Uniform confidence bands for pricing kernels

Wolfgang Karl Hardle, Yarema Okhrin, W. Wang

Angaben zur Veröffentlichung / Publication details:

Hardle, Wolfgang Karl, Yarema Okhrin, and W. Wang. 2015. "Uniform confidence bands for pricing kernels." *Journal of Financial Econometrics* 13 (2): 376–413. https://doi.org/10.1093/jjfinec/nbu002.

Uniform Confidence Bands for Pricing Kernels

WOLFGANG KARL HÄRDLE
Humboldt-Universität zu Berlin, CASE - Center for Applied Statistics and Economics

YAREMA OKHRIN
University of Augsburg

WEINING WANG
Humboldt-Universität zu Berlin

ABSTRACT
Pricing kernels implicit in option prices play a key role in assessing the risk aversion over equity returns. We deal with nonparametric estimation of the pricing kernel (PK) given by the ratio of the risk-neutral density estimator and the historical density (HD). The former density can be represented as the second derivative w.r.t. the European call option price function, which we estimate by nonparametric regression. HD is estimated nonparametrically too. In this framework, we develop the asymptotic distribution theory of the Empirical Pricing Kernel (EPK) in the L^∞ sense. Particularly, to evaluate the overall variation of the pricing kernel, we develop a uniform confidence band of the EPK. Furthermore, as an alternative to the asymptotic approach, we propose a bootstrap confidence band. The developed theory is helpful for testing parametric specifications of pricing kernels and has a direct extension to estimating risk aversion patterns. The established results are assessed and compared in a Monte-Carlo study. As a real application, we test risk aversion over time induced by the EPK. (*JEL*: C14, J01, J31)

KEYWORDS: empirical pricing kernel, confidence band, bootstrap, kernel smoothing, nonparametric fitting

---

The financial support from the Deutsche Forschungsgemeinschaft via SFB 649 "Ökonomisches Risiko", Humboldt-Universität zu Berlin is gratefully acknowledged. Address correspondence to Yarema Okhrin, Department of Statistics, University of Augsburg, D-86159 Augsburg, Germany, or e-mail: yarema.okhrin@wiwi.uni-augsburg.de

A challenging task in financial econometrics is to understand investors' attitudes toward market risk in its evolution over time. Such a study naturally involves stochastic discount factors, empirical pricing kernels (EPK), and state price densities, see Cochrane (2001). Asset pricing kernels (PKs) summarize investors' risk preferences and the so called "EPK paradox" exhibits when PKs are estimated from data, as several studies including Aït-Sahalia and Lo (2000), Rosenberg and Engle (2002), Brown and Jackwerth (2004) have shown. Although in all these studies the EPK paradox (nonmonotonicity) became evident, a test for the nonmonotone behavior of the pricing kernel has not been devised yet. A uniform confidence band is a very simple tool for such shape inspection. Confidence bands drawn around an EPK based on asymptotic theory and bootstrap is the subject of our study. In addition, we relate critical values of our test to changing market conditions given by exogenous time series.

The common difficulty is that the investors' preference is implicit in the goods traded in the market and thus can not be directly observed from the path of returns. A profound martingale-based pricing theory provides us one approach to tackle the problem from a probabilistic perspective. An important concept involved is the State Price Density (SPD) or Arrow-Debreu prices reflecting fair prices of one unit gain or loss for the whole market. Under no arbitrage assumption, there exists at least one SPD, and when a market is complete, this SPD is unique. Assuming a market is complete, pricing is done by taking expected payoff under the risk neutral measure, which is related to the pdf of the historical measure by multiplying with a stochastic discount factor, see Section 1 for a detailed illustration. From an economic perspective, the price is formed according to the utility maximization theory, which admits that the risk preference of consumers is connected to the shape of utility functions. Specifically, a concave, convex, or linear utility function describes the risk averse, risk seeking, or risk neutral behavior. Importantly, a stochastic discount factor can be expressed via a utility function (Marginal Rate of Substitution), which links the shape of pricing kernel to the risk patterns of investors, see Kahneman and Tversky (1979), Jackwerth (2000), Rosenberg and Engle (2002) and others.

The above mentioned theory allows us to relate price processes of assets to risk preference of investors. This amounts to fitting a flexible model and making inference on the dynamics of EPKs over time. A well-known but restrictive approach is to assume the underlying following a geometric Brownian motion. In this setting, SPDs and HDs are log normal distributions with different drifts, and the parametrization of PK coincides with the conditional expectation of marginal utilities when assuming a power utility function. Thus it is decreasing in return and implies overall risk-averse behavior. However, across different markets, one observes quite often a nondecreasing pattern for EPKs, a phenomenon called the EPK paradox, see Chabi-Yo, Garcia, and Renault (2008), Christoffersen, Heston, and Jacobs (2011).

Two plots of pricing kernels are shown in Figures 1 and 2. Figure 1 depicts inter-temporal pricing kernels with fixed maturity, while Figure 2 depicts pricing

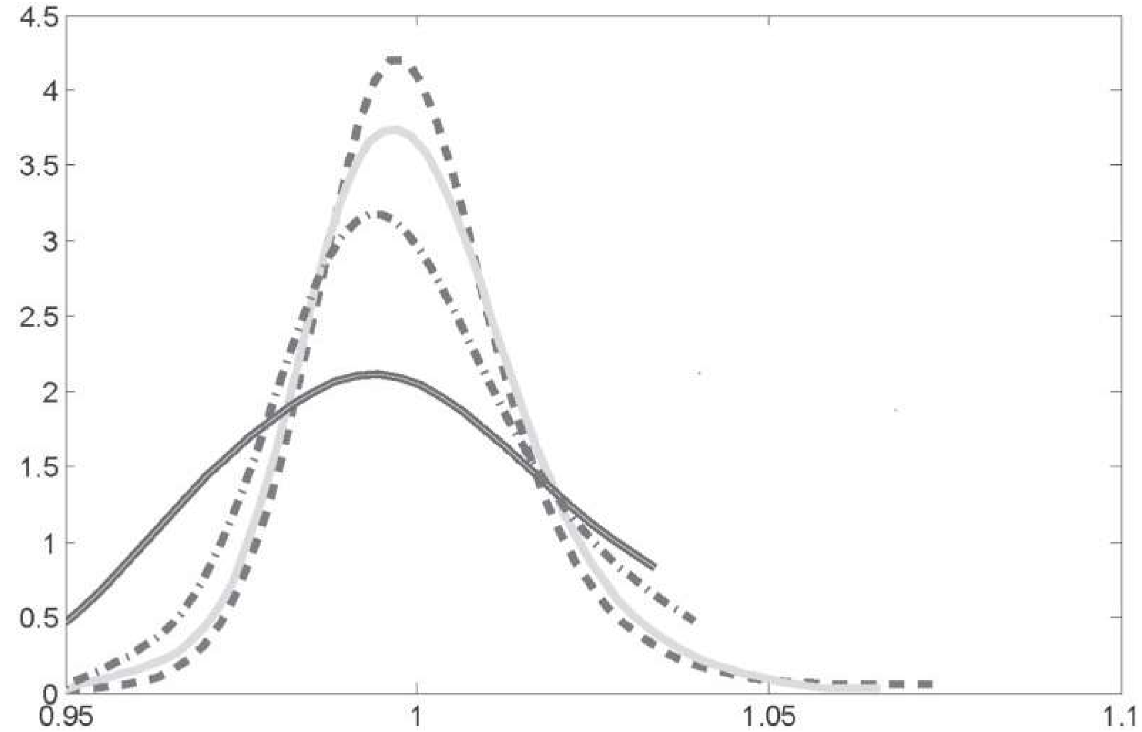

Figure 1 Examples of estimated inter-temporal pricing kernels (as functions of moneyness) with fixed maturity 0.00833 (3 days) in years respectively on January 17, 2006 (dashed), April 18, 2006 (light gray), May 16, 2006 (dark gray), June 13, 2006 (dash-dotted), see Grith, Härdle, and Park (2013).

kernels with two different maturities and their confidence bands. The figures are shown on return scale. The curves present a bump in the middle and a switch from convexity to concavity in all cases. Especially, this shows that very unlikely the bands contain a monotone decreasing curve.

Besides the shape of the confidence bands, the time varying coverage probability of a uniform confidence band has implications on risk attitudes of investors. At a fixed point in time, it helps us to test against alternatives for a PK and thus yields insights into time varying risk patterns. The extracted time varying parameter, realized either from a low-dimensional model for PKs or given by the coverage probability, may thus be economically analyzed in connection with exogenous macroeconomic business cycle indicators, e.g., credit spread, yield curve, etc., see also Grith, Härdle, and Park (2013).

To our knowledge, there are no comparable approaches developed for uniform testing of the shape of EPK or of any continuous transformations of it. Golubev, Härdle, and Timofeev (2014) suggest a test of monotonicity of the PK, while other literature on testing PKs, for example, Jagannathan and Wang (1996), Wang (2002), Wang (2003), serves different purposes like verifying the significance of pricing errors. In contrary to these papers we do not address the issue of mispricing, but provide a solid statistical tool to testing the validity of any parametric shape of the PK.

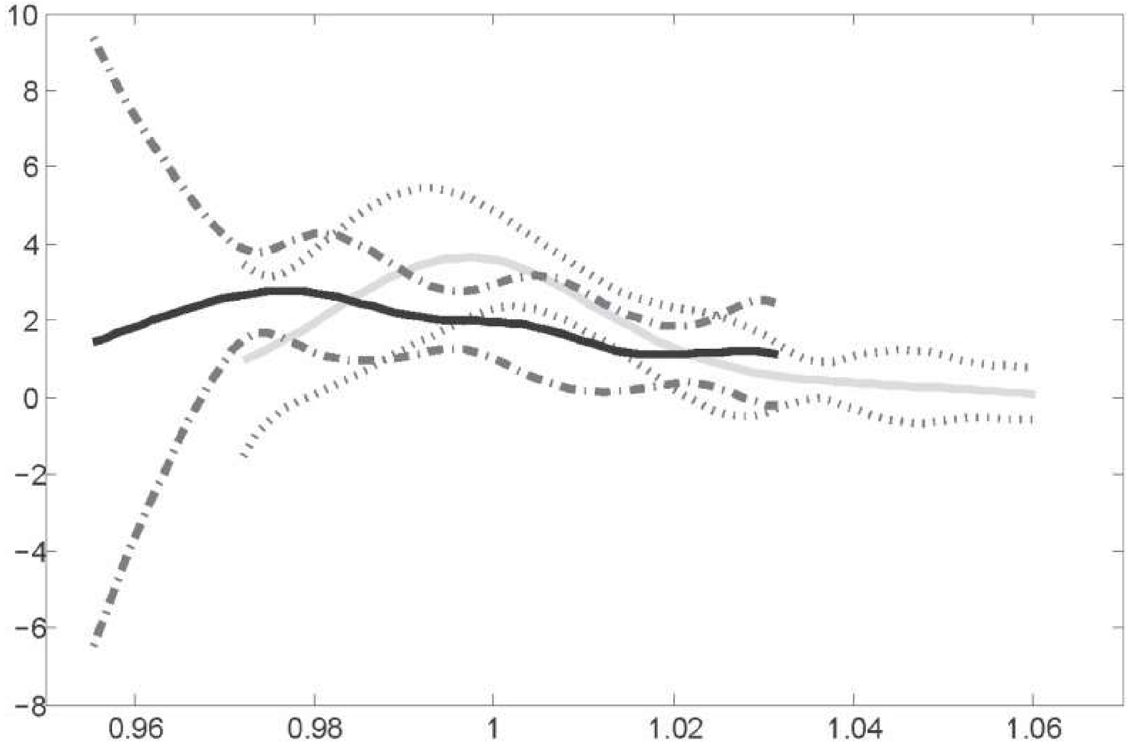

Figure 2 Examples of estimated inter-temporal pricing kernels with various maturities in years: 0.02222 (8 days, gray) 0.1 (36 days, black) on January 12, 2006 and their confidence bands.

Several econometric studies are concerned with estimating PKs by estimating a SPD and HD separately. See Section 1 for details. It is stressed in Aït-Sahalia and Lo (1998) that nonparametric inference from pricing kernels gives unbiased insights into the properties of asset markets. The stochastic fluctuation of EPK as measured by the maximum deviation has not been studied yet. Nevertheless, the asymptotic distribution of the maximum deviation and the uniform confidence band linked to it are very useful for model check.

Uniform confidence bands for smooth curves have first been developed for kernel density estimators by Bickel and Rosenblatt (1973), extension to regression smoothing can be found in Liero (1982) and Härdle (1989). But only recently, the results have been carried over to derivative smoothing by Claeskens and Van Keilegom (2003). Our theoretical path follows largely their results, but our results are applied to a ratio estimator instead of a local polynomial estimator. Also we have a realistic data situation that relates coverage to economic indicators. In addition we perform the smoothing in an implied volatility space which brings by itself an interesting modification of the results of that paper.

This article is organized as follows: In Section 1, we describe the theoretical connection between utility functions and pricing kernels. In Section 2, we present a nonparametric framework for the estimation of both the HD and the SPD and derive the asymptotic distribution of the maximum deviation. In Section 3, we simulate the asymptotic behavior of the uniform confidence band and compare it with the bootstrap method. Moreover, we also compare the results with other parametric estimation procedures. In Section 4, we conclude and discuss our results.

1 EMPIRICAL PRICING KERNEL ESTIMATION

Consider an arbitrary risky financial security with the price process $\{S_t\}_{t\in[0,T]}$. The interest rate process r is deterministic. We assume that the market is complete, so there exists a unique risk neutral measure. By the change-of-measure argument the price at time t for the nonnegative payoff $\psi(S_T)$ is

$$P_t \stackrel{\text{def}}{=} \mathsf{E}^{\mathsf{Q}}\left[e^{-r\tau}\psi(S_T)\right] = \mathsf{E}\left[e^{-r\tau}\psi(S_T)\mathcal{K}(S_T)\right], \tag{1}$$

where $\mathcal{K}(S_T)$ is defined as *the pricing kernel* or *stochastic discount factor* at time t, E is the expectation under the historical measure $\mathbb{P}_{S_T|S_t}(x)$ and E^{Q} is the expectation under the risk neutral measure $\mathbb{Q}_{S_T|S_t}(x)$, τ is the time to maturity. Thus the price of the security at time point t equals the expected net present value of its future payoffs, computed with respect to the risk-neutral measure. More explicitly,

$$P_t = \int_0^{+\infty} e^{-r\tau}\psi(x)q(x)dx = \int_0^{+\infty} e^{-r\tau}\psi(x)\mathcal{K}(x)p(x)dx, \tag{2}$$

where $p(x)$ and $q(x)$ are the pdf of the historical measure and the risk neutral density or state price density (SPD) of S_T, respectively. Note that $p(x)$ and $q(x)$ are conditional on the current price S_t and potentially may depend on other parameters as discussed below. We skip the indication of conditioning to keep the notation simple. Thus all expectations hereafter are conditional on S_t if not stated otherwise.

It follows from (2), that $\mathcal{K} = \frac{q}{p}$ and both the pdf of the future payoff and the SPD are required to compute the pricing kernel. Several approaches are available to determine the EPK explicitly. First, we can impose strict parametric restrictions on the dynamics of the asset prices and on the distribution of the future payoff. Mixture normal distributions are an example, see Jackwerth (2000). In the case of more complex stochastic processes, usually no explicit solution is available. A possible technique though is to use the Brownian motion setup as a prior model. Subsequently the SPD is estimated by minimizing the distance to the prior SPD subject to the constraints characterizing the underlying securities, see Rubinstein (1994) and Jackwerth and Rubinstein (1996).

Another important perspective of specifying PK is done via the utility function in the consumption based pricing model, see Heaton and Lucas (1992).

Let the aim of the investor be to solve the problem:

$$\max_{W_t}\{u(W_t) + \mathsf{E}[\beta u(W_T)]\},$$

where $u(\cdot)$ denotes the utility function, W_t the wealth and β the subjective discount factor. The current price of an asset is

$$P_t = \mathrm{E}[\beta_t \psi(S_T)], \tag{3}$$

where β_t is the stochastic discount factor and it equals the inter-temporal marginal rate of substitution. If both ways of pricing in (1) and (3) are admissible, then they lead to the same price in a complete market. Then the stochastic discount factor is given by $\beta \frac{u'(s)}{u'(S_t)}$ and is proportional to the PK. This implies that by fixing the utility of the investor we can determine the PK, which is related to the standard risk aversion measures. In practice, however, usually the opposite procedure is applied. The PK is estimated via a ratio of $\hat{q}$ and $\hat{p}$ and used to determine the utility function or the risk aversion coefficient of the investor. Assessment of the temporal dynamics of the latter allows for inferences on the market risk behavior.

