Journal of Econometrics 179 (2014) 1–15

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

On implied volatility for options—Some reasons to smile and more to correct

Song Xi Chen [a,*], Zheng Xu [b]

[a] Guanghua School of Management and Center for Statistical Science, Peking University, China

[b] Department of Statistics, Iowa State University, United States

ARTICLE INFO

Article history:
Received 6 February 2012
Received in revised form
3 October 2013
Accepted 7 October 2013
Available online 23 October 2013

JEL classification:
C50
C58

Keywords:
Bias correction
Implied volatility
Kernel estimator
Pricing errors

ABSTRACT

We analyze the properties of the implied volatility, the commonly used volatility estimator by direct option price inversion. It is found that the implied volatility is subject to a systematic bias in the presence of pricing errors, which makes it inconsistent to the underlying volatility. We propose an estimator of the underlying volatility by first estimating nonparametrically the option price function, followed by inverting the nonparametrically estimated price. It is shown that the approach removes the adverse impacts of the pricing errors and produces a consistent volatility estimator for a wide range of option price models. We demonstrate the effectiveness of the proposed approach by numerical simulation and empirical analysis on S&P 500 option data.

1. Introduction

The implied volatility is an important concept associated with option contracts written on an underlying asset. It is calculated by inverting an option price via the Black–Scholes price formula (Black and Scholes, 1973; Merton, 1973). As the Black–Scholes pricing model assumes a constant volatility in the return of the underlying asset, it is expected that the implied volatility would be largely constant across the moneyness and the time to maturity. However, in numerous empirical studies, the implied volatility show sharp differences across moneyness and time to maturity, displaying either the so-called volatility smile or sneer. Rubinstein (1994) and Bakshi et al. (1997) found that for S&P 500 options the implied volatility exhibited a smile pattern and the extend of the smile changed along with the time to maturity. Dumas et al. (1998) found that after the 1987 stock market crash the implied volatility of S&P 500 options was monotone with respect to the moneyness, displaying the so-called volatility sneer. See also Bates (1996, 2000) for the analyses that provide further insights.

Understanding the implied volatility patterns has been a focal point in the theory and practice of pricing options. A common explanation to the volatility smile or sneer is the lack of fit of the Black–Scholes framework; see Macbeth and Merville (1979), Rubinstein (1985) and Mayhew (1995) for comprehensive studies. One consensus was that the Black–Scholes model performs reasonably well for at-the-money short-maturity options, which motivated the choice of using short maturity at-the-money options in the implied volatility estimation.

Alternative models have been proposed for the underlying asset process to produce certain forms of the implied volatility, which include the deterministic volatility model by Dupire (1994), Derman and Kani (1994) and Rubinstein (1994); Merton (1976)'s jump diffusion model and Hull and White (1987) and Heston (1993)'s stochastic volatility model. See Bates (1996, 2000) and Bakshi et al. (1997) for comprehensive pricing frameworks that can accommodate the stochastic volatility, the stochastic interest rate and jumps.

The existing practice of inverting one option price at a time in the implied volatility calculation is based on the assumption that the option prices are observed without errors. However, pricing errors are widely present in option data, which are due to a range of causes including the bid–ask spread, non-synchronicity between the option and the spot markets, the discreteness in the quoted prices and other random errors. The pricing errors may also be due to a lack of consensus among the market participants on the value of options. This is especially the case for deep-in-the-money or deep-out-of-the-money options. The errors are particularly pro-

* Corresponding author. Tel.: +86 10 62750427; fax: +86 10 62753590.
E-mail address: csx@gsm.pku.edu.cn (S.X. Chen).

0304-4076/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jeconom.2013.10.007

nounced for out-of-the-money calls or puts as shown in our empirical study on the S&P 500 option data. The practitioners are aware of the errors and have used largely at-the-money options in the implied volatility calculation, hoping that the impacts of the pricing errors would be the smallest there.

The effects of the pricing errors on the implied volatility have been considered in the literature. For foreign exchange future options, Jorion (1995) considered the measurement errors by regressing the realized volatility against the implied volatility in an effort to evaluate the predictive power of the implied volatility. He showed that the errors can distort the statistical inference. Hentschel (2003) evaluated the impacts of the measurement errors in variables via the first-order approximation to the option pricing function, which is used to quantify the impacts of the error on the implied volatility estimation. Hentschel hinted via a scenario analysis that even when the Black–Scholes formula is correct, the measurement error may induce a smile in the implied volatility. Aït-Sahalia and Lo (1998) included pricing errors in option prices in their nonparametric estimation of the state-price density based on the options prices. They also proposed a kernel smoothing of the implied volatility volatility estimator. Bates (1996, 2000) and Bakshi et al. (1997) have also included the pricing errors in their analyses.

