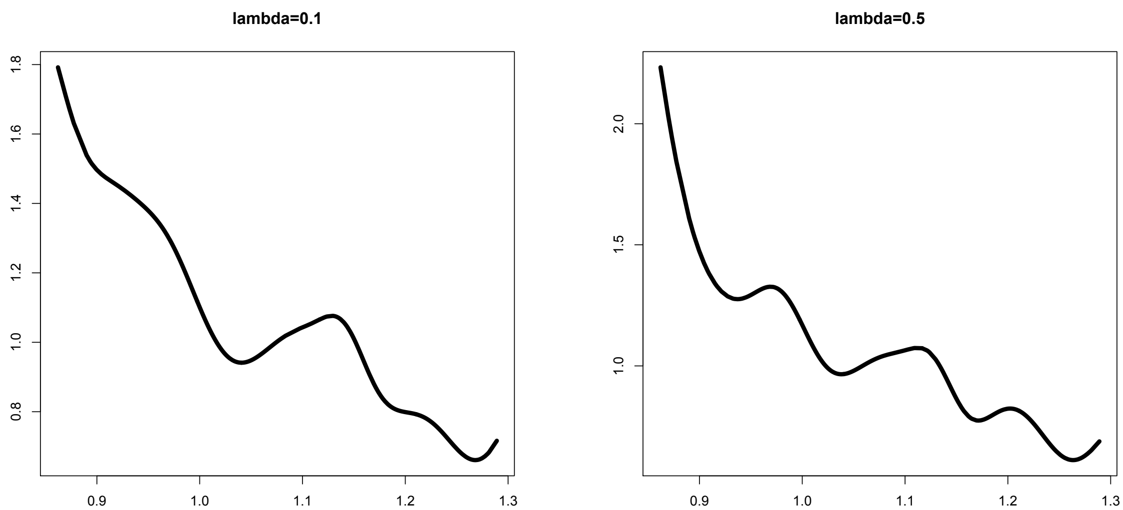

Figure 9: The plots of the empirical kernel pricing function against moneyness for the Strike-Call dataset

References

Aït-Sahalia, Y. and Lo, A. W. (2000). Nonparametric risk management and implied risk aversion. *Journal of Econometrics* 94, 9-51.

Beare, B. (2011). Measure preserving derivatives and the pricing kernel puzzle. *Journal of Mathematical Economics* 47, 689-697.

Beare, B. and Schmidt, L. (2014). An empirical test of pricing kernel monotonicity. *Journal of Applied Econometrics,* accepted, DOI: 10.1002/jae.2422.

Beare, B. and Schmidt, L. (2015). Empirical implications of the pricing kernel puzzle for the return on contingent claims. Working paper, University of California, San Diego.

Black, F. and Scholes M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy* 81, 637-659.

Breeden, D. T. and Litzenberger, R. H. (1987). Prices of state-contingent claims implicit in option prices. *Journal of Business*, 51, 621-651.

Eggermont, P. P. B. and LaRiccia, V. N. (2001). Maximum penalized likelihood estimation. Vol. I. Density estimation. Springer Series in Statistics. Springer-Verlag, New York.

Golubev, Y., Härdle, W., and Timofeev, R. (2014).Testing monotonicity of pricing kernels. *AStA Advances in Statistical Analysis*, 98, 305-326, DOI: 10.1007/s10182-014-0225-5.

Grith, M., Härdle, W., and Park, J. (2013). Shape invariant modeling of pricing kernels and risk aversion. *Journal of Financial Econometrics,* 11, 370-399, DOI: 10.1093/jjfinec/nbs019.

Grith, M., Härdle, W., and Krätschmer, V. (2015). Reference dependent preferences and the EPK puzzle. *Review of Finance,* invited resubmission.

Härdle, W., Okhrin, Y., and Wang, W. (2014). Uniform confidence bands for pricing kernels. *Journal of Financial Econometrics*, DOI: 10.1093/jjfinec/nbu002.

Renault, E. (1997). Econometric models of option pricing errors, in Advances in Economics and Econometrics, Seventh World Congress, edited by D.M. Kreps and K.F. Wallis, Econometric Society Monographs, Cambridge University Press, 223-278.

Spokoiny, V. and Zhilova, M. (2013). Sharp deviation bounds for quadratic forms. *Mathematical Methods of Statistics*, 22, 100-113.

Vapnik, V. (1995). *The Nature of Statistical Learning Theory.* Springer, New York.

Vapnik, V. (1998). *Statistical Learning Theory.* Wiley, New York.