1.1 EPK and Option Pricing

EPK is calibrated from the data via an estimation of the ratio of the SPD q and the HD p respectively. In this section we describe the details for this calibration. The latter can easily be estimated either parametrically or nonparametrically from the time series of payoffs. On the contrary, the SPD depends on risk preferences and therefore the past observed stock price time series do not contain enough information. Option prices do reflect preferences and, therefore, can be used to estimate the SPD q. Let $C(S_t, X, \tau, r, \sigma^2)$ denote the European call-option price as a function of the strike price X, price S_t, maturity τ, interest rate r and volatility σ. Following Breeden and Litzenberger (1978) the SPD can be determined from the pricing equation by

$$q(S_T) = e^{r\tau} \frac{\partial^2 C}{\partial X^2}\bigg|_{X=S_T}. \tag{4}$$

This result is very general and is valid for all European call options with the payoff function $(S_T - X)_+$ and with the single assumption that the price is twice differentiable. No additional restrictions on the stochastic process for the underlying or on the preferences of market participants are needed. In a Black–Scholes (BS) framework, where the underlying asset price S_t follows a geometric Brownian motion, the European options are priced via:

$$C(S_t, X, \tau, r, \sigma^2) = S_t \Phi(d_1) - X e^{r\tau} \Phi(d_2),$$

where d_1 and d_2 are known functions of σ^2, τ, X, and S_t. This implies that both $q(S_T)$ and $p(S_T)$ are the densities of lognormal distributions:

$$q(S_T) = \frac{1}{S_T\sqrt{2\pi\sigma^2\tau}} \exp\left[-\frac{\{\log(S_T/S_t) - (r - \sigma^2/2)\tau\}^2}{2\sigma^2\tau}\right] \tag{5}$$

and $p(S_T)$ with μ replacing r in (5).

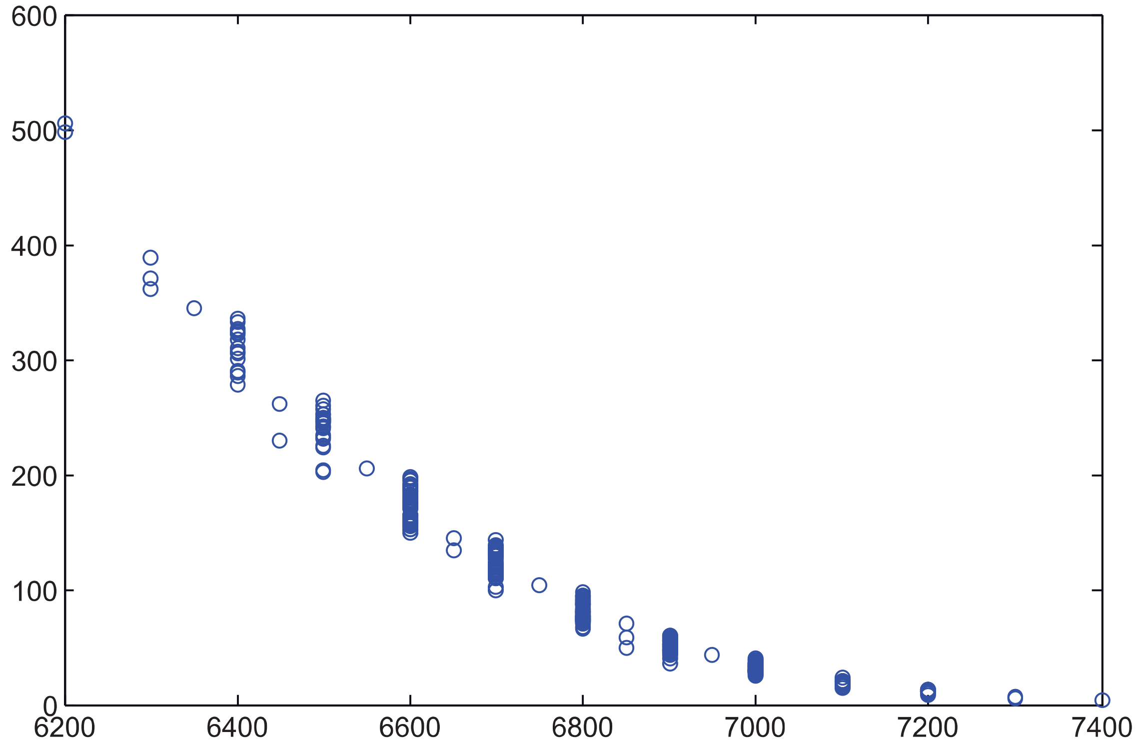

Figure 3 Plot of call option prices against strike prices on January 17, 2001

The restrictive parametric form of BS model may often induce modeling bias when fitted to the data, especially it is not possible to reflect the implicit volatility smile (surface) as a function of X and τ via (5), see Renault (1997). The latter may be derived in a stochastic volatility model, such as those of Heston or Bates type. In order to study unbiased risk patterns, we need to guarantee models for the pricing kernel that are rich enough to reflect local risk aversion in time and space. This leads naturally to a smoothing approach.

We now describe how to estimate $q(\cdot)$ nonparametrically. Consider call options with maturity τ. We consider the following heteroscedastic nonparametric model for the observed option prices Y_i

$$Y_i = C_\tau(X_i) + \sigma(X_i)\varepsilon_i, \quad i = 1, \dots, n_q, \tag{6}$$

where Y_i denotes the observed option price, X_i the strike price and $C_\tau(\cdot)$ is a smooth function of the strike price. For simplicity of notation, we write $C(\cdot)$ for $C_\tau(\cdot)$. The informational content of the model is similar to that of Yuan (2009). We argue that the perceived errors are due to neglected heterogeneity factors, rather than mispricings exploited by arbitrage strategies, see Renault (1997). Thus the pricing errors ε_i are assumed to be i.i.d. in the cross section. Figure 3 depicts the call option prices used to calculate a SPD. The observations are distributed, with different variances, at discrete grid points of strike prices.

As from (4), estimation of $q(\cdot)$ boils down to the estimation of the second derivative of $C(\cdot)$. The following local polynomial approach allows us to estimate

$C(\cdot)$ and the derivatives of $C(\cdot)$ simultaneously. Assuming that $C(X)$ is continuously differentiable of order $d=3$, it can be locally approximated by

$$C(X, X_0) = \sum_{j=0}^{d} C_j(X_0)(X_0 - X)^j, \tag{7}$$

where $C_j(X) = C^{(j)}(X)/j!$, $j = 0, \dots, d$. See Cleveland (1979), Fan (1992), Fan (1993), Ruppert and Wand (1994) for more details. By assuming a local Gaussian quasi-likelihood model with $\mathcal{L}\{Y, C(X)\} = \{Y - C(X)\}^2/\{2\sigma^2(X)\}$ and maximizing the local likelihood, the function $C(\cdot)$ can be approximated by $\hat{C}(X, X_0)$ with:

$$\hat{\boldsymbol{C}}(X) = \underset{\boldsymbol{C}(X)}{\arg\max} \frac{1}{n_q} \sum_{i=1}^{n_q} w_i K_{h_{n_q}}(X_i - X) \mathcal{L}\{Y_i, C(X, X_i)\}, \tag{8}$$

where $\boldsymbol{C}(X) = \{C_0(X), C_1(X), \dots, C_d(X)\}^\top$, $K_{h_{n_q}}(X_i - X) = K\{(X_i - X)/h_{n_q}\}/h_{n_q}$ is a kernel function with bounded support and a bandwidth sequence h_{n_q}. Following Aït-Sahalia and Duarte (2003) and Yuan (2009) the weights w_i may reflect the relative liquidity of different options, putting more weight on more heavily traded options. To simplify the exposition we assume $w_i = 1$ for all i. The above maximum likelihood approach is equivalent to a minimum weighted square loss approach, since maximizing the local Gaussian likelihood function leads numerically to the same solution as minimizing the weighted least squares. The advantage of the likelihood framework is that it can be easily adapted to non-Gaussian distributions and to heteroscedastic pricing equations.

Solving the above optimization problem in (8) is equivalent to solving:

$$\mathbf{A}_{n_q}(X) \stackrel{\text{def}}{=} \frac{1}{n_q} \sum_{i=1}^{n_q} K_{h_{n_q}}(X_i - X) \frac{\partial Q\{Y_i, C(X, X_i)\}}{\partial C} \mathbf{X}_i = \mathbf{0}, \tag{9}$$

with $\mathbf{X}_i \stackrel{\text{def}}{=} (1, X_i - X, (X_i - X)^2, (X_i - X)^3)^\top$. We are concerned with $2!\hat{C}_2(x) = \frac{\partial^2 C(X)}{\partial X^2}\big|_{X=x}$, which is shown by Breeden and Litzenberger (1978) to be proportional to $q(x)$. Note that the described procedure does not guarantee the feasibility of the estimator as a density. The constrained estimator of Aït-Sahalia and Duarte (2003) makes the large sample results below invalid. Therefore, we rely on the consistency and asymptotic validity of $\hat{q}$ as a density estimator. This approach is justified by large samples available in financial applications. A multiplicative renormalizing of the estimator will shift the EPK curve and the corresponding confidence bands, while keeping the results of the monotonicity test unchanged. Furthermore, the renormalization introduces a bias, which is difficult to tackle analytically. Additional improvement of the estimator is elaborated in Section 2.1.

Note that we assume the parameter $C(\cdot)$ and $\sigma(\cdot)$ to be orthogonal to each other. Thus we can estimate them separately as in a single parameter case. The following lemma states the results on the existence of the solution and its consistency.

Lemma 1: *Under conditions* $(A1)-(A7)$, *there exists a sequence of solutions to the equations*

$$\mathbf{A}_{n_q}(x)=\mathbf{0},$$

with x being an element of a compact set E, such that

$$\sup_{x\in E}|\hat{q}(x)-q(x)|=\mathcal{O}\big[h_{n_q}^{-2}\{\log n_q/(n_q h_{n_q})\}^{1/2}+h_{n_q}^2\big] \quad a.s.$$

Proof. The statement follows from Theorem 2.1 of Claeskens and Van Keilegom (2003). ■

The HD $p(x)$ can be estimated separately from the SPD using historical prices $S_t,\dots,S_{t+n_p+\tau-1}$ $(t+n_p+\tau-1<T)$ of the underlying asset. The nonparametric kernel estimator of $p(x)$ is given similarly to Aït-Sahalia and Lo (2000) by

$$\hat{p}_{return}(x)=n_p^{-1}\sum_{j=0}^{n_p-1}K_{h_{n_p}}\left\{x-\log(S_{t+j+\tau}/S_{t+j})\right\},$$

where $K_{h_{n_p}}$ is a kernel function with bounded support and the bandwidth h_{n_p}. This kernel should necessarily coincide with the kernel for estimating SPD $q(\cdot)$. Also as in Aït-Sahalia and Lo (2000), the density of log-returns can be estimated as:

$$\hat{p}_{return}(x)=S_t\exp(x)\hat{p}\{S_t\exp(x)\}.$$

Alternatively, to eliminate the impact of serial dependence of overlapping returns over τ periods, we can simulate independent pathes of the price process and use it to estimate the density of S_T, then n_p will become the number of paths simulated for the time T. Under assumption (A5), we know that

$$\sup_{x\in E}|\hat{p}(x)-p(x)|=\mathcal{O}\{(n_p h_{n_p}/\log n_p)^{-1/2}+h_{n_p}^2\} \quad a.s. \tag{10}$$

Remark The uniform convergence results for estimation of HD in the i.i.d case follows from Bickel and Rosenblatt (1973), and recently extended by Liu and Wu (2010) (Theorem 2.3) to the serial dependent data case.

The EPK is then given by the ratio of the estimated SPD and the HD $p(x)$ i.e. $\hat{\mathcal{K}}(x)=\hat{q}(x)/\hat{p}(x)$. The next lemma provides the linearization of the ratio, which is important for further statements about the uniform confidence band of the EPK.

Lemma 2: *Under conditions* $(A1)-(A7)$ *it holds*

$$\begin{aligned}&\sup_{x\in E}|\hat{\mathcal{K}}(x)-\mathcal{K}(x)|\\ =&\sup_{x\in E}\Big|\frac{\hat{q}(x)-q(x)}{p(x)}-\frac{\hat{p}(x)-p(x)}{p(x)}\cdot\frac{q(x)}{p(x)}-\frac{\{\hat{q}(x)-q(x)\}\{\hat{p}(x)-p(x)\}}{p^2(x)}\Big|\\ &+\mathcal{O}[\max\{(n_p h_{n_p}/\log n_p)^{-1/2}+h_{n_p}^2,h_{n_q}^{-2}\{n_q h_{n_q}/\log n_q\}^{-1/2}+h_{n_q}^2\}] \quad a.s. \end{aligned}\tag{11}$$

This lemma implies that the stochastic deviation of $\widehat{\mathcal{K}}$ from $\mathcal{K}$ can be linearized into a stochastic part containing the estimator of the SPD and a deterministic part containing $\mathrm{E}[\hat{p}(x)]$. The uniform convergence can be proved by dealing separately with the two parts. The convergence of the deterministic part is shown by imposing mild smoothness conditions, while the convergence of the stochastic part is proved by following the approach of Claeskens and Van Keilegom (2003). Theorem 1 formalizes this uniform convergence of the EPK.

Theorem 1: *Under conditions* $(A1)-(A7)$, *it holds*

$$\sup_{x\in E}|\hat{\mathcal{K}}(x)-\mathcal{K}(x)|=\mathcal{O}[\max\{(n_p h_{n_p}/\log n_p)^{-1/2}+h_{n_p}^2, h_{n_q}^{-2}\{n_q h_{n_q}/\log n_q\}^{-1/2}+h_{n_q}^2\}] \quad a.s.$$

The proof is given in the Appendix. The theoretical optimal rate of bandwidth follows by minimizing the bias and variance term together in Theorem 1 leading to $(n_q \log n_q)^{-1/9}$. In our simulation and applications, we use the bandwidth sequence which minimizes coverage error by performing a grid search, see Claeskens and Van Keilegom (2003).

2 CONFIDENCE INTERVALS AND CONFIDENCE BANDS

Confidence intervals characterize the precision of the EPK for a given fixed value of the payoff. This allows to make inference about the PK at each particular strike price, but does not allow conclusions about the global shape. The confidence bands, however, characterize the whole EPK curve and offer therefore the possibility to test for shape characteristics. In particular, it is a way to check the persistence of the bump as observed. Given a certain shape, one may verify the restriction imposed by the power utility and obtain insights on the market risk aversion. In addition, the confidence bands can be used to measure the global variability of the EPK. Also, the proportion of the BS-based EPK covered in nonparametric bands can be used as a measure of global risk aversion. The global variability is measured by the variance function of EPK and the BS-based EPK means the parametric fitting achieved by assuming that the underlying follows the geometric Brownian motion.

A confidence interval for the EPK at a fixed value x requires the asymptotic distribution of $\hat{p}(x)$ and $\hat{q}(x)$. Hereafter, we use $\mathcal{L}$ to denote the convergence in law. Under $(A1)-(A7)$:

$$\sqrt{n_p h_{n_p}}\{\hat{p}(x)-p(x)\} \xrightarrow{\mathcal{L}} \mathrm{N}\{0, p(x)\int K^2(u)du\} \tag{12}$$

and

$$\sqrt{n_q h_{n_q}^5}\{\hat{q}(x)-q(x)\} \xrightarrow{\mathcal{L}} \mathrm{N}\{0, \sigma_q^2(x)\}, \tag{13}$$

where $\sigma_q^2(x)=[B(x)^{-1}\boldsymbol{L}^{-1}\boldsymbol{T}\boldsymbol{L}^{-1}]_{(3,3)}$, with $B(x)$ equal to the product of the density $f_X(x)$ of the strike price and the local Fisher information matrix $I\{C(x)\}$. The matrices

$\mathbf{L}$ and $\mathbf{T}$ are given by $\mathbf{L} \stackrel{\text{def}}{=} [\int u^{i+j}K(u)du]_{i,j}$ and $\mathbf{T} \stackrel{\text{def}}{=} [\int u^{i+j}K^2(u)du]_{i,j}$ with $i,j = 0,\ldots,3$. This implies the asymptotic normality of the EPK at a fixed payoff x. More precisely

$$\sqrt{n_q h_q^5}\{\hat{\mathcal{K}}(x) - \mathcal{K}(x)\} \xrightarrow{\mathcal{L}} \mathrm{N}\{0, \sigma_q^2(x)/p^2(x)\}. \tag{14}$$

The variance of $\hat{\mathcal{K}} = \hat{\mathcal{K}}(x)$ is given by

$$\mathrm{Var}\{\hat{\mathcal{K}}(x)\} \approx \{p(x)\}^{-2} B^{-1}(x)\mathbf{L}^{-1}\mathbf{T}\mathbf{L}^{-1}. \tag{15}$$

The above results on the limiting distribution of $\hat{p}$, $\hat{q}$, and $\hat{\mathcal{K}}$ can directly used to establish confidence intervals and the Bonferroni-type confidence bands for the considered densities. This approach, as argued by Eubank and Speckman (1993), are asymptotically conservative even though they do not explicitly account for potential bias. These bands are taken as benchmarks for comparison purposes in the simulation study.

Let $\hat{\mathcal{D}}_{n_q}(x)$ be the standardized process with the estimated variance of the EPK:

$$\hat{\mathcal{D}}_{n_q}(x) \stackrel{\text{def}}{=} n_q^{1/2} h_{n_q}{}^{5/2}\{\hat{\mathcal{K}}(x) - \mathcal{K}(x)\} \,/\, [\widehat{\mathrm{Var}}\{\hat{\mathcal{K}}(x)\}]^{1/2}.$$

Relying on the linearization in Lemma 2, we derive the confidence band for $\mathcal{K}$.