In this paper, we first analyze the properties of the conventional implied volatility estimator and show that it is generally biased. Our analysis reveals that the bias depends on the moments of the errors and the derivatives of the underlying pricing function with respective to the price, which provides more statistical insights beyond (Jorion, 1995; Hentschel, 2003). We show that the pricing errors can cause implied volatility inconsistent even under a correct pricing model, for instance the Black–Scholes.

The main purpose of this paper is to develop a consistent implied volatility estimator that automatically corrects for the pricing errors for a wide range of option pricing models. The proposed approach first estimates nonparametrically the option price function by the kernel smoothing method (Härdle, 1990; Fan and Gijbels, 1996). Then, we invert the kernel option price estimate instead of a raw price. It is shown that this implied kernel (IK) implied volatility estimator is consistent for general option pricing models including the Black–Scholes framework. The kernel smoothing method has been employed in Aït-Sahalia and Lo (1998) for estimating the risk-neutral state price density and for smoothing the implied volatilities; see also Aït-Sahalia and Duarte (2003) for a state price density estimator that respects the monotonicity and convexity of option price functions. Fan and Mancini (2009) employed the smoothing technique to develop a model-guided nonparametric approach to estimate the state price survival function. Our proposed volatility estimator is closely related to the kernel smoothed volatility estimator of Aït-Sahalia and Lo (1998), except that we reverse the order of the kernel smoothing and the volatility inversion by engaging in the kernel smoothing to estimate the price function first and then carry out the volatility inversion. This re-ordering makes the proposed approach consistent.

The paper is organized as follows. Section 2 provides the basics on option pricing and implied volatility. The effects of pricing errors on the implied volatility is evaluated in Section 3. Section 4 discusses the stylish patterns of the implied volatility due to the pricing errors. Section 5 proposes the implied kernel estimator and presents its properties. Simulation studies are reported in Section 6. An empirical study on the S&P 500 option data is reported in Section 7.

2. Implied volatility

A European call (put) option gives the owner of the option contract the right, but not the obligation, to buy (sell) a security at a pre-specified strike price on the expiration date (Hull, 1999). The celebrated works of Black and Scholes (1973) and Merton (1973) lay down the foundation of the option pricing theory. Their framework is rooted on the conditions of market equilibrium, frictionless market, and no arbitrage opportunity, which constitute most of the qualifications for the option being "rational" by Merton (1973). Specifically, the price of the underlying asset S_t follows a geometric Brownian motion

$$dS_t = \mu S_t dt + \sigma S_t dB_t \tag{2.1}$$

where μ and σ are respectively the mean and the volatility of the instantaneous relative return, and B_t is the standard Brownian Motion. By solving a partial differential equation, they derived the following pricing formula for a call that has a strike price K, the time to maturity τ when the underlying price and the interest rate are S and r respectively:

$$C_{BS}(\sigma; S, K, \tau, r) = S\Phi(d_1) - Ke^{-r\tau}\Phi(d_2), \tag{2.2}$$

where Φ is the standard normal distribution function,

$$d_1 = \frac{ln\left(\frac{S}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)(\tau)}{\sigma\sqrt{\tau}} \quad \text{and} \quad d_2 = d_1 - \sigma\sqrt{\tau}.$$

Merton (1973) provided a systematic treatment on the so-called "rational" options and allow time-dependent volatility σ_t in (2.1). He also derived general properties of "rational" option prices, including one that the price should be bounded within

$$[\max(S - Ke^{-r\tau}, 0), S]. \tag{2.3}$$

Let $X = (S, K, \tau, r)$ be the covariate in a European call. For a general call price function $C(X)$, which may differ from the Black–Scholes (BS) price, we define the volatility function implied by the Black–Scholes as

$$\sigma(X) \equiv C_{BS}^{-1}\{C(X); X\}, \tag{2.4}$$

where C_{BS}^{-1} is the inverse of the BS price C_{BS} with respect to σ. Since $C(X)$ and $C_{BS}(X)$ share the same bounds in (2.3), $C(X)$ can always be inverted by C_{BS}^{-1}, and hence $\sigma(X)$ in (2.4) is always defined. As the BS price is monotone with respect to σ, any two different pricing formulae always lead to two different implied volatility functions. Thus, $\sigma(X)$ identifies uniquely the underlying price $C(X)$, and vice versa.

Let Y be the observed price of a call with covariates X. Regardless of what the underlying price $C(X)$ might be, the conventional implied volatility of Y at X is defined as

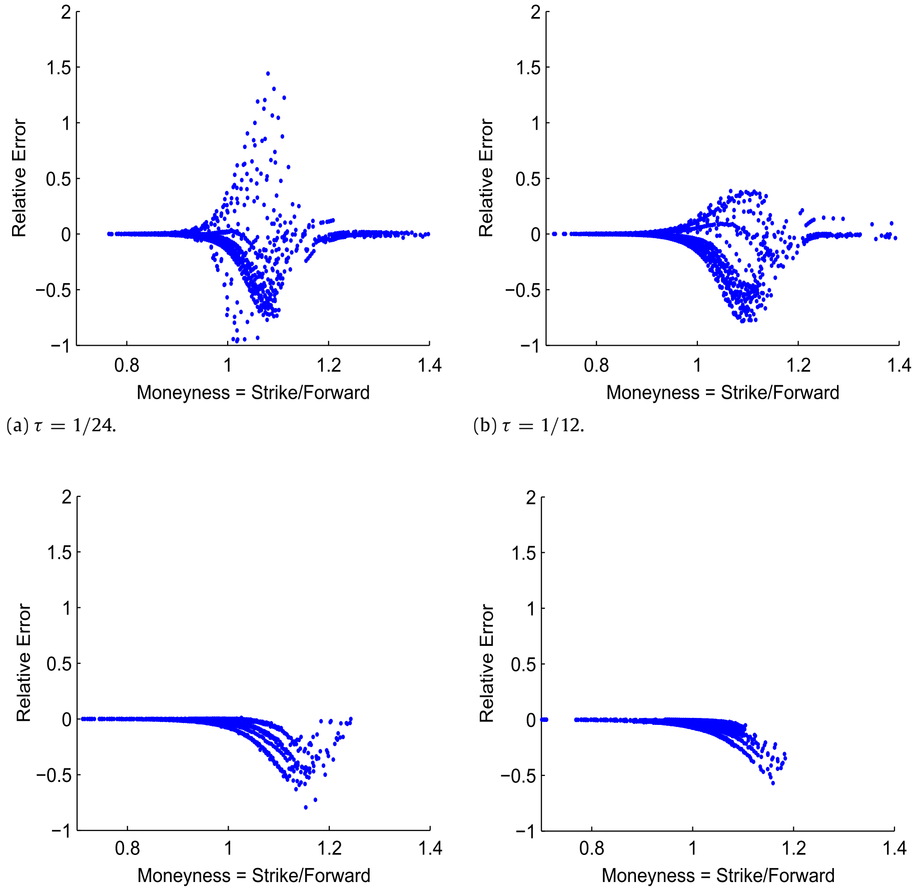
$$\hat{\sigma}_I(Y; X) \equiv C_{BS}^{-1}(Y; X).$$

Since the BS price is monotone with respect to σ, $\hat{\sigma}_I(Y; X)$ is available as long as $Y \in [\max(S - Ke^{-r\tau}, 0), S]$. When Y is outside this interval, the implied volatility cannot be obtained. We call such a case a truncation, which happens more likely for deep-in-the-money calls ($K/S \ll 1$), where the allowable range for Y narrows. A major cause for truncation is the pricing errors, an issue we will discuss extensively later.

3. Effects of pricing errors

Suppose at a given time period in a market there are n option contracts with price Y_i at $X_i = (S_i, K_i, \tau_i, r_i)$ for the ith contract and $i = 1, \ldots, n$. Let $C(X)$ be the "agreed" price among the market participants. The form of $C(x)$ is likely to be unknown but is assumed to be "rational" as specified in Merton (1973). Empirically, the observed price Y_i is rarely exactly the underlying price $C(X_i)$ but rather

$$Y_i = C(X_i) + \epsilon_i, \tag{3.1}$$



(a) $\tau = 1/24$.
(b) $\tau = 1/12$.
(c) $\tau = 1/6$.
(d) $\tau = 1/4$.