Theorem 2: *Under assumptions (A1)-(A5) it follows*

$$\mathrm{P}\left[(-2\log h_{n_q})^{1/2}\left\{\sup_{x\in E}|\hat{\mathcal{D}}_{n_q}(x)| - c_{n_q}\right\} < z\right] \longrightarrow \exp\{-2\exp(-z)\},$$

where $c_{n_q} = (-2\log h_{n_q})^{1/2} + (-2\log h_{n_q})^{-1/2}\{x_\alpha + \log(R/2\pi)\}$.

The $(1-\alpha)100\%$ confidence band for the pricing kernel $\mathcal{K}$ is thus:

$$[f : \sup_{x\in E}\{|\hat{\mathcal{K}}(x) - f(x)| \,/\, \widehat{\mathrm{Var}}(\hat{\mathcal{K}})^{1/2}\} \le L_\alpha],$$

where $L_\alpha \stackrel{\text{def}}{=} 2!(n_q h_{n_q}^5)^{-1/2} c_{n_q}$, $x_\alpha = -\log\{-1/2\log(1-\alpha)\}$ and
$R \stackrel{\text{def}}{=} (\boldsymbol{L}^{-1}\boldsymbol{P}\boldsymbol{L}^{-1})_{3,3}/(\boldsymbol{L}^{-1}\boldsymbol{T}\boldsymbol{L}^{-1})_{3,3}$
with $\boldsymbol{P} \stackrel{\text{def}}{=} [\int u^{i+j}K'(u)\}^2 du - \frac{1}{2}\{i(i-1) + j(j-1)\}\int u^{i+j-2}K^2(u)du]_{i,j=0,\ldots,3}$.

For the implementation with real data we need a consistent estimator of $\mathrm{Var}(\hat{\mathcal{K}})$. For fixed τ, we rely on the delta method and use the empirical sandwich estimator, see Carroll, Ruppert, and Welsh (1998). The latter method provides the variance estimator for the parameters obtained from estimating equations given by (9).

To estimate the variance function of the EPK we consider time series of the option prices and the corresponding strike prices, (X_{it}, Y_{it}), $i = 1,\ldots,n_q$;

$t = t+1, \ldots, t+\tau$, we have

$$\widehat{\mathrm{Var}}\{\hat{\mathcal{K}}(x)\} = \{\hat{p}(x)\}^{-2} \boldsymbol{V}(x)^{-1} \boldsymbol{U}(x) \boldsymbol{V}(x)^{-1}, \tag{16}$$

where

$$\boldsymbol{V}(x) \stackrel{\text{def}}{=} \frac{1}{n_q \tau} \sum_{i=1}^{n_q} \sum_{j=t+1}^{t+\tau} K^2_{h_{n_q}}(X_{ij} - x) \left[\frac{\partial}{\partial C} Q\{Y_{ij}; \hat{C}(x, X_{ij})\} \right]^2 (\boldsymbol{H}^{-1}_{n_q} \mathbf{X}_{ij})(\boldsymbol{H}^{-1}_{n_q} \mathbf{X}_{ij})^\top, \tag{17}$$

$$\boldsymbol{U}(x) \stackrel{\text{def}}{=} \frac{1}{n_q \tau} \sum_{i=1}^{n_q} \sum_{j=t+1}^{t+\tau} K^2_{h_{n_q}}(X_{ij} - x) \left[\frac{\partial^2}{\partial^2 C} Q\{Y_{ij}; \hat{C}(x, X_{ij})\} \right] (\boldsymbol{H}^{-1}_{n_q} \mathbf{X}_{ij})(\boldsymbol{H}^{-1}_{n_q} \mathbf{X}_{ij})^\top, \tag{18}$$

where $\mathbf{X}_{ij} \stackrel{\text{def}}{=} (1, \ldots, (X_{ij} - x)^3)^\top$ and $\boldsymbol{H}_{n_q} \stackrel{\text{def}}{=} \mathrm{diag}\{1, \ldots, h^3_{n_q}\}$. The estimator is consistent in our setup as motivated in Appendix A.2 of Carroll, Ruppert, and Welsh (1998).

It is important to note that the nonparametric estimators are biased, which leads to potentially wrongly centered confidence bands and misleading coverages.

To overcome this problem we deploy the bias-correcting technique of Xia (1998), which is based on the local polynomial estimation. It is used to correct the bias in estimated SPD, while the bias in the HD is corrected using the additive bias correction method mentioned in Jones, Linton, and Nielson (1995). In the next step we correct the bias in the EPK using the linearization in Lemma 2. The estimated leading term bias for EPK consists of the estimated bias of $\hat{q}(x)$ and of $\hat{p}(x)$ with a bigger bandwidth than what used in estimation. This is the oversmoothing idea proposed by Eubank and Speckman (1993).

2.1 Bootstrap Confidence Bands

In this subsection, we discuss a bootstrap version of the confidence band to obtain possibly better finite sample performance. The slow rate of convergence is known to us by Hall (1991), who showed that for density estimators, the supremum of $\{\widehat{q}(x) - q(x)\}$ converges at the slow rate $(\log n_q)^{-1}$ to the Gumbel extreme value distribution. Therefore the confidence band may exhibit poor performance in finite samples. An alternative approach is to use the bootstrap method. Claeskens and Van Keilegom (2003) used smooth bootstrap for the numerical approximation to the critical value. Here we consider the bootstrap technique of the leading term in Lemma 2

$$\sup_{x \in E} \left| \frac{\widehat{q}(x) - q(x)}{p(x)} \right|.$$

We resample data from the smoothed bivariate distribution of (X, Y) with the density estimator given by estimator is:

$$\hat{f}(x, y) = \frac{\hat{\sigma}_X}{n_q h_{n_q} h_{n_q} \hat{\sigma}_Y} \sum_{i=1}^{n_q} K\left\{ \frac{X_i - x}{h_{n_q}}, \frac{(Y_i - y)\hat{\sigma}_X}{h_{n_q} \hat{\sigma}_Y} \right\},$$

where $\hat{\sigma}_X$ and $\hat{\sigma}_Y$ are the estimated standard deviations of the distributions of X and Y. The motivation of using the smooth bootstrap procedure is that the Rosenblatt transformation requires the resampled data (X^*, Y^*) to be continuously distributed.

From the re-sampled data sets, we calculate the bootstrap analogue of the leading term in Lemma 2:

$$\sup_{x\in E} |\frac{\widehat{q}^*(x) - \widehat{q}(x)}{\widehat{p}(x)}|.$$

One may argue that this resampling technique does not correctly reflect the bias arising in estimating q. Therefore, Härdle and Marron (1991) use therefore a resampling procedure based on a larger bandwidth. This refined bias correcting bootstrap method does not need to be applied in our case, since our bandwidth conditions ensure a negligible bias.

Correspondingly, we define the one-step estimator for the stochastic deviation:

$$h_{n_q}^2 \{\widehat{\mathcal{K}}^*(x) - \widehat{\mathcal{K}}(x)\} = -\{p(x)\}^{-2}\{\boldsymbol{U}^*(x)^{-1}\boldsymbol{H}_{n_q}^{-1}\boldsymbol{A}_{n_q}^*(x)\}_{3,3},$$

with $\boldsymbol{U}^*(x)$ and $\boldsymbol{A}_{n_q}^*(x)$ as $\boldsymbol{U}(x)$ and $\boldsymbol{A}_{n_q}(x)$ defined previously with bootstrap data (X_i^*, Y_i^*) and the variance given by:

$$\mathsf{Var}\{\widehat{\mathcal{K}}(x)\} \approx \{p(x)\}^{-2} B(x)^{-1}\boldsymbol{L}^{-1}\boldsymbol{P}\boldsymbol{L}^{-1}, \tag{19}$$

where $B(x)$ is defined after equation (13).

Corollary 1: *Assume conditions (A1)-(A7), a* $(1-\alpha)100\%$ *bootstrap confidence band for the EPK* $\mathcal{K}(x)$ *is:*

$$[f(x) : \sup_{x\in E}\{|\widehat{\mathcal{K}}(x) - f(x)|\widehat{Var}(\widehat{\mathcal{K}})^{-1/2}\} \le L_\alpha^*],$$

where the bound L_α^* *satisfies*

$$\mathrm{P}^*[-\{\boldsymbol{U}(x)^{-1}\boldsymbol{H}_{n_q}^{-1}\boldsymbol{A}_{n_q}^*(x)\}_{3,3}/\{B(x)^{-1}\boldsymbol{L}^{-1}\boldsymbol{P}\boldsymbol{L}^{-1}\}_{3,3} \le L_\alpha^*] = 1-\alpha.$$

The estimator $\widehat{Var}(\widehat{\mathcal{K}})$ is computed in a similar fashion as in the previous section.

2.2 Confidence Bands based on Smoothing Implied Volatility

Although the nonparametric estimator of the PK is reasonable in theoretical sense, it often fails to provide stable and economically treatable estimators with real data. One way to stabilizing the empirical SPD is the use of data-driven local bandwidths (see Vieu 1993) or a multiple-testing-type adaptive technique of Lepski and Spokoiny (1997). These alternative methods are tools of general purpose and address the bias-variance trade-off locally. They are known to be either asymptotically optimal or to have a near oracle property. Although the adaptive bandwidth provides us with optimal estimators, it is still possible that the noise is too large and the algorithm fails to provide a curve with a small bias and easy interpretation. This point is stressed in Rookley (1997): “implied volatilities on the other hand tend to be less volatile and differences in implied volatilities convey much more economic information than option prices alone, as implied volatilities already embed much of the fundamental information available.”

We follow, however, an alternative approach and stabilize the empirical SPD by a two-step procedure as in Rookley (1997) and Fengler (2005). At the first step, we estimate the implied volatility (IV) function by a local polynomial regression. At the second step, we plug the smoothed IV into the BS formula to obtain a semiparametric estimator of the option price. This approach relies on a bijective transformation of the call prices to the IV space and reflects the tendency of investors to quote the options in terms of IV. Aït-Sahalia and Lo (1998) used a similar semiparametric technique for dimension reduction purposes. Note that the procedure does not require the BS model to hold, but leads to finite sample improvements, while being asymptotically equivalent to the original estimator (see Theorem 3). Thus we impose more assumptions on the functional form of the call price function and focus only on the nonparametric structure of the volatility surface. Thus the noise is relevant only in the estimation of the volatility surface. However, it is well recognized that the volatility is less noisy and its shape is more tractable and easy to interpret economically. Moreover, we can improve our two-step procedure further by adopting adaptive techniques for the volatility surface.

Formally we smooth the IV using a local polynomial regression in moneyness M, with the implicit assumption on the pricing formula is homogenous of degree 1 w.r.t. the asset price and the strike price as proved in Renault (1997). In the absence of dividends, the moneyness is defined at time t as $M_{it} = S_t / X_i$. The heteroscedastic model for the IV is given by:

$$\sigma_i = \sigma(M_{it}) + \sqrt{\eta(M_{it})}\upsilon_i, \quad i = 1, \ldots, n_q, \tag{20}$$

where υ_i are the i.i.d. errors with zero mean, unit variance and $\eta(\cdot)$ is the volatility function. We make the same assumptions about the implied volatility $\sigma(\cdot)$ as we did for the option prices $C(\cdot)$ in Section 1.1.

Defining the rescaled call option price $c(M_{it}) = C(X_i)/S_t$, we obtain from the BS formula

$$c(M_{it}) = c\{M_{it}; \sigma(M_{it})\} = \Phi\{d_1(M_{it})\} - \frac{e^{-r\tau}\,\Phi\{d_2(M_{it})\}}{M_{it}},$$

where

$$d_1(M_{it}) = \frac{\log(M_{it}) + \left\{r + \frac{1}{2}\sigma(M_{it})^2\right\}\tau}{\sigma(M_{it})\sqrt{\tau}}, \qquad d_2(M_{it}) = d_1(M_{it}) - \sigma(M_{it})\sqrt{\tau}.$$

Combining the result of Breeden and Litzenberger (1978) with the expression for $c(M_{it})$ leads to the SPD

$$q(x) = e^{r\tau} \frac{\partial^2 C}{\partial X^2}\Big|_{X=x} = e^{r\tau} S_t \frac{\partial^2 c}{\partial X^2}\Big|_{X=x} \tag{21}$$

with

$$\frac{\partial^2 c}{\partial X^2} = \frac{\mathrm{d}^2 c}{\mathrm{d}M^2}\left(\frac{M}{X}\right)^2 + 2\frac{\mathrm{d}c}{\mathrm{d}M}\frac{M}{X^2}. \tag{22}$$

As it is shown in the Appendix the derivatives in the last expression can be determined explicitly and are functions of $V \stackrel{\text{def}}{=} \sigma(M)$, $V' \stackrel{\text{def}}{=} \partial\sigma(M)/\partial M$ and $V'' \stackrel{\text{def}}{=} \partial^2\sigma(M)/\partial M^2$. We estimate the latter quantities by the nonparametric local polynomial regression for the IV of the from

$$\sigma(M_{it}) \approx V(M) + V'(M)(M_{it} - M) + \frac{1}{2}V''(M)(M_{it} - M)^2,$$

for M near M_{it}. The respective estimators are denoted by $\hat{V}$, $\hat{V}'$ and $\hat{V}''$. Plugging the results into (21)-(22) we obtain the estimator of SPD in the smoothed IV space. Assuming that the IV process fulfills the $(A1)-(A7)$ in the appendix instead of $C(\cdot)$, we conclude that Theorem 2.1 of Claeskens and Van Keilegom (2003) holds also for $\hat{V}$, $\hat{V}'$ and $\hat{V}''$. Note that the convergence rate of $\hat{V}$ and $\hat{V}'$ is lower than of $\hat{V}''$. Relying on this fact, we state the asymptotic behavior of $\hat{q}(x) - q(x)$ in the next theorem.

Theorem 3: *Let $\sigma(\cdot)$ satisfy the assumptions (A1)-(A7). Then with $M = S_t/x$ it holds*

$$\sqrt{n_q h_{n_q}^5}\{\hat{q}(x) - q(x)\} \xrightarrow{\mathcal{L}} \mathrm{N}\{0, r(M)^2\sigma_V^2(M)\}, \tag{23}$$

where

$$\begin{aligned} r(M) &\stackrel{def}{=} e^{r\tau} S_t \frac{M^2}{x^2}\Big[\varphi\{d_1(M)\}\{\sqrt{\tau}/2 - \frac{\log(M) + r\tau}{V(M)^2\sqrt{\tau}}\} \\ &- e^{-r\tau}\varphi\{d_2(M)\}\{-\sqrt{\tau}/2 - \frac{\log(M) + r\tau}{V(M)^2\sqrt{\tau}}\}/M\Big] \end{aligned}$$

and $\sigma_V^2(M) \stackrel{def}{=} [B_V(M)^{-1}\boldsymbol{L}^{-1}\boldsymbol{T}\boldsymbol{L}^{-1}]_{(3,3)}$, with $\sigma_V^2(M)$ defined as in (15).

Proof. The proof is given in the Appendix. ■

Theorem 3 allows us to construct the confidence bands of the SPD estimated semiparametrically using the confidence bands for the IV. The variance of the estimator is obtained by the delta method in the following way

$$\mathrm{Var}\{\hat{q}(x) - q(x)\} = \left(\frac{\partial q}{\partial V''}\right)^2 \mathrm{Var}\{\hat{V}''(M) - V''(M)\}.$$

The variance $\text{Var}\{\hat{V}''(M)-V''(M)\}$ is estimated using a sandwich estimator similarly to (16), and $\frac{\partial q}{\partial V''}=e^{r\tau}S_t\frac{M^2}{x^2}\Big[\varphi\{d_1(M)\}\{\sqrt{\tau}/2-\frac{\log(M)+r\tau}{V(M)^2\sqrt{\tau}}\}-e^{-r\tau}\varphi\{d_2(M)\}\{-\sqrt{\tau}/2-\frac{\log(M)+r\tau}{V(M)^2\sqrt{\tau}}\}/M\Big]$. Here we have proved that it is sufficient to consider only the variance of the second derivative of V, as the other terms involved are of higher order.

2.3 Extension to Dependent Data

In the previous sections we have assumed independent data in the estimation of both the historical density and the SPD. Violation of this assumption may lead to misspecified asymptotic results and wrong confidence bands. The assumption is, however, feasible in our study for both densities. The confidence bands for a simple density estimator of time series data are analysed by Liu and Wu 2010, see Section 2.1 and can be directly transferred to the historical density in our setup. Note however, that the impact of these results on the confidence bands for the EPK is flattened by the higher convergence rate of the historical density. Regarding the SPD note that assuming sufficient liquidity we can use options only with a given maturity traded on a single day. This implies that the data used to estimate SPD is not a time series data and there is no need to take the serial correlation into account.

Nevertheless, to serve a general purpose estimation, it is still interesting to generalize our theoretical results to dependent data. Liu and Wu (2010) developed uniform confidence bands for kernel density estimators and Nadaraya–Watson estimation for a general class of time series models. In this section we adopt their approach to our problem. We extend the setup to time dependence and consider the model as in (6)

$$Y_i=C(X_i)+\sigma(X_i)\varepsilon_i, \qquad i=1,\dots,n_q,$$

with the strike prices being a causal stationary process $X_i=G(\dots,\eta_{i-1},\eta_i)$, η_i are i.i.d. and independent with ε_i.