Fig. 1. Nonparametrically estimated relative pricing errors vs moneyness and maturity (τ) for S&P 500 call option data in the month from August 15, 2008 for the moneyness and maturity.

where ϵ_i is a pricing error and is assumed to satisfy

$$E(\epsilon_i|X_i) = 0. \tag{3.2}$$

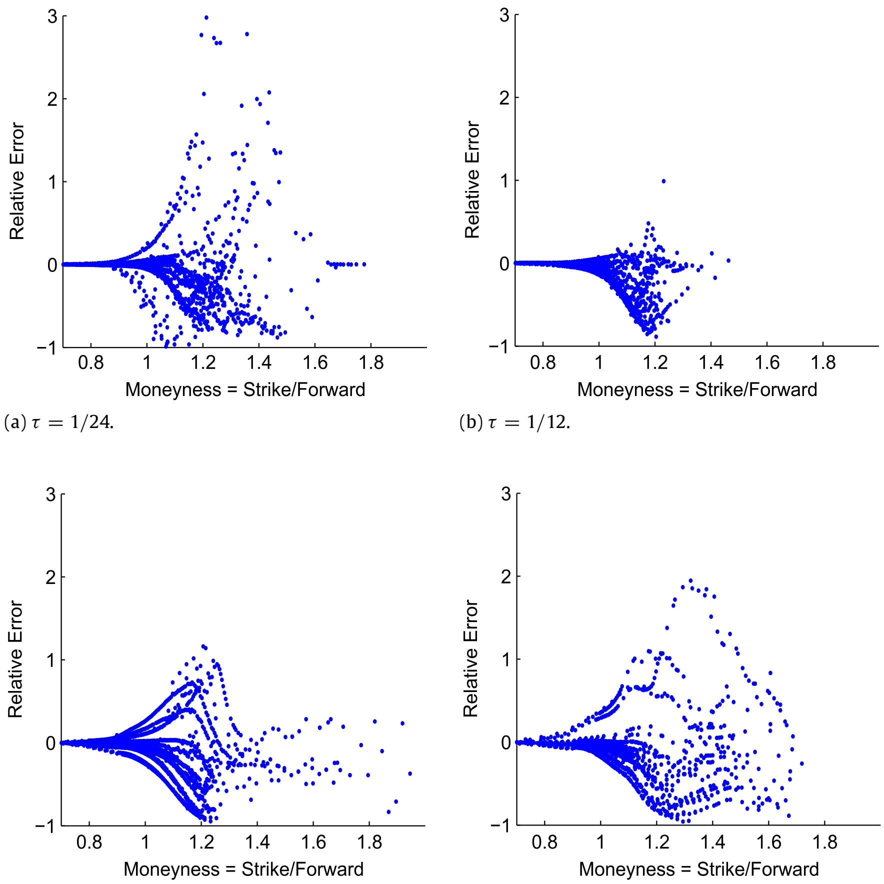
Model (3.1) has been considered in Aït-Sahalia and Lo (1998) in proposing a nonparametric estimator for the risk-neutral state space density. Bates (1996, 2000) has also incorporated them in his analyses on the behavior of option pricing post the 1987 crash.

The pricing error is largely present in the empirical prices of options, which is most eminently reflected in the bid–ask spreads, the discreteness in the quoted prices as opposed to the continuous $C(X_i)$ and the market dis-synchronization between the derivative and the spot markets. The error can also be understood as the natural fluctuations around the fair price $C(X)$, as reflected in the random movement of the bid and ask prices.

To demonstrate the presence of the pricing errors, we display in Figs. 1 and 2 the estimated relative pricing errors $\hat{C}^{-1}(X_i)\{Y_i - \hat{C}(X_i)\}$ plotted against the moneyness at different maturities for S&P 500 call option data for the month *before* and the month *after* September 15, 2008, the day Lehman Brothers declared bankruptcy. In the above, $\hat{C}(X_i)$ is a nonparametric kernel estimate to $C(X_i)$ whose details will be given in the next section. We also studied the estimated pricing errors for puts for the two months, which showed similar evidence of the errors. For both calls and puts, the amount of the pricing errors was the biggest for the short maturity out-of-the-money options. The relative errors became smaller but more negative as the time to maturity increased, indicating an under-pricing for longer maturity out-of-the-money options. We also observe that the pricing error was smaller for in-the-money options, partly because of higher underlying price $\hat{C}(X_i)$. Most importantly, we observed that the pricing errors for the short maturity at-the-money options were not negligible as commonly believed.