We focus on the estimation of $C(\cdot)$ as in (8) and keep $\sigma(\cdot)$ known for the derivations in this subsection. The physical dependence measure $\theta_{(i,\gamma)}\overset{\text{def}}{=}\|X_i-X_i'\|_\gamma=(\mathsf{E}|X_i-X_i'|^\gamma)^{1/\gamma}$, where X_i' is a coupled process of X_i with η_0 is replaced by an i.i.d. copy of η_0', i.e. $X_i'=G(\eta_0',\dots,\eta_{i-1},\eta_i)$. Additionally define the dependence measure with coupled whole past as $\Psi_{i,\gamma}\overset{\text{def}}{=}\|G(\eta_0,\dots,\eta_{i-1},\eta_i)-G(\eta_0',\dots,\eta_{i-1}',\eta_i')\|_\gamma$. Suppose that $||X_i||_\kappa\leq\infty$ for some $\kappa>0$. Let $\kappa'=\min(\kappa,2)$ such that $\Theta_{n_q}=\sum_{i=1}^{n_q}\theta_{(i,\kappa')}^{\kappa'/2}$. Define $\mathcal{Z}_{n_q}\overset{\text{def}}{=}\sum_{k=-n_q}^{\infty}(\Theta_{n_q+k}-\Theta_k)^2$ and $\tilde{\xi}_i\overset{\text{def}}{=}(\dots,\varepsilon_{i-1},\varepsilon_i,\dots,\eta_{i-1},\eta_i)$.

(A8) Assume $(||X_i||_\kappa\leq\infty$ for $\kappa>0$. The density of η_i is positive and uniformly bounded over its whole support up to the third derivative. There exists a constant $M<\infty$ such that $\sup[|f_{X_{n_q}|\tilde{\xi}_{n_q-1}}(x)|+|f'_{X_{n_q}|\tilde{\xi}_{n_q-1}}(x)|+|f''_{X_{n_q}|\tilde{\xi}_{n_q-1}}(x)|]\leq M$ almost surely. ε_i has bounded fourth moments. $\Psi_{n_q,\gamma}=\mathcal{O}(n_q^{-r})$ for some γ and $r>\delta_1/(1-\delta_1)$,

$0 < \delta_1 < 1/4$. There exists a constant δ such that $0 < \delta \leq \delta_1 < 1$ and $h_{n_q} = \mathcal{O}(n_q^{-\delta})$, $n_q^{-\delta} = \mathcal{O}(h_{n_q})$. Furthermore $\theta_{(n_q,\kappa)} = \mathcal{O}(\rho^{n_q})$ for some $\kappa > 0$ and $0 < \rho < 1$.

Let $\hat{\mathcal{F}}_{n_q}(x)$ be the standardized process:

$$\hat{\mathcal{F}}_{n_q}(x) \stackrel{\text{def}}{=} n_q^{1/2} h_{n_q}{}^{5/2} \{\hat{q}(x) - q(x)\} / [\hat{\sigma_q}(x)]^{1/2}.$$

Theorem 4: *Under assumptions (A1)-(A6), (A8), $h_{n_q} = \mathcal{O}\{(n_q \log n_q)^{-1/9}\}$, $\mathcal{Z}_{n_q} h_{n_q}^3 = \scriptstyle{\mathcal{O}}(n_q \log n_q)$, it follows*

$$\mathrm{P}\left[(-2\log h_{n_q})^{1/2} \left\{ \sup_{x \in E} |\hat{\mathcal{F}}_{n_q}(x)| - c_{n_q} \right\} < z \right] \longrightarrow \exp\{-2\exp(-z)\},$$

where $c_{n_q} = (-2\log h_{n_q})^{1/2} + (-2\log h_{n_q})^{-1/2}\{x_\alpha + \log(R/2\pi)\}$.

Liu and Wu (2010) note an interesting dichotomy phenomenon, where the rate of convergence is the consequence of an interplay between the strength of dependency and the bandwidth h_{n_q}. Accordingly, we suggest to undersmooth as a smaller bandwidth would both reduce the bias and the effect of dependency. The rate of h_{n_q} is set to $\mathcal{O}\{(n_q \log n_q)^{-1/9}\}$.

3 MONTE-CARLO STUDY

The practical performance of the above theoretical considerations is investigated via two Monto-Carlo studies. The first simulation aims at evaluating the performance under the BS hypothesis, while the second simulation setup does the same under a realistically calibrated surface. The confidence bands are applied to DAX index options. We first study the confidence bands under a BS null model (Section 3.1). Naturally, without volatility smile, both the BS estimator and nonparametric estimator are expected to be covered by the bands. While in the presence of volatility smile (Section 3.2), we expect our tests to reject the BS hypothesis in most cases.

3.1 How Well is the BS Model Covered?

In the first setting, we calibrate a BS model on day 20010117 with the interest rate set equal to the short rate $r = 0.0481$, $S_0 = 6500$, strike prices in the interval $[6000, 7400]$. We refer to Aït-Sahalia and Duarte (2003) on the sources of the noise and use an identical simulation setting, with the noise being uniformly distributed in the interval $[0, 6]$. Figure 4 is a scatter plot of generated observations of European call option prices against strikes, the data is clustered in discrete values of the strike price. Recall that bandwidths in the following context are all selected to

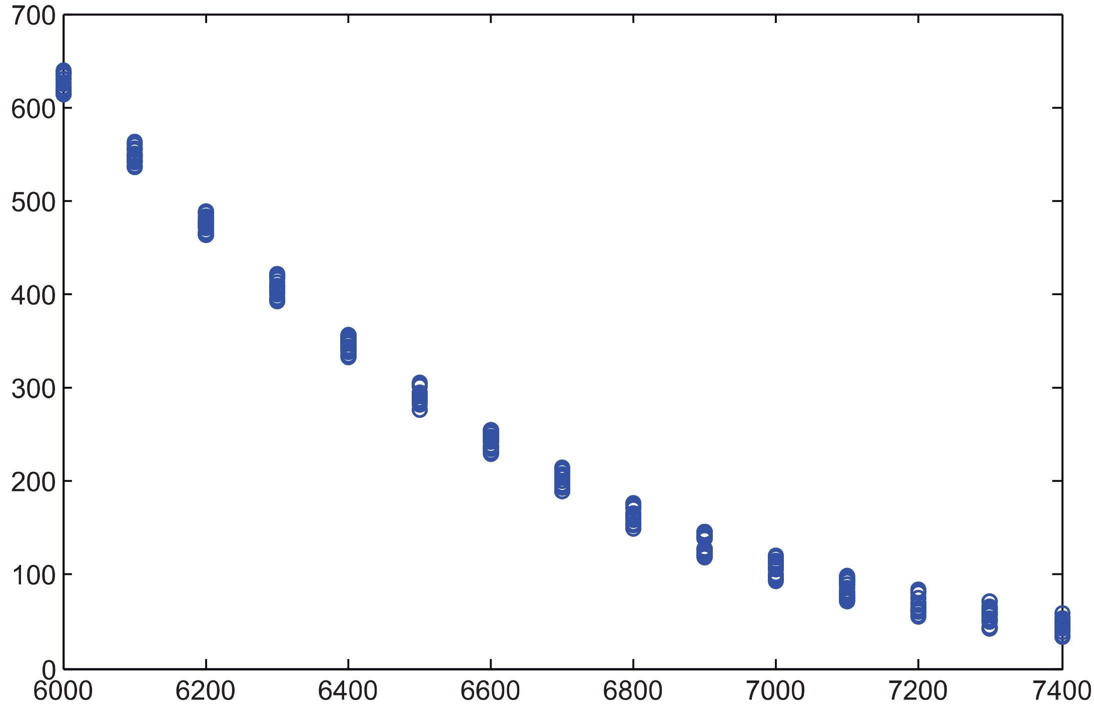

Figure 4 Generated noisy BS call option prices against strike prices.

minimize pricing error using a leave-one-out approach on the bivariate grid $[1/n_p; 1] \times [1/n_q; 1]$.

Figure 5 shows a nonparametric estimator for the SPD and a parametric BS estimator. The two estimators roughly coincide except for a small wiggle, thus the bands drawn around the nonparametric curve also fully cover the parametric one. The accuracy is evaluated by calculating the coverage probabilities and average area within the bands, see Table 1 (see the rows labeled "null"). The coverage probabilities is determined via 500 simulations, whenever the hypothesized curve calculated on a grid of 100. The coverage probability approaches its nominal level with increasing sample size, but never reaches it. This may well be attributed to the above mentioned poor convergence of Gaussian maxima to the Gumbel distribution. The area within the bands reflects the stability of the estimation procedure. The bands get narrower with increasing sample size.

The bias correction for the SPD follows the approach of Xia (1998). The HD is corrected as in Jones, Linton, and Nielson (1995). The bias correction for EPK relies on the linear term from Lemma 2. The correction of the bands mimics the Bonferroni correction in Eubank and Speckman (1993) and is based on the asymptotic confidence intervals in (13) and (14). We conclude that the bias correction approach and the Bonferroni correction are not better than the proposed method for all sample sizes.

HDs are estimated from simulated stock prices following geometric Brownian motion with $\mu = 0.23$. A BS EPK estimator could be tested using the above procedure. Due to boundary effects, we concentrate on moneyness ($M_t = S_t/X$) in

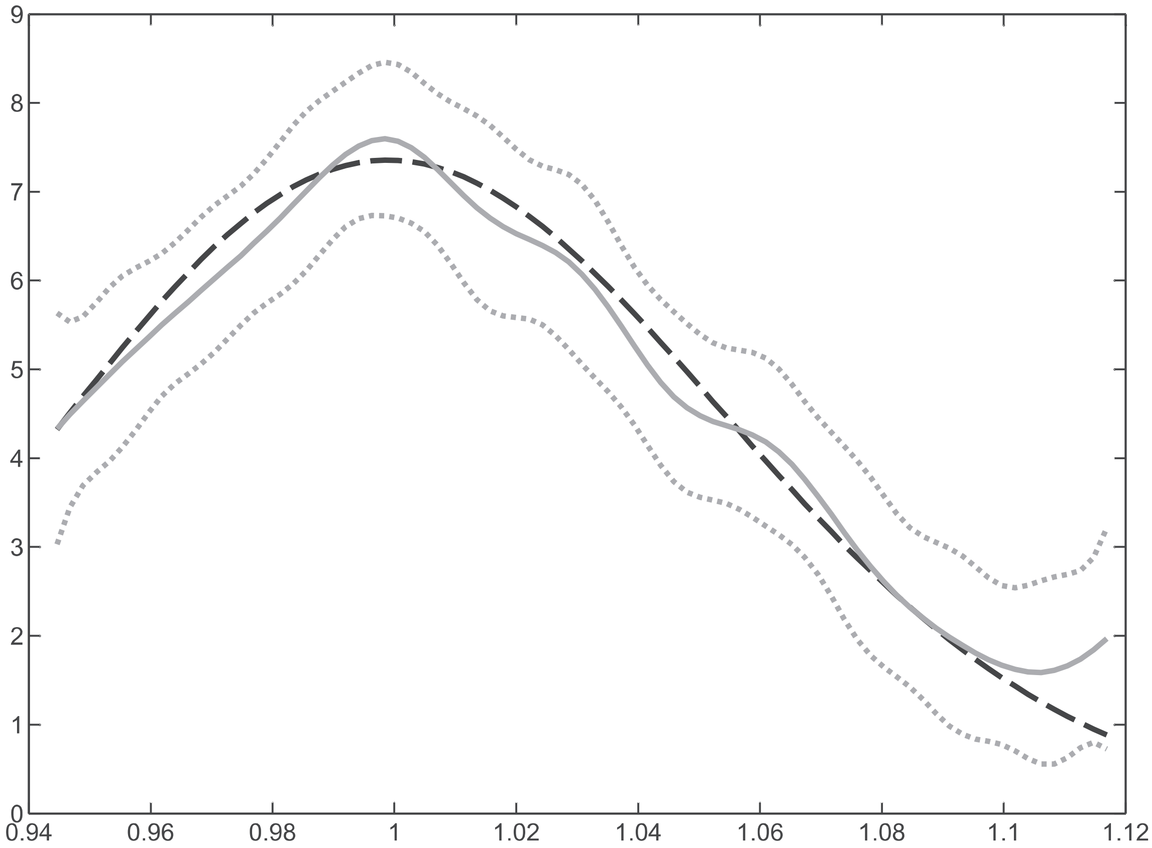

Figure 5 Estimation of SPD (gray), bands (dotted) and the BS SPD (dashed), with $h_{n_q} = 0.085$, $\alpha = 0.05$, $n_q = 300$.

[0.95, 1.1]. Figure 6 displays the nonparametric EPK with confidence band and the BS EPK covered in the band. We observe that the BS EPK is strictly monotonically decreasing. The summary statistics are given in Table 1, due to the additional source of randomness introduced through the estimation of $p(x)$, the coverage probabilities are less precise than the corresponding coverage probabilities for SPD. Nevertheless, the probabilities are getting closer to their nominal values and the bands get narrower when the sample size increases.

3.2 How Well is the Band in Reality?

Section 3.1 studied the performance of the bands under the null hypothesis with BS assumption, while this section is designed to investigate the performance of the bands when the null hypothesis is violated by a realistic volatility smile observed in the market. Keeping the parameters identical to the setup of the first study, we generated the data with a smoothed volatility function based on data for or options traded on 20010117 with $\tau = 3M, 6M$ to maturity.

Figures 7 and 8 report the estimators for SPD and EPK. The bands do not cover the BS estimator. Correspondingly, Table 1 (see the rows labeled "alter.") show the

Table 1 Averaged, coverage probability (area) of the uniform confidence band over 500 simulations in different cases

Level	Maturity	Method	$n_q = 300$	450	600
5% (null)	3M	EPK	0.782 (2.54)	0.798 (2.49)	0.802 (2.38)
		EPK (bias)	0.798 (2.51)	0.800 (2.43)	0.802 (2.31)
		EPK (Bonfer.)	0.673 (2.98)	0.697 (2.88)	0.754 (2.76)
		SPD	0.906 (2.40)	0.914 (2.20)	0.923 (1.99)
		SPD (Bonfer.)	0.873 (2.68)	0.924 (2.57)	0.929 (2.44)
	6M	EPK	0.860 (2.50)	0.875 (2.43)	0.890 (2.41)
		EPK (bias)	0.862 (2.53)	0.883 (2.41)	0.899 (2.42)
		EPK (Bonfer.)	0.785 (2.90)	0.801 (2.67)	0.824 (2.74)
		SPD	0.896 (2.44)	0.906 (2.13)	0.920 (2.07)
		SPD (Bonfer.)	0.883 (2.73)	0.894 (2.69)	0.903 (2.52)
10%(null)	3M	EPK	0.706 (2.47)	0.736 (2.34)	0.762 (2.23)
		EPK (bias)	0.712 (2.45)	0.737 (2.33)	0.771 (2.23)
		EPK (Bonfer.)	0.673 (2.33)	0.686 (2.12)	0.734 (2.01)
		SPD	0.795 (2.17)	0.812 (2.06)	0.853 (1.88)
		SPD (Bonfer.)	0.764 (2.12)	0.801 (2.00)	0.833 (1.98)
	6M	EPK	0.729 (2.50)	0.774 (2.23)	0.829 (2.31)
		EPK (bias)	0.713 (2.47)	0.753 (2.26)	0.835 (2.30)
		EPK (Bonfer.)	0.671 (2.86)	0.745 (2.88)	0.798 (2.72)
		SPD	0.800 (2.34)	0.814 (2.08)	0.860 (1.94)
		SPD (Bonfer.)	0.763 (2.55)	0.800 (2.46)	0.847 (2.36)
5% (alter.)	3M	EPK	0.512 (2.43)	0.178 (2.23)	0.050 (2.02)
		EPK (bias)	0.543 (2.42)	0.235 (2.27)	0.145 (1.99)
		EPK (Bonfer.)	0.372 (2.51)	0.239 (2.37)	0.099 (2.12)
	6M	EPK	0.592 (2.53)	0.410 (2.17)	0.178 (2.02)
		EPK (bias)	0.541 (2.49)	0.349 (2.12)	0.251 (2.01)
		EPK (Bonfer.)	0.331 (2.34)	0.136 (2.16)	0.150 (2.15)
10% (alter.)	3M	EPK	0.258 (2.12)	0.050 (2.04)	0.030 (2.01)
		EPK (bias)	0.268 (2.13)	0.043 (2.01)	0.001 (2.00)
		EPK (Bonfer.)	0.148 (2.78)	0.030 (2.61)	0.001 (2.54)
	6M	EPK	0.375 (2.22)	0.410 (2.13)	0.178 (2.00)
		EPK (bias)	0.362 (2.21)	0.432 (2.13)	0.176 (2.01)
		EPK (Bonfer.)	0.231 (2.46)	0.221 (2.35)	0.110 (2.25)

(bias) means bias correction, (Bonfer.) means Bonferoni correction.

coverage probabilities, which rapidly decrease when sample sizes are increasing. However, the area within the bands does not change significantly when compared with the results of Section 3.1. We conclude that the confidence bands are useful for detecting the deviation from the BS model.