To illustrate the impacts of the errors, we assume that the BS price is correct and plot in Fig. 4 $C_{BS}^{-1}\{C_{BS}(\sigma; X)(1+\delta), X\}$ against the moneyness with the standardized price $S = 1$ and $\sigma = 0.3$, where δ denotes the relative pricing error. The baseline relative pricing error δ is displayed in Fig. 3, enumerated based on the anonymous referee's suggestion which was based on a one-factor stochastic volatility model with price jumps with parameters that reproduced a realistic IV curve based on the average bid–ask spread of the S&P 500 option data, basically the model and estimation approach used in Pan (2002). We multiply the baseline δ by ± 0.5, ± 0.8 and ± 1 at four different maturities $\tau = 1/12, 1/6, 1/4$ and $1/3$, respectively. Fig. 4 shows the impacts of the pricing error on the estimated implied volatility (IV). The IV curves display substantial deviation from the underlying volatility ($\sigma = 0.3$) for the out-of-the money (OTM) options at all maturities. Although the IV curve deviates little from the underlying $\sigma = 0.3$ for ITM calls with longer maturity, at the short maturity of $\tau = 1/12$, there are also noticeable impacts of the pricing error for in-the-money (IMT) calls.

The pricing errors can cause Y_i fall outside $[\max(S_i - K_i e^{-r\tau_i}, 0), S_i]$ so that the implied volatility inversion is not attainable. While Y_i is hardly larger than S_i, it is likely that Y_i may be below $S_i - K_i e^{-r\tau_i}$ for deep-in-the-money options ($M_i = K_i/S_i \ll 1$) as smaller strikes make the lower limit larger and hence more likely



(a) $\tau = 1/24$.

(b) $\tau = 1/12$.

(c) $\tau = 1/6$.

(d) $\tau = 1/4$.

Fig. 2. Nonparametrically estimated relative pricing errors vs moneyness and maturity (τ) for S&P 500 call option data in the month from September 15, 2008 for the moneyness and maturity.

for a price to fall below. For the S&P 500 option data, the percentages of the prices fell below the lower limit for the four quarters starting on March 15, June 15, September 15 and December 15 2008 were 9.95%, 6.25%, 4.53% and 5.03% respectively. A common treatment is to ignore such prices. However, this may cause the remaining errors biased such that $E(\epsilon_i|X_i) \neq 0$, which can influence the behavior of the implied volatility as demonstrated shortly.

Without loss of generality, we assume the price function $C(X)$ is homogeneous of degree 1 (Merton, 1973) such that $C(S_1, K_1, \tau, r) = qC(S_1/q, K_1/q, \tau, r)$ for any positive constant q. Thus, we standardize Model (3.1) by dividing S_i to obtain

$$\tilde{Y}_i = C(1, M_i, \tau_i, r_i) + \tilde{\epsilon}_i \tag{3.3}$$

where $\tilde{Y}_i \equiv Y_i/S_i$ is the standardized price, $M_i = K_i/S_i$ is the moneyness and $\tilde{\epsilon}_i \equiv \epsilon_i/S_i$ is the standardized residual. Let $Z_i \equiv (M_i, \tau_i, r_i)$ and $C(Z_i) = C(1, M_i, \tau_i, r_i)$. Then, (3.3) may be written as

$$\tilde{Y}_i = C(Z_i) + \tilde{\epsilon}_i. \tag{3.4}$$

As noted in Aït-Sahalia and Lo (1998), (3.4) is a model with one fewer covariate than X_i. The "standardized residual" $\tilde{\epsilon}_i$ may still depend on X_i. In particular, we note that

$$E(\tilde{\epsilon}_i|X_i) = \frac{E(\epsilon_i|X_i)}{S_i} \quad \text{and} \quad Var(\tilde{\epsilon}_i|X_i) = \frac{V(X_i)}{S_i^2} = \tilde{V}(X_i).$$

Since the Black–Scholes formula is homogeneous of degree 1, $C_{BS}(\sigma; x) = sC_{BS}(\sigma; z)$ and C_{BS}^{-1} is scale-invariant in the sense that $C_{BS}^{-1}(y; x) = C_{BS}^{-1}(\tilde{y}; z)$ for $\tilde{y} = y/s$. Let $C_{BS}^{-1}(\cdot; Z_i)$ be the inverse function with respect to σ at Z_i. Then, $\hat{\sigma}_I(X_i) = C_{BS}^{-1}(Y_i; X_i) = C_{BS}^{-1}(\tilde{Y}_i; Z_i) = \hat{\sigma}_I(Z_i)$. Thus, our discussion can be focused on $\hat{\sigma}_I(Z_i)$.

As $C_{BS}^{-1}\{C(Z_i); Z_i\}$ is infinitely differentiable with respect to C, and if ϵ_i has finite third moment, we have by Taylor expansion

$$\begin{aligned}\hat{\sigma}_I(Z_i) &= C_{BS}^{-1}\{C(Z_i) + \tilde{\epsilon}_i; Z_i\} \\ &= \sigma(Z_i) + \frac{\partial C_{BS}^{-1}\{C(Z_i); Z_i\}}{\partial C}\tilde{\epsilon}_i + \frac{1}{2}\frac{\partial^2 C_{BS}^{-1}\{C(Z_i); Z_i\}}{\partial C^2}\tilde{\epsilon}_i^2 \\ &\quad + \frac{1}{3!}\frac{\partial^3 C_{BS}^{-1}\{C(Z_i) + q\tilde{\epsilon}_i; Z_i\}}{\partial C^3}\tilde{\epsilon}_i^3 \\ &\quad \text{for some } q \in (0, 1).\end{aligned} \tag{3.5}$$

Taking conditional expectation on both sides of (3.5),

$$\begin{aligned}E\{\hat{\sigma}_I(Z_i)|Z_i\} &= \sigma(Z_i) + \frac{\partial C_{BS}^{-1}\{C(Z_i); Z_i\}}{\partial C}E(\tilde{\epsilon}_i|Z_i) \\ &\quad + \frac{1}{2}\frac{\partial^2 C_{BS}^{-1}\{C(Z_i); Z_i\}}{\partial C^2}E(\tilde{\epsilon}_i^{\,2}|Z_i) \\ &\quad + \frac{1}{3!}E\left[\frac{\partial^3 C_{BS}^{-1}\{C(Z_i) + q\tilde{\epsilon}_i; Z_i\}}{\partial C^3}\tilde{\epsilon}_i^3|Z_i\right].\end{aligned} \tag{3.6}$$

Since $C(X_i)$ is less than S_i, the standardized residual $|\tilde{\epsilon}_i| < 1$. This means that the bias is very much determined by the first two terms on the right hand side of (3.6) so that