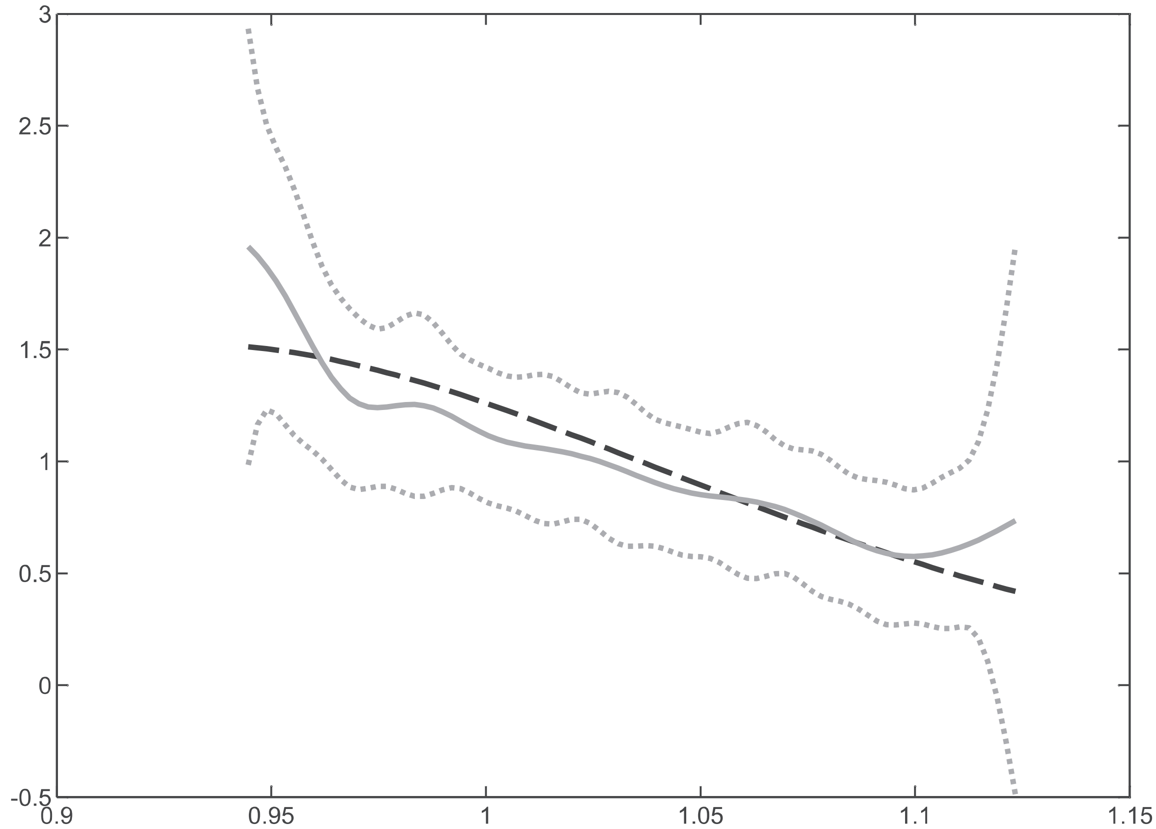

Figure 6 Estimation of EPK (gray), bands (dotted), and the BS EPK (dashed), with $h_{n_q}=0.085$, $h_{n_p}=0.060$ and $\alpha=0.05$, $n_p=2000$, $n_q=300$.

Note that the suggested method works for other processes, for example, the Heston model, which allows for flexible forms of the pricing kernels. The simulation study confirms the good performance of the confidence bands. The results are not reported here for the brevity of presentation.

4 AN ILLUSTRATION WITH DAX DATA

This section aims at illustrating the functionality of our bands by checking the coverage of BS EPK, which indicates how much the market risk behavior deviates from the BS model. The procedure can be seen as a test of monotonicity of pricing kernels. The available tests for monotonicity (Ghosal, Sen, and van der Vaart (2000), Lee, Linton, and Whang (2009), Chetverikov (2012)) work for (regression) functions and not for derivative estimation as required here. We take a dynamic point of view by considering the EPK estimated at different dates.

4.1 Data

In contrast to previous studies that are mainly based on S&P500 data, we focus on intraday European options on the DAX options. The source is the

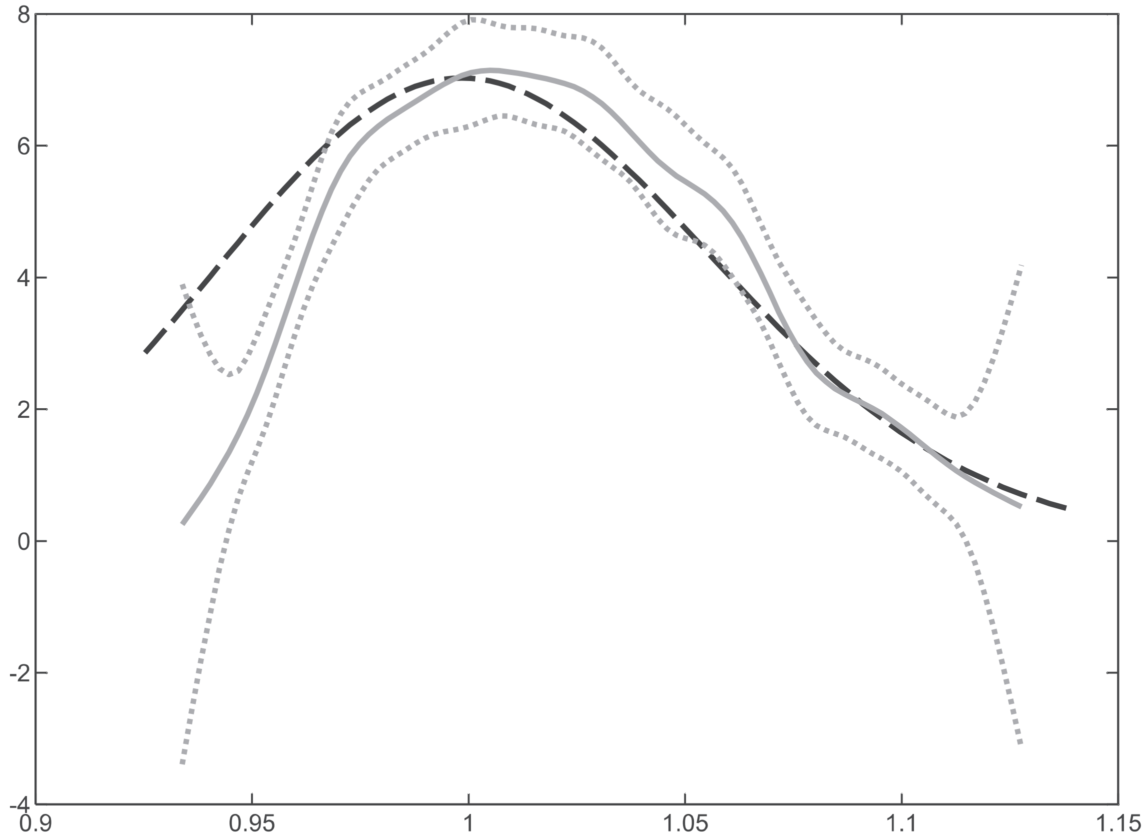

Figure 7 Plot of confidence bands (dotted), nonparametrically estimated SPD (gray), the fitted BS (dashed) SPD with simulated volatility smile, $n_q = 300$, $h_{n_q} = 0.066$, $\alpha = 0.05$.

European Exchange EUREX and data available by C.A.S.E., RDC SFB 649 (http://sfb649.wiwi.hu-berlin.de) in Berlin. The extracted observations for our analysis cover the period between 1998 and 2008. The smoothing in volatility approach described in Section 2.2 is applied to estimate the EPK (denoted as Rookley method). As we cannot find traded options with the same maturity on each day, we consider options with maturity 15 days (10 trading days) across several years. Specifically, we extract a time series of options for every month from January 2001 to December 2006; this adds up to 63 days.

To make sure that the data correctly represents the market conditions, we use several cleaning criteria. In our sample, we eliminate the observations with $\tau < 1D$ and $IV > 0.7$. Also, we skip the option quotes violating general no-arbitrage condition i.e., $S > C > \max\{0, S - Xe^{-r\tau}\}$. Due to the put-call parity, both out-of-the-money call options and in-the-money puts are used to compute the smoothed volatility surface. The median of intra day stock prices is used to compute the SPD. We use a window of 500 returns for nonparametric kernel density estimators of HD.

Figure 9 describes the relative position of the HD and SPD on a specific day, the EPK peak is apparently created through the different probability mass contributions at different moneyness states.

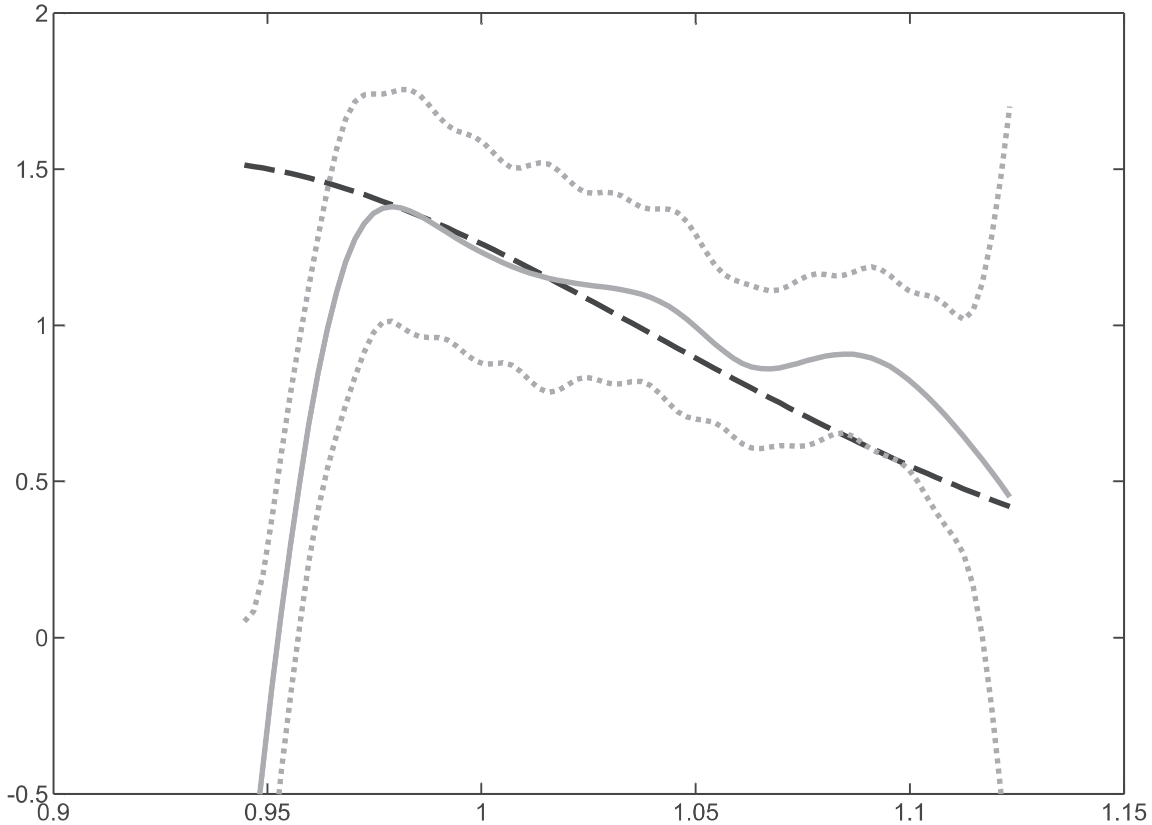

Figure 8 Plot of confidence bands (dotted), nonparametrically estimated EPK (gray), the fitted BS (dashed) EPK with simulated volatility smile $n_p = 2000$, $n_q = 300$, $h_{n_q} = 0.063$, $h_{n_p} = 0.011$, $\alpha = 0.05$.

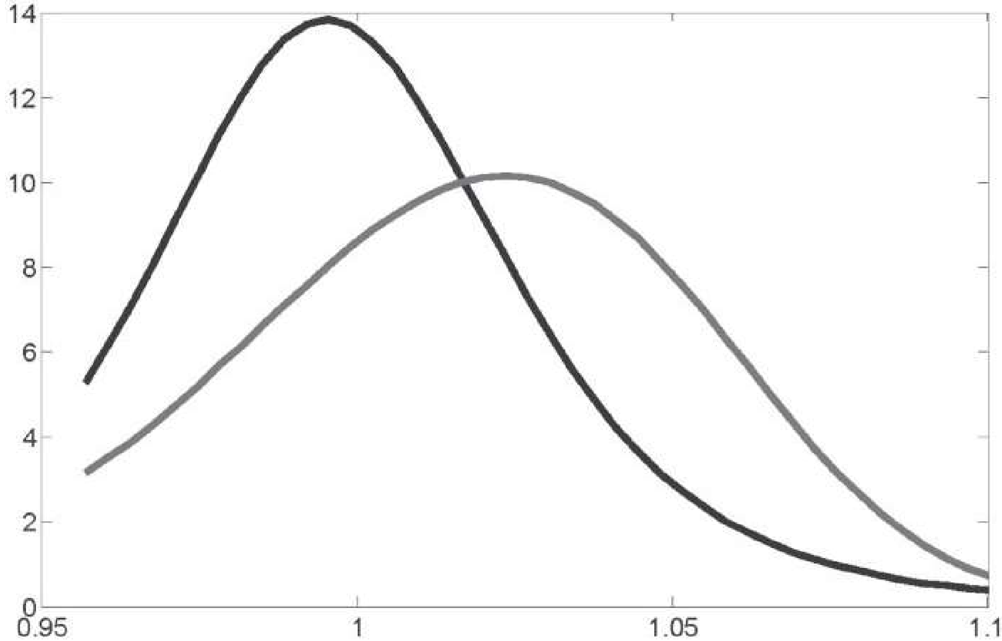

Figure 9 Plot of estimated SPD on February 28, 2006 (Rookley, $h_{n_q} = 0.063$, black) and HD ($h_{n_p} = 0.0106$, gray).

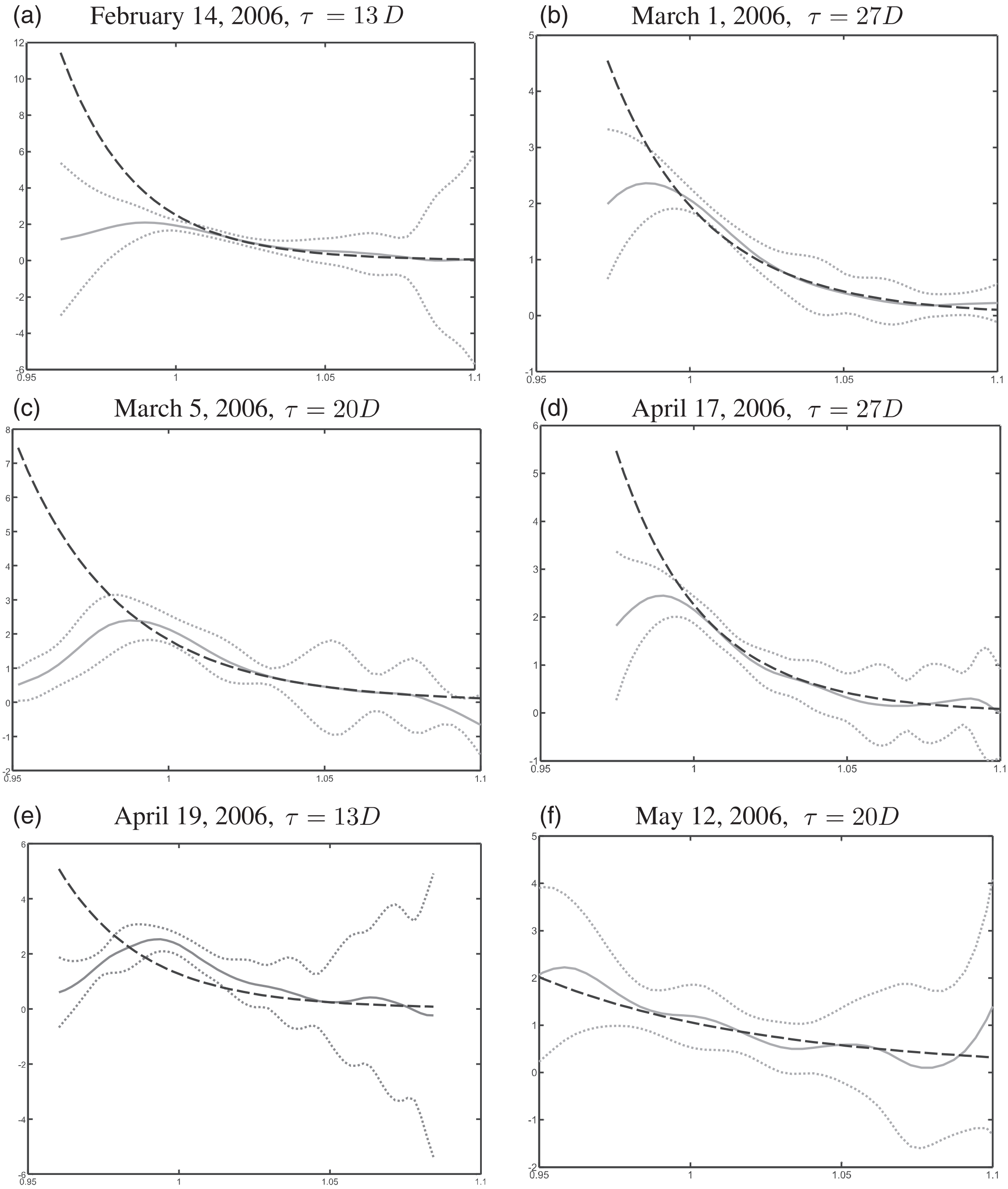

Figure 10 Estimated BS EPK (dashed), Rookley EPK (gray), uniform confidence band (dotted), $\alpha = 0.05$.

4.2 Estimation of DAX EPK and its Uniform Confidence Band

We consider two specifications for the pricing kernels. In the first specification, the BS pricing kernels have a marginal rate of substitution with power utility function:

$$\mathcal{K}(M) = \beta_0 M^{-\beta_1}, \tag{24}$$

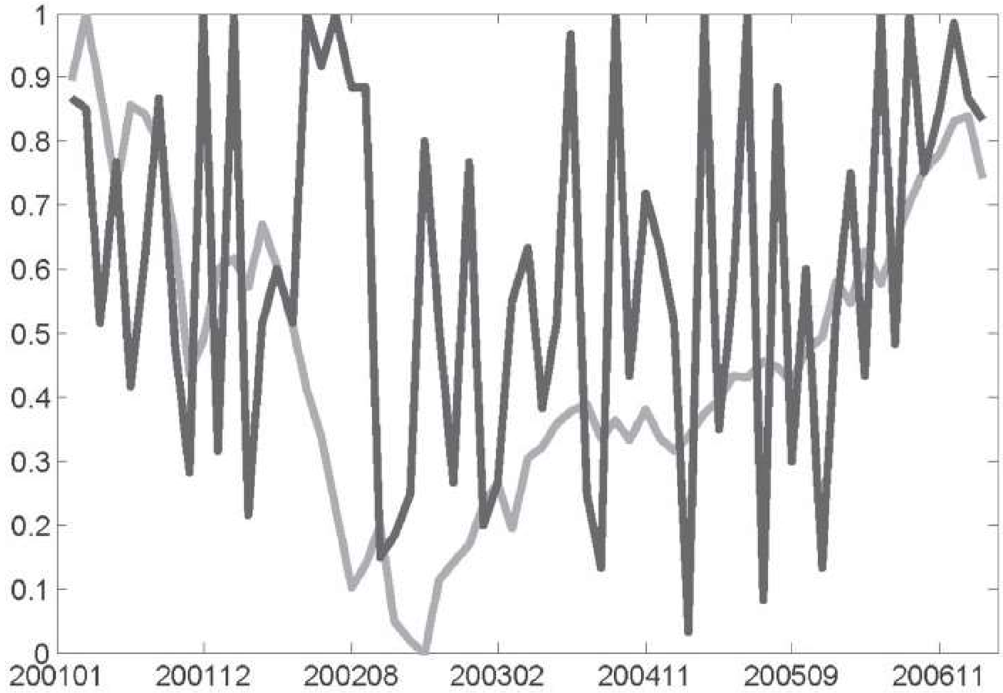

Figure 11 Coverage probability ($\alpha = 0.05$) estimated at 63 trading days and the DAX index (gray, rescaled to $[0, 1]$), $\tau = 3M$.

where β_0 is a scaling factor and β_1 determines the slope of pricing kernel. Thus the BS calibration is realized by linearly regressing the (ordered) log-EPK on log-moneyness. In the second specification, we construct the nonparametric confidence bands as described in Section 2.2. A sequence of EPKs and corresponding bands are shown in Figure 10. In most of the cases, the BS EPKs are rejected via the confidence bands. The amount of deviation from the hypothesized BS specification though provides us valuable information about how risk hungry investors are. Besides, the area of between the bands varies over time, which gives us insights into the variabilities of the prevailing risk patterns. In sum, the bands do not only provide a simple test for hypothesizes EPKs, but also help us to study the dynamics of risk patterns over time.

4.3 Linking Economic Conditions to EPK Dynamics

We use two different indicators for the deviation from a simple BS model. As an approximation to the coverage probability, we calculate the proportion of grid points of the band which covers the BS EPK. As a second measure, we introduce the average width of the confidence bands over the moneyness interval $[0.95, 1.1]$ as a proxy for the area between the confidence bands. This provides us with a measure of variability, see also Theorem 2.

The first risk pattern time series is given in Figure 11, where we display the DAX index (scaled to $[0, 1]$) together with the coverage probability. We discover that

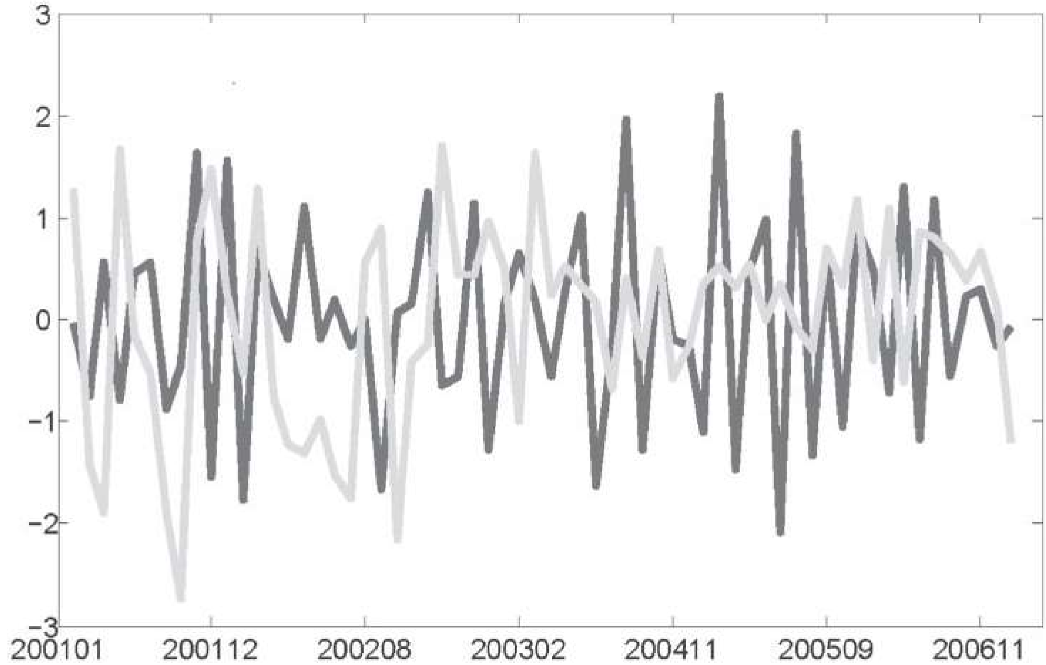

Figure 12 First difference of coverage probability and the DAX index return (gray, standardized).

the coverage probability becomes less volatile when the DAX index level is high. Figure 12 shows the differenced time series. From a simple correlation analysis, we argue that the change in coverage probability and DAX return (with a lag of 3M) are highly negatively correlated (correlation −0.3543) when the DAX index goes down (200101-200302). On the contrary, in the period when the DAX goes up, one observes a large positive correlation (0.3151). What does this mean economically? This implies in a period of worsening economic condition, a positive part of the monthly DAX returns induces a greater hunger for risk in a delay horizon of 3 months. Positive returns have just the opposite effects. With boosting and bullish markets, the positive correlation indicates a 3-month horizon of decreasing risk aversion. The exercise we have done so far support these economic reasoning. Risk aversion seems to be higher in recessions and lower in boom times. This corresponds to the findings in the economics literature. Economic agents (e.g., corporate firms, banks, households) with higher risk aversion tend to hold more liquid assets, driving down the interest rate. At the same time, the higher risk aversion calls for a higher rate of return on risky assets. A lower interest rate and a higher rate on risky assets generate a higher risk premium, Gilchrist and Zakrajšek (2012).

As far as the average width of the bands is concerned, we may conclude from Figure 13 and Figure 14 that in periods of clearly bullish or bearish momentum, the volatility of the width of the confidence band is higher. This may be caused by the uncertainty of the market participants about the long-term persistence of the trend. The lag effect on risk hunger is also detectable for this constructed indicator. Over the whole observation interval, the correlation between the monthly DAX return and the change in the average width is −0.3230 for a 1M lag and −0.2717 for 3M.

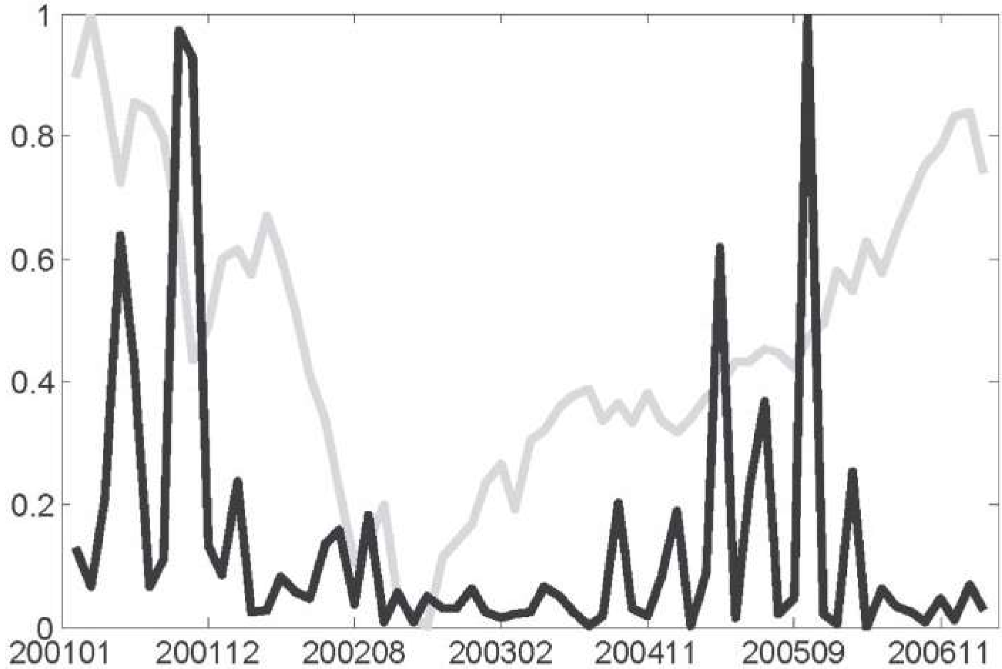

Figure 13 Area of the confidence bands ($\alpha = 0.05$) estimated at 63 trading days and the DAX index (gray, rescaled to $[0, 1]$).

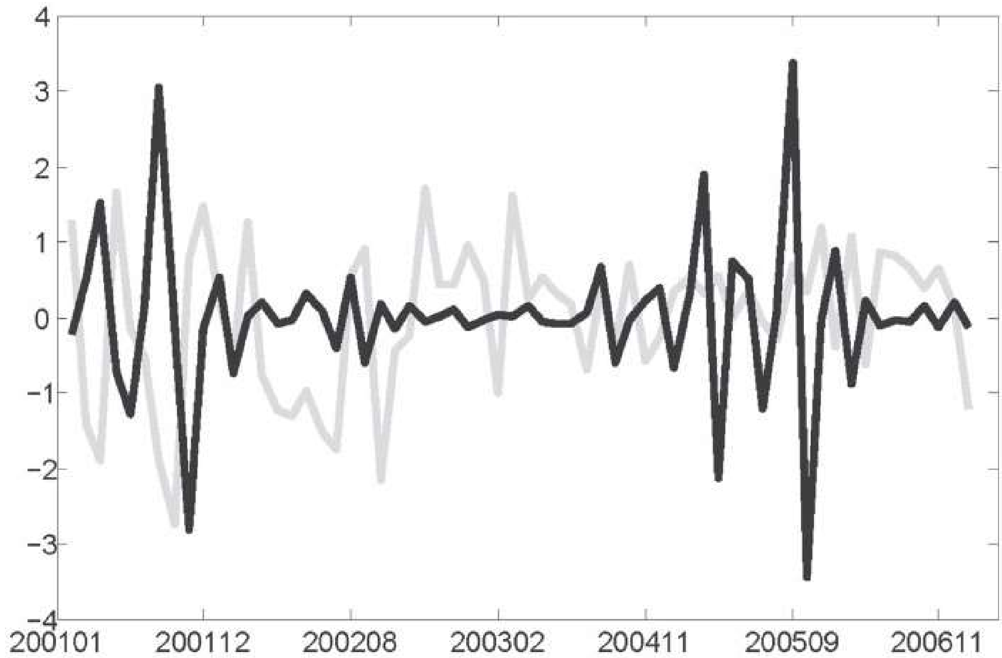

Figure 14 First difference of area and the DAX index return (gray, standardized).

5 CONCLUSIONS

Pricing Kernels are important elements in understanding investment behavior since they reflect the relative weights given by investors' states of nature (Arrow–Debreu securities). Pricing kernels may be deduced in either parametric or nonparametric approaches. Parametric approaches like a simple BS model are too restrictive to account for the dynamics of the risk patterns, which induces the well-known EPK paradox. Nonparametric approaches allow more flexibility and reduce the modeling bias. Simple tools like uniform confidence bands help us to conduct tests against any parametric assumption of the EPKs i.e., shape inspection. Considering the numerical stability, we smooth the IV surface via the Rookley's method, and obtain SPD estimator.

We have studied systematically the methodology of constructing the uniform bands for both semiparametric or nonparametric estimators. Based on the confidence bands, we explored two indicators to measure risk aversion over time and linked it with DAX index; the first one is the coverage probability measuring the proportion of the BS curve covered in bands, while the second one is the area indicator measuring the variability of the estimator. We found out that there are strong correlations between DAX index and our indicators with lag effects. The smooth bootstrap is also studied without a significant improvement in finite sample performance. One interest further extension is employing robust smoothers to improve the bootstrap performance.

6 APPENDIX

Assumptions:

(A1) $h_{n_q} \to 0$ in such a way that $\{\log n_q/(n_q h_{n_q})\}^{1/2} \cdot h_{n_q}^3 \to 0$, and the optimal rate bandwidth, to guarantee undersmoothing, would be $\mathcal{O}\{(\log n_q \cdot n_q)^{-1/9}\}$.

(A2) The kernel functions $K \in C^{(1)}[-1,1]$ (adopted for estimating both HD $p(x)$ and SPD $q(x)$) are symmetric and takes value 0 on the boundary.

(A3) For the likelihood function $\mathcal{L} \in C^{(1)}(E)$ it holds that $\inf_{x \in E} \mathcal{L}(x) > 0$. $C(x) \in C^{(4)}(E)$. Additionally the third partial derivatives of $\mathcal{L}(Y,C)$ with respect to C exists and is continuous in C for every y. The Fisher information $I(C(x))$ has a continuous derivative and $\inf_{x \in E} I\{C(x)\} > 0$.

(A4) There exists a neighborhood $N(C(x))$ such that

$$\max_{k=1,2} \sup_{x \in E} || \sup_{C \in N\{C(x)\}} \frac{\partial^k}{\partial C^k} \mathcal{L}(y;C) ||_\lambda < \infty$$

for some $\lambda \in (2,\infty]$. Furthermore

$$\sup_{x \in E} \mathrm{E}[\sup_{C \in N\{C(x)\}} | \frac{\partial^3}{\partial C^3} \mathcal{L}(y;C) |] < \infty.$$

(A5) The HD of underlying $p(x)$ is three times continuously differentiable and is bounded by a positive constant from below on the compact set E.

(A6) Let $a_{n_p} = (n_p h_{n_p}/\log n_p)^{-1/2} + h_{n_p}^2$ from (10) and $b_{n_q} = h_{n_q}^{-2}(n_q h_{n_q}/\log n_q)^{-1/2} + h_{n_q}^2$ from Lemma 1. We assume that $n_q/n_p = \mathcal{O}(1)$, $h_{n_q}^5/h_{n_p} = \scriptstyle\mathcal{O}(1)$.

(A7) The pricing errors ε_i are independent and identically distributed random variables.

(A1) is a bandwidth assumption for estimating SPD. We can undersmooth to reduce the bias. (A2) is the assumption on kernel function which facilitates the derivation of results. In a typical setting, h_{n_q} is chosen to be as in (A1), while h_{n_p} is chosen to be $\mathcal{O}(n_p^{-1/5})$, and h_{n_q} is larger than h_{n_p}. (A4) contains moment conditions defined via likelihood functions. (A5) is an assumption imposed on the smoothness of $p(x)$. In our empirical setting, $n_q = 715$, $n_p = 200$, thus for a typical data situations, (A6) is reasonable. The assumptions (A1) and (A2) ensure $a_{n_p}/b_{n_q} = \scriptstyle\mathcal{O}(1)$.