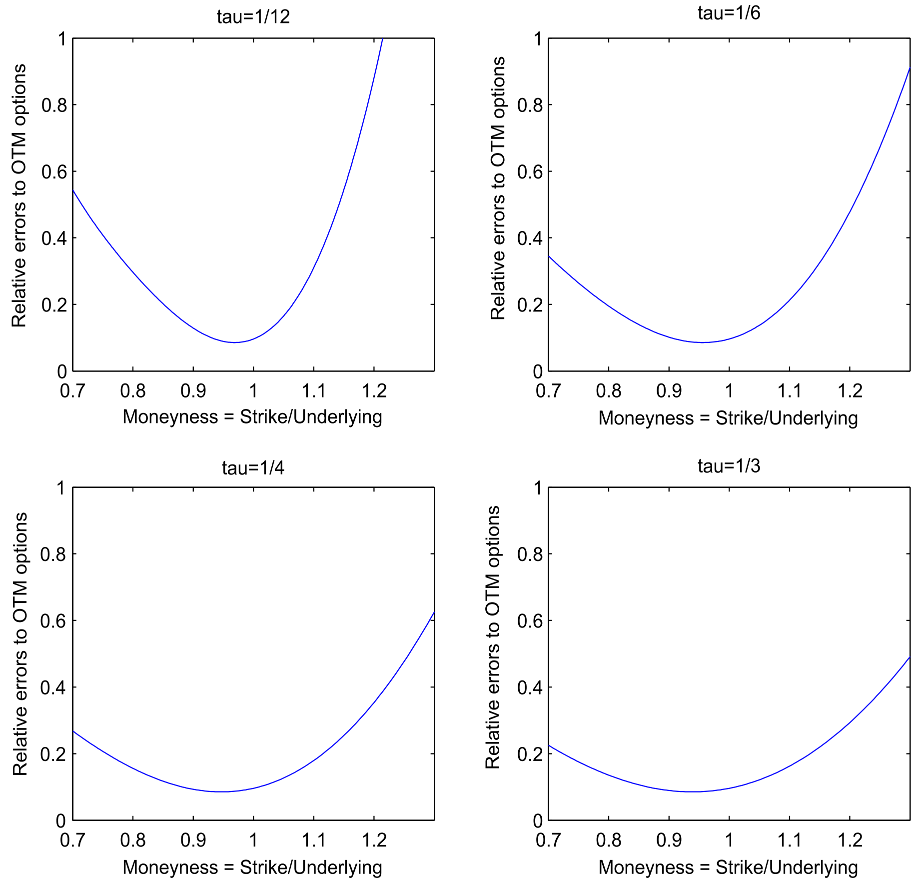


Fig. 3. Calibrated relative price deviation to OTM options, against moneyness for time to maturity $\tau = 1/12$ (upper left panel), $1/6$ (upper right panel), $1/4$ (bottom left panel) and $1/3$ (bottom right panel).

$$E\{\hat{\sigma}_I(Z_i)|Z_i\} - \sigma(Z_i) \approx \frac{\partial C_{BS}^{-1}\{C(Z_i);Z_i\}}{\partial C} E(\tilde{\epsilon}|Z_i) + \frac{1}{2}\frac{\partial^2 C_{BS}^{-1}\{C(Z_i);Z_i\}}{\partial C^2} E(\tilde{\epsilon}^2|Z_i). \quad (3.7)$$

If the error ϵ_i is unbiased so that $E(\epsilon_i|X_i) = 0$, then the bias of the implied volatility will be largely impacted by the quadratic term, namely

$$E\{\hat{\sigma}_I(Z_i)|Z_i\} - \sigma(Z_i) \approx \frac{1}{2}\frac{\partial^2 C_{BS}^{-1}\{C(Z_i);Z_i\}}{\partial C^2} E(\tilde{\epsilon}^2|Z_i). \quad (3.8)$$

Both (3.7) and (3.8) reveal that the implied volatility is biased, which causes the implied volatility inconsistent to the underlying $\sigma(Z_i)$.

Taking conditional variance on (3.5),

$$Var\{\hat{\sigma}_I(Z_i)|Z_i\} \approx \left[\frac{\partial C_{BS}^{-1}\{C(Z_i);Z_i\}}{\partial C}\right]^2 Var(\tilde{\epsilon}_i|Z_i) \quad (3.9)$$

which is influenced by the first partial derivative of $C_{BS}^{-1}\{C;Z_i\}$ with respect to C and the conditional variance of $\tilde{\epsilon}_i$. We see here that the bias and the variance of $\hat{\sigma}_I(Z_i)$ are largely influenced by the derivatives of C_{BS}^{-1} and the conditional moments of the pricing error. Importantly, both the bias and the variance do not get smaller as the number of options used is increased, indicating that the implied volatility $\hat{\sigma}_I(Z_i)$ is not statistically inconsistent.

4. Stylish patterns in implied volatility

We see from (3.7) and (3.9) that the first two derivatives of $C_{BS}^{-1}(C;Z_i)$ with respect to C contribute significantly to the behaviors of the implied volatility. Hence, they deserve a further analysis. Let $\sigma(C;z) = C_{BS}^{-1}(C;z)$ be the implied volatility of C at z for a general price C. If $C(z) = C_{BS}(z;\sigma)$, then $\sigma(C;z) \equiv \sigma$.

A derivation in Chen and Xu (2013) shows that at $z = (m, \tau)$

$$\frac{\partial C_{BS}^{-1}(C;z)}{\partial C} = \left(\sqrt{\tau}\phi[d_1\{\sigma(c;z)\}]\right)^{-1}, \quad (4.1)$$

$$\frac{\partial^2 C_{BS}^{-1}(C;z)}{\partial C^2} = \tau^{-1}\phi^{-2}[d_1\{\sigma(c;z)\}] \times \left(\frac{\tau\sigma(c;z)}{4} - \frac{\{-\ln(m)+r\tau\}^2}{\tau\{\sigma(c;z)\}^3}\right) \quad (4.2)$$

where ϕ is the standard normal density and

$$d_1\{\sigma(C;z)\} = \frac{-\ln(m) + (r + \frac{\sigma^2(c;z)}{2})\tau}{\sigma(C;z)\sqrt{\tau}}.$$

The behavior of the first derivative depends on the underlying price C via $d_1\{\sigma(C;z)\}$. It will be the smallest over a set of z satisfying $m = e^{-\{t+\sigma^