6.1 Proof of Lemma 2

Recall from Lemma 1 and (10) that

$$\sup_{x\in E} |\hat{p}(x) - p(x)| = \mathcal{O}\{\{\log n_q/(n_q h_{n_q})\}^{1/2} + h_{n_p}^2\} = \mathcal{O}(a_{n_p}),$$

$$\sup_{x\in E} |\hat{q}(x) - q(x)| = \mathcal{O}[h_{n_q}^{-2}\{\log n_q/(n_q h_{n_q})\}^{1/2} + h_{n_q}^2] = \mathcal{O}(b_{n_q}).$$

To determine the order of the EPK we linearize the ratio $q(x)/p(x)$.

$$\hat{\mathcal{K}}(x) - \mathcal{K}(x) = \frac{\hat{q}(x)}{\hat{p}(x)} - \frac{q(x)}{p(x)} = \frac{\hat{q}(x)p(x) - \hat{p}(x)q(x)}{p^2(x)} \cdot \frac{1}{1 + \frac{\hat{p}(x)-p(x)}{p(x)}}. \tag{25}$$

We decompose the first factor as $\hat{q}(x)p(x) - \hat{p}(x)q(x) = \{\hat{q}(x) - q(x)\}p(x) - \{\hat{p}(x) - p(x)\}q(x)$, while for the second factor we use the first order Taylor expansion. Putting together we obtain

$$\begin{aligned}\sup_{x\in E} |\hat{\mathcal{K}}(x) - \mathcal{K}(x)| &= \sup_{x\in E} \Big| \frac{\hat{q}(x) - q(x)}{p(x)} - \frac{\hat{p}(x) - p(x)}{p(x)} \cdot \frac{q(x)}{p(x)} \\ &\quad - \frac{\{\hat{q}(x) - q(x)\}\{\hat{p}(x) - p(x)\}}{p^2(x)} + \frac{\{\hat{p}(x) - p(x)\}^2}{p^2(x)} \cdot \frac{q(x)}{p(x)} \Big|.\end{aligned}$$

The first two elements are of order $\mathcal{O}(b_{n_q})$ and $\mathcal{O}(a_{n_p})$ respectively, while the last element is of order $\scriptstyle\mathcal{O}(a_{n_p})$. Summarizing we conclude that

$$\sup_{x\in E} |\hat{\mathcal{K}}(x) - \mathcal{K}(x)| = \mathcal{O}[\max\{a_{n_p}, b_{n_q}\}].$$

6.2 Proof of Theorem 2

The basic idea of the proof is to approximate the process

$$D_{n_q}(x) = n_q^{1/2} h_{n_q}^{5/2} \{\hat{\mathcal{K}}(x) - \mathcal{K}(x)\} / [\widehat{\mathrm{Var}}\{\hat{\mathcal{K}}(x)\}]^{1/2}$$

by a process with nonstochastic variance term, which will then be further approximated by a process that can be treated with the tools of Claeskens and Van Keilegom (2003). Here we have dropped for the simplicity of notation the q in n_q and the n_q in h_{n_q}. More precisely, we define, as first approximation,

$$D_n^{(1)}(x) \stackrel{\text{def}}{=} n^{1/2} h^{5/2} \{\hat{\mathcal{K}}(x) - \mathcal{K}(x)\} / \{\mathrm{Var}\{\hat{\mathcal{K}}(x)\}\}^{1/2},$$

where $\mathrm{Var}\{\hat{\mathcal{K}}(x)\}$ is given in (15). Lemma 2 ensures that the approximation by

$$n^{1/2} h^{5/2} \{\hat{q}(x) - q(x)\} / \{p(x) \mathrm{Var}\{\hat{\mathcal{K}}(x)\}^{1/2}\} \tag{26}$$

is uniformly of order $\mathcal{O}_p\{(\log n)^{-1/2}\}$. The process in equation (26) can be approximated as in Claeskens and Van Keilegom (2003) by

$$2! \exp(r\tau) h^2 f_X(x)^{-1/2} \mathrm{Var}\{\hat{\mathcal{K}}(x)\}^{-1/2} I\{C(x)\}^{-1/2} \sum_{i=0}^{3} f_X(x)^{-1/2} I\{C(x)\}^{-1/2} \{\boldsymbol{L}^{-1}\}_{3,i+1} A_{n,i}(x) \tag{27}$$

For the definition of the local Fisher information, $I\{C(x)\}$, the matrix $\boldsymbol{L}$ and the process $A_{ni}(x)$, we refer to Section 2, and Section 6. Define

$$Z_{ni}(x) \stackrel{\text{def}}{=} (nh)^{1/2} h^{-i} [I\{C(x)\} f_X(x)]^{-1/2} A_{ni}(x).$$

Then equation (27) can be written as

$$F_n(x) \stackrel{\text{def}}{=} 2! \exp(r\tau) h^2 \{f_X(x)\}^{-1/2} \mathrm{Var}\{\hat{\mathcal{K}}(x)\}^{-1/2} I\{C(x)\}^{-1/2} \sum_{i=0}^{3} h^i \{\boldsymbol{L}^{-1}\}_{3,i+1} Z_{ni}(x)$$

Please note that $\boldsymbol{L}$ is not a function of x as Claeskens and Van Keilegom (2003) erroneously write. Following their line of thoughts, we replace $Z_{ni}(x)$ (uniformly) by

$$Z'_{ni}(x) = h^{1/2} \int K_h(z-x) \left(\frac{z-x}{h}\right) dz$$

In order to apply corollary $A1$ of Bickel and Rosenblatt (1973), define the covariance function $r(x)$ of the Gaussian process $F_{n_q}(x)$, and we know that

$$\begin{aligned} r(x) &= \mathrm{Cov}(Z'_{nj}(x), Z'_{nj}(0)) \\ &= C_1 - C_2 |x|^2 + \mathcal{O}(|x|^2), \end{aligned}$$

for $x \in E$, where C_1 and C_2 are two constants, so the regularity conditions satisfies, the result follows.

Finally, we have to show that $\sup_{x\in E}|\widehat{\text{Var}}\{\hat{\mathcal{K}}(x)\}-\text{Var}\{\hat{\mathcal{K}}(x)\}|=\mathcal{O}_p(1)$.

$$\begin{aligned}&\sup_{x\in E}|\widehat{\text{Var}}\{\hat{\mathcal{K}}(x)\}-\text{Var}\{\hat{\mathcal{K}}(x)\}|\\ =&\sup_{x\in E}|\widehat{\text{Var}}\left\{\frac{\hat{q}(x)-q(x)}{p(x)}\right\}-\text{Var}\left\{\frac{\hat{q}(x)-q(x)}{p(x)}\right\}|+\mathcal{O}_p\{(nh)^{-(1/2+\alpha)}(\log n)^{1+\alpha}\},\end{aligned}$$

where $0<\alpha<1$.

According to corollary 2.1 in Claeskens and Van Keilegom (2003), for $j=3,k=3$,

$$\sup_{x\in E}|\widehat{\text{Var}}\{\hat{q}(x)\}-\text{Var}\{\hat{q}(x)\}|=\mathcal{O}_p\{(nh\log n)^{-1/2}\}.$$

So we have,

$$\sup_{x\in E}|\widehat{\text{Var}}\{\hat{\mathcal{K}}(x)\}-\text{Var}\{\hat{\mathcal{K}}(x)\}|=\mathcal{O}_p(1).$$

6.3 Expressions for the Semiparametric Estimator of SPD and Proof of Theorem 3

To proof the statement we show that $\sqrt{n_q h_{n_q}}(\hat{q}(x)-q(x))$ has asymptotically the same distribution as $\sqrt{n_q h_{n_q}}\{\hat{V}''(M)-V''(M)\}$ with proper scaling. Thus we drive the following equation.

$$\begin{aligned}\hat{q}(x)-q(x) &= e^{r\tau}S_t\frac{M^2}{x^2}\Big[\varphi\{\hat{d}_1(M)\}\big\{\sqrt{\tau}/2-\frac{\log(M)+r\tau}{\hat{V}(M)^2\sqrt{\tau}}\big\} \qquad (28)\\ &\quad -e^{-r\tau}\varphi\{\hat{d}_2(M)\}\big\{-\sqrt{\tau}/2-\frac{\log(M)+r\tau}{\hat{V}(M)^2\sqrt{\tau}}\big\}/M\Big]\{\hat{V}''(M)-V''(M)\}\\ &\quad +\mathcal{O}\{\hat{V}''(M)-V''(M)\}, \qquad (29)\end{aligned}$$

where $\hat{d}_i$ and $\hat{c}$ are the terms defined in Section 2.2 with $\hat{V}(M)$ replaced by the true function. We now describe how to derive (29) Taking the derivatives of $c(M_{it})$ with respect to moneyness (M) and noting that both $d_1(M_{it})$ and $d_2(M_{it})$ depend on M_{it} we obtain

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

Computing the first and second order differentials for d_1 and d_2 using the notation $V=\sigma(M)$, $V'=\partial\sigma(M)/\partial M$ and $V''=\partial^2\sigma(M)/\partial M^2$, we obtain

$$\begin{aligned}
\frac{\mathrm{d}d_1}{\mathrm{d}M} &= \frac{1}{MV\sqrt{\tau}}+\Big\{-\frac{\log(M)+r\tau}{V^2\sqrt{\tau}}+\sqrt{\tau}/2\Big\}V',\\
\frac{\mathrm{d}d_2}{\mathrm{d}M} &= \frac{1}{MV\sqrt{\tau}}+\Big\{-\frac{\log(M)+r\tau}{V^2\sqrt{\tau}}-\sqrt{\tau}/2\Big\}V',\\
\frac{\mathrm{d}^2d_1}{\mathrm{d}M^2} &= -\frac{1}{MV\sqrt{\tau}}\Big\{\frac{1}{M}+\frac{V'}{V}\Big\}+V''\Big\{\frac{\sqrt{\tau}}{2}-\frac{\log(M)+r\tau}{V^2\sqrt{\tau}}\Big\}\\
&\quad +V'\Big\{2V'\frac{\log(M)+r\tau}{V^3\sqrt{\tau}}-\frac{1}{MV^2\sqrt{\tau}}\Big\},\\
\frac{\mathrm{d}^2d_2}{\mathrm{d}M^2} &= -\frac{1}{MV\sqrt{\tau}}\Big\{\frac{1}{M}+\frac{V'}{V}\Big\}+V''\Big\{-\frac{\sqrt{\tau}}{2}-\frac{\log(M)+r\tau}{V^2\sqrt{\tau}}\Big\}\\
&\quad +V'\Big\{2V'\frac{\log(M)+r\tau}{V^3\sqrt{\tau}}-\frac{1}{MV^2\sqrt{\tau}}\Big\}.
\end{aligned}$$

To prove (29), we know from (22) and (21) that

$$q(x)-\hat{q}(x)=\left(\frac{\mathrm{d}^2c}{\mathrm{d}M^2}-\frac{\mathrm{d}^2\hat{c}}{\mathrm{d}M^2}\right)\left(\frac{M}{X}\right)^2+2\left(\frac{\mathrm{d}c}{\mathrm{d}M}\frac{M}{X^2}-\frac{\mathrm{d}\hat{c}}{\mathrm{d}M}\frac{M}{X^2}\right). \tag{30}$$

The stochastic terms involved are

$$\begin{aligned}
&\frac{\mathrm{d}^2c}{\mathrm{d}M^2}-\frac{\mathrm{d}^2\hat{c}}{\mathrm{d}M^2}=\Big\{\varphi(d_1)\frac{\mathrm{d}^2d_1}{\mathrm{d}M^2}-\varphi(\hat{d}_1)\frac{\mathrm{d}^2\hat{d}_1}{\mathrm{d}M^2}\Big\}-\Big\{\varphi(d_1)d_1\left(\frac{\mathrm{d}d_1}{\mathrm{d}M}\right)^2-\varphi(\hat{d}_1)\hat{d}_1\left(\frac{\mathrm{d}\hat{d}_1}{\mathrm{d}M}\right)^2\Big\}\\
&-\Big\{\frac{e^{-r\tau}\varphi(d_2)}{M}\frac{\mathrm{d}^2d_2}{\mathrm{d}M^2}-\frac{e^{-r\tau}\varphi(\hat{d}_2)}{M}\frac{\mathrm{d}^2\hat{d}_2}{\mathrm{d}M^2}\Big\}+\Big\{\frac{e^{-r\tau}\varphi(d_2)}{M}\frac{2}{M}\frac{\mathrm{d}d_2}{\mathrm{d}M}-\frac{e^{-r\tau}\varphi(\hat{d}_2)}{M}\frac{2}{M}\frac{\mathrm{d}\hat{d}_2}{\mathrm{d}M}\Big\}\\
&+\Big\{\frac{e^{-r\tau}\varphi(d_2)}{M}d_2\left(\frac{\mathrm{d}d_2}{\mathrm{d}M}\right)^2-\frac{e^{-r\tau}\varphi(\hat{d}_2)}{M}\hat{d}_2\left(\frac{\mathrm{d}\hat{d}_2}{\mathrm{d}M}\right)^2\Big\}-\Big\{\frac{2e^{-r\tau}\Phi(d_2)}{M^3}-\frac{2e^{-r\tau}\Phi(\hat{d}_2)}{M^3}\Big\}\\
&\stackrel{\text{def}}{=} g_{t1}-g_{t2}-g_{t3}+g_{t4}+g_{t5}-g_{t6}
\end{aligned}$$

and

$$\begin{aligned}
\frac{\mathrm{d}c}{\mathrm{d}M}-\frac{\mathrm{d}\hat{c}}{\mathrm{d}M} &= \Big\{\varphi(d_1)\frac{\mathrm{d}d_1}{\mathrm{d}M}-\varphi(\hat{d}_1)\frac{\mathrm{d}\hat{d}_1}{\mathrm{d}M}\Big\}-\Big\{e^{-r\tau}\frac{\varphi(d_2)}{M}\frac{\mathrm{d}d_2}{\mathrm{d}M}-e^{-r\tau}\frac{\varphi(\hat{d}_2)}{M}\frac{\mathrm{d}\hat{d}_2}{\mathrm{d}M}\Big\}\\
&\quad +\Big\{e^{-r\tau}\frac{\Phi(d_2)}{M^2}-e^{-r\tau}\frac{\Phi(\hat{d}_2)}{M^2}\Big\}\\
&\stackrel{\text{def}}{=} g'_{t1}-g'_{t2}+g'_{t3}.
\end{aligned}$$

Now we analyze each term obtaining,

$$
\begin{aligned}
g_{t1} &= \frac{\mathrm{d}^2 d_1}{\mathrm{d}M^2}\Big\{\varphi(d_1)-\varphi(\hat{d}_1)\Big\}-\varphi(\hat{d}_1)\Big\{\frac{\mathrm{d}^2\hat{d}_1}{\mathrm{d}M^2}-\frac{\mathrm{d}^2 d_1}{\mathrm{d}M^2}\Big\}\\
g_{t2} &= d_1\Big(\frac{\mathrm{d}d_1}{\mathrm{d}M}\Big)^2\Big\{\varphi(d_1)-\varphi(\hat{d}_1)\Big\}+\Big(\frac{\mathrm{d}d_1}{\mathrm{d}M}\Big)^2\varphi(\hat{d}_1)\Big\{d_1-\hat{d}_1\Big\}\\
&\quad+\hat{d}_1\varphi(\hat{d}_1)\Big\{\Big(\frac{\mathrm{d}d_1}{\mathrm{d}M}\Big)^2-\Big(\frac{\mathrm{d}\hat{d}_1}{\mathrm{d}M}\Big)^2\Big\},
\end{aligned}
$$

and similarly for $g_{t3}, g_{t4}, g_{t5}, g_{t6}, g'_{t1}, g'_{t2}, g'_{t3}$. The further analysis of the rate boils down to the analysis of the rates for $\frac{\mathrm{d}\hat{d}_1}{\mathrm{d}M}-\frac{\mathrm{d}d_1}{\mathrm{d}M}, \frac{\mathrm{d}^2 d_1}{\mathrm{d}M^2}-\frac{\mathrm{d}^2\hat{d}_1}{\mathrm{d}M^2}$, $d_1-\hat{d}_1$, $\Phi(d_1)-\Phi(\hat{d}_1)$ and similar terms for d_2, as $d_2(M)=(d_1(M)-\sigma(M))\sqrt{\tau}$.

By the mean value theorem

$$
\begin{aligned}
\Phi(d_1)-\Phi(\hat{d}_1) &= \varphi(d_0)(d_1-\hat{d}_1)\\
d_1-\hat{d}_1 &= \frac{\{\log(M)+r_t\tau\}(\hat{V}-V)}{V\hat{V}\sqrt{\tau}}+\sqrt{\tau}(V-\hat{V})/2\\
\frac{\mathrm{d}d_1}{\mathrm{d}M}-\frac{\mathrm{d}\hat{d}_1}{\mathrm{d}M} &= \frac{\hat{V}-V}{MV\hat{V}\sqrt{\tau}}+\Big\{\frac{(\log(M)+r\tau)(V^2-\hat{V}^2)}{V^2\hat{V}^2\sqrt{\tau}}+\sqrt{\tau}/2\Big\}V'\\
&\quad+\Big\{\frac{(\log(M)+r\tau)}{\hat{V}^2\sqrt{\tau}}+\sqrt{\tau}/2\Big\}(\hat{V}'-V'),\\
&= \mathcal{O}(\Big\{\frac{(\log(M)+r\tau)}{\hat{V}^2\sqrt{\tau}}+\sqrt{\tau}/2\Big\}(\hat{V}'-V'))\\
\frac{\mathrm{d}^2 d_1}{\mathrm{d}M^2}-\frac{\mathrm{d}^2\hat{d}_1}{\mathrm{d}M^2} &= -\Big\{\frac{1}{M\sqrt{\tau}}\frac{\hat{V}-V}{V\hat{V}}\frac{1}{M}+\frac{1}{M}\frac{\hat{V}-V}{V\hat{V}\sqrt{\tau}}\frac{V'}{V}+\frac{1}{M\hat{V}\sqrt{\tau}}\frac{V'(\hat{V}-V)+V(V'-\hat{V}')}{V\hat{V}M}\Big\}\\
&\quad+\Big\{V''-\hat{V}''\Big\}\Big\{\frac{\sqrt{\tau}}{2}-\frac{\log(M)+r\tau}{V^2\sqrt{\tau}}\Big\}+\hat{V}''\Big\{-\frac{(\log(M)+r\tau)(\hat{V}^2-V^2)}{V^2\hat{V}^2\sqrt{\tau}}\Big\}\\
&\quad+2\hat{V}^3\{V'^2-\hat{V}'^2\}\frac{\log(M)+r\tau}{V^3\hat{V}^3\sqrt{\tau}}+2\hat{V}'^2\{\hat{V}^3-V^3\}\frac{\log(M)+r\tau}{V^3\hat{V}^3\sqrt{\tau}}\\
&\quad-\Big\{V^2(V'-\hat{V}')\frac{1}{MV^2\hat{V}^2\sqrt{\tau}}+\hat{V}'(V^2-\hat{V}^2)\frac{1}{MV^2\hat{V}^2\sqrt{\tau}}\Big\}\\
&= \mathcal{O}(\Big\{V''-\hat{V}''\Big\}\Big\{\frac{\sqrt{\tau}}{2}-\frac{\log(M)+r\tau}{V^2\sqrt{\tau}}\Big\}).
\end{aligned}
$$

So the dominant term in the equation (30) is $S_t e^{r\tau}\left(\frac{\mathrm{d}^2 c}{\mathrm{d}M^2} - \frac{\mathrm{d}^2 \hat{c}}{\mathrm{d}M^2}\right)\left(\frac{M}{X}\right)^2$ and this term is dominated by g_{t1} and g_{t3}.

$$
\begin{aligned}
q(x)-\hat{q}(x) &= \mathcal{O}(S_t e^{r\tau}\left(\frac{M}{X}\right)^2\left\{\varphi(d_1)\frac{\mathrm{d}^2 d_1}{\mathrm{d}M^2} - \varphi(\hat{d}_1)\frac{\mathrm{d}^2 \hat{d}_1}{\mathrm{d}M^2}\right\} \\
&\quad -\left(\frac{M}{X}\right)^2 S_t e^{r\tau}\left\{\frac{e^{-r\tau}\varphi(d_2)}{M}\frac{\mathrm{d}^2 d_2}{\mathrm{d}M^2} - \frac{e^{-r\tau}\varphi(\hat{d}_2)}{M}\frac{\mathrm{d}^2 \hat{d}_2}{\mathrm{d}M^2}\right\}) \\
&= \mathcal{O}(-S_t e^{r\tau}\left(\frac{M}{X}\right)^2 \varphi(\hat{d}_1)\left\{\frac{\mathrm{d}^2 \hat{d}_1}{\mathrm{d}M^2} - \frac{\mathrm{d}^2 d_1}{\mathrm{d}M^2}\right\} \\
&\quad +S_t e^{r\tau} e^{-r\tau}\left(\frac{M}{X}\right)^2 \varphi(\hat{d}_2)/M\left\{\frac{\mathrm{d}^2 \hat{d}_2}{\mathrm{d}M^2} - \frac{\mathrm{d}^2 d_2}{\mathrm{d}M^2}\right\}) \\
&= \mathcal{O}(-S_t e^{r\tau}\left(\frac{M}{X}\right)^2 \varphi(\hat{d}_1)\left\{\hat{V}'' - V''\right\}\left\{\frac{\sqrt{\tau}}{2} - \frac{\log(M)+r\tau}{V^2\sqrt{\tau}}\right\} \\
&\quad +S_t\left(\frac{M}{X}\right)^2 \varphi(\hat{d}_2)/M\left\{\hat{V}'' - V''\right\}\left\{\frac{-\sqrt{\tau}}{2} - \frac{\log(M)+r\tau}{V^2\sqrt{\tau}}\right\}).
\end{aligned}
$$

6.4 Proof of Theorem 4

Firstly, the expansion under (A8) in Corollary 2.1 in Claeskens and Van Keilegom (2003) is still valid with remainder term,

$$
\mathbf{H}_{n_q}\{\hat{\mathbf{C}}(x)-\mathbf{C}(x)\} = \mathbf{J}(x)^{-1}\mathbf{H}_{n_q}^{-1}\mathbf{A}_{n_q}(x) + \mathbf{R}_{n_q}(x), \tag{31}
$$

where using $I\{C(x)\}$, $\mathbf{L}$ defined in Section 2,

$$
\mathbf{J}(x) \stackrel{\text{def}}{=} f_X(x) I\{C(x)\}\mathbf{L} \tag{32}
$$

and

$$
\begin{aligned}
\mathbf{R}_{n_q}(x) &= -\mathbf{B}_{n_q}^{-1}(x)\mathbf{J}^{-1}(x)\{\mathbf{J}(x)+\mathbf{B}_{n_q}(x)\}\mathbf{H}_{n_q}^{-1}\mathbf{A}_{n_q}(x) && (33)\\
&\quad +\{\mathbf{B}_{n_q}^{-1}(x)\mathbf{J}^{-1}(x)\mathbf{B}_{n_q}(x) - \mathbf{J}^{-1}(x)\}\mathbf{H}_{n_q}^{-1}\mathbf{A}_{n_q}(x) - \mathbf{B}_{n_q}^{-1}(x)\mathbf{D}_{n_q}(x). && (34)
\end{aligned}
$$

Define

$$\mathbf{D}_{n_q}(x) \stackrel{\text{def}}{=} \frac{1}{2n_q}\sum_{i=1}^{n_q} K_{h_{n_q}}(X_i - x)\frac{\partial^3}{\partial C^3}\log f\{Y_i;\xi(x,X_i)\}(\hat{\mathbf{C}}(x)-\mathbf{C}(x))$$
$$\mathbf{X}_i^\top \mathbf{X}_i(\hat{\mathbf{C}}(x)-\mathbf{C}(x))\mathbf{H}_{n_q}^{-1}\mathbf{X}_i$$

$$\mathbf{B}_{n_q}(x) \stackrel{\text{def}}{=} \frac{1}{n_q}\sum_{i=1}^{n_q} K_{h_{n_q}}(X_i - x)\frac{\partial^2}{\partial C^2}\log f\{Y_i;C(x,X_i)\}\mathbf{H}_{n_q}^{-1}\mathbf{X}_i(\mathbf{H}_{n_q}^{-1}\mathbf{X}_i)^\top.$$

where $\xi(x,X_i)$ is in between $C(x,X_i)$ and $\hat{C}(x,X_i)$. Masry (1996) proved that the remainder term is theoretically ignorable uniformly under strong mixing conditions and under (A8) it can be shown that

$$\sup_{x\in E}\mathbf{R}_{n_q j}(x) = \mathcal{o}_p((h_{n_q}\log n_q/n_q)^{1/2}). \tag{35}$$

Thus we concentrate on the scaled first-order term

$$S_{nj} \stackrel{\text{def}}{=} (n_q h_{n_q})^{1/2} h_{n_q}^{-j}\{g(x)\}^{-1/2}\mathbf{A}_{n_q j}(x)$$

with $g(x) \stackrel{\text{def}}{=} I(C(x))f_X(x)$.

Recall that with a Gaussian likelihood, the components involved in $\mathbf{J}(x)^{-1}\mathbf{H}_n^{-1}\mathbf{A}_n(x)$ are

$$\begin{aligned} S_{n_q j} &= -(n_q)^{-1/2}h_{n_q}^{1/2}(g(x))^{-1/2}\sum_{i=1}^{n_q}K_h(X_i-x)\{(X_i-x)^j/h_{n_q}^j\}\{Y_i - C(x,X_i)\}/\sigma^2(x) \\ &= T_{n_q 1}(x) + T_{n_q 2}(x), \qquad j = 0,\ldots,d, \end{aligned}$$

where

$$T_{n_q 1} \stackrel{\text{def}}{=} -(n_q)^{-1/2}h_{n_q}^{1/2}g(x)^{-1/2}\sum_{i=1}^{n_q}K_{h_{n_q}}(X_i-x)\{(X_i-x)^j/h_{n_q}^j\}(C(x)-C(x,X_i))/\sigma^2(x),$$

$$T_{n_q 2} \stackrel{\text{def}}{=} -(n_q)^{-1/2}h_{n_q}^{1/2}g(x)^{-1/2}\sum_{i=1}^{n_q}K_{h_{n_q}}(X_i-x)\{(X_i-x)^j/h_{n_q}^j\}(\sigma(X_i)\varepsilon_i)/\sigma^2(x).$$

Then we rely on an easy modification of Theorem 2.4 in Liu and Wu (2010), which implies

$$\sup_{x\in E}|T_{n_q 1}(x)| = \mathcal{O}_p\left(h_{n_q}\sqrt{\log n_q} + \frac{\mathcal{Z}_{n_q}^{1/2}h_{n_q}^{3/2}}{n_q^{1/2}}\right), \tag{36}$$

$$\sup_{x\in E}\left|T_{n_q 2}(x) - \frac{h_{n_q}^{1/2}}{(n_q f_X(x))^{1/2}}\sum_{i=1}^{n_q}K_{h_{n_q}}(X_i-x)(X_i-x)^j\varepsilon_i/h_{n_q}^j\right| = \mathcal{O}_p(h_{n_q}\sqrt{\log n_q}). \tag{37}$$

We need a linear combination of the component $S_{n_q j}$ to analyze $\hat{q}(x) - q(x)$, in particular $\sum_{j=0}^{p} \mathbf{L}_{3,j+1}^{-1} S_{n_q j}$.

The rest of the proof then follows from the modification of the proof of Proposition 2.1 in Liu and Wu (2010) and is similar to Theorems 4 and 5 in Zhao and Wu (2008).

Received April 15, 2011; revised December 4, 2013; accepted January 8, 2014.

REFERENCES

Aït-Sahalia, Y., and J. Duarte. 2003. Nonparametric Option Pricing under Shape Restrictions. *Journal of Econometrics* 116: 9–47.

Aït-Sahalia, Y., and A. W. Lo. 1998. Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices. *The Journal of Finance* 53: 499–547.

Aït-Sahalia, Y., and A. W. Lo. 2000. Nonparametric Risk Management and Implied Risk Aversion. *Journal of Econometrics* 94: 9–51.

Bickel, P. J., and M. Rosenblatt. 1973. On Some Global Measures of the Deviations of Density Function Estimates. *The Annals of Statistics* 1: 1071–1095.

Breeden, D., and R. Litzenberger. 1978. Prices of State-Contingent Claims Implicit in Options Prices. *The Journal of Business* 51: 621–651.

Brown, D. P., and J. C. Jackwerth. 2004. The Pricing Kernel Puzzle: Reconciling Index Option Data and Economic Theory. *Manuscript*.

Carroll, R., D. Ruppert, and A. Welsh. 1998. Local Estimating Equations. *Journal of American Statistical Association* 93: 214–227.

Chabi-Yo, Y., R. Garcia, and E. Renault. 2008. State Dependence can Explain the Risk Aversion Puzzle. *Review of Financial Studies* 21: 973–1011.

Chetverikov, D. (2012). Testing Regression Monotonicity in Econometric Models. *Technical report*, MIT.

Christoffersen, P., S. Heston, and K. Jacobs. 2011. A GARCH Option Model with Variance-Dependent Pricing Kernel. *Technical report*, University of Maryland.

Claeskens, G., and I. Van Keilegom. 2003. Bootstrap Confidence Bands for Regression Curves and their Derivatives. *The Annals of Statistics* 31: 1852–1884.

Cleveland, W. 1979. Robust Locally Weighted Regression and Smoothing Scatter-plots. *Journal of the American Statistical Association* 74: 829–836.

Cochrane, J. H. 2001. *Asset Pricing*. Princeton and Oxford: Princeton University Press.

Eubank, R. L., and P. L. Speckman. 1993. Confidence Bands in Nonparametric Regression. *Journal of the American Statistical Association* 88: 1287–1301.

Fan, J. 1992. Design-Adaptive Nonparametric Regression. *Journal of the American Statistical Association* 87: 998–1004.

Fan, J. 1993. Local Linear Regression Smoothers and their Minimax Efficiencies. *The Annals of Statistics* 21: 196–216.

Fengler, M. 2005. *Semiparametric Modeling of Implied Volatility*. Berlin: Springer Verlag.

Ghosal, S., A. Sen, and A. van der Vaart. 2000. Testing Monotonicity of a Regression Function. *The Annals of Statistics* 28: 1054–1082.

Gilchrist, S., and E. Zakrajšek. 2012. Credit Spreads and Business Cycle Fluctuations. *American Economic Review* 102: 1692–1720.

Golubev, Y., W. Härdle, and R. Timofeev. 2014. Testing Monotonicity of Pricing Kernels. *Advances in Statistical Analysis* 98: 305–326.

Grith, M., W. Härdle, and J. Park. 2013. Shape Invariant Modelling Pricing Kernels and Risk Aversion. *Journal of Financial Econometrics* 11: 370–399.

Hall, P. 1991. *Edgeworth Expansions for Nonparametric Density Estimators, with Applications*, Vol. 22. New York: Academic Press.

Härdle, W. 1989. Asymptotic Maximal Deviation of *M*-smoothers. *Journal of Multivariate Analysis* 29: 163–179.

Härdle, W., and J. Marron. 1991. Bootstrap Simultaneous Error Bars for Nonparametric Regression. *The Annals of Statistics* 19: 778–796.

Heaton, J., and D. Lucas. 1992. The Effects of Incomplete Insurance Markets and Trading Costs in a Consumption-based Asset Pricing Model. *Journal of Economic Dynamics and Control* 16: 601–620.

Jackwerth, J., and M. Rubinstein. 1996. Recovering Probability Distributions from Option Prices. *Journal of Finance* 51: 1611–1631.

Jackwerth, J. 2000. Recovering Risk Aversion from Option Prices and Realized Returns. *Review of Financial Studies* 13: 433–451.

Jagannathan, R., and Z. Wang. 1996. The Conditional CAPM and the Cross-Section of Expected Returns. *Journal of Finance* 51: 3–53.

Jones, M. C., O. Linton, and J. P. Nielson. 1995. A Simple Bias Reduction Method for Density Estimations. *Biometrika* 82: 327–38.

Kahneman, D., and A. Tversky. 1979. Prospect Theory: An Analysis of Decision Under Risk. *Econometrica* 47: 263–291.

Lee, S., O. Linton, and Y. Whang. 2009. Testing for Stochastic Monotonicity. *Econometrica* 77: 585–602.

Lepski, O. V., and V. G. Spokoiny. 1997. Optimal Pointwise Adaptive Methods in Nonparametric Estimation. *The Annals of Statistics* 25: 2512–2546.

Liero, H. 1982. On the Maximal Deviation of the Kernel Regression Function Estimate. *Mathematische Operationsforschung und Statistik, Serie Statistics* 13: 171–182.

Liu, W., and W. Wu. 2010. Simultaneous Nonparametric Inference of Time Series. *The Annals of Statistics* 38: 2388–2421.

Masry, E. 1996. Multivariate Local Polynomial Regression for Time Series: Uniform Strong Consistency and Rates. *Journal of Time Series Analysis* 17: 571–599.

Renault, E. (1997). "Econometric Models of Option Pricing Errors." In D. M. Kreps and K. F. Wallis (eds.), *Proceedings the Seventh World Congress of the Econometric Society*, Econometric Society Monographs. London: Cambridge University Press, pp. 223–278.

Rookley, C. 1997. Fully Exploiting the Information Content of Intra Day Option Quotes: Applications in Option Pricing and Risk Management.

Rosenberg, J., and R. F. Engle. 2002. Empirical Pricing Kernels. *Journal of Financial Economics* 64: 341–372.

Rubinstein, M. 1994. Implied Binomial Trees. *Journal of Finance* 49: 771–818.

Ruppert, D. and M. Wand. 1994. Multivariate Locally Weighted Least Squares Regression. *The Annals of Statistics* 22: 1346–1370.

Vieu, P. 1993. Nonparametric Regression: Optimal Local Bandwidth Choice. *Journal of the Royal Statistical Society, Series B* 53: 453–464.

Wang, K. 2002. Nonparametric Tests of Conditional Mean-Variance Efficiency of a Benchmark Portfolio. *Journal of Empirical Finance* 9: 133–169.

Wang, K. 2003. Asset Pricing with Conditioning Information: a New Test. *Journal of Finance* 58: 161–196.

Xia, Y. 1998. Bias-Corrected Confidence Bands in Nonparametric Regression. *Journal of the Royal Statistical Society: Series B* 4: 797–811.

Yuan, M. 2009. State Price Density Estimation via Nonparametric Mixtures. *Journal of Applied Statistics* 3: 963–984.

Zhao, Z., and W. Wu. 2008. Confidence Bands in Nonparametric Time Series Regression. *The Annals of Statistics* 36: 1854–1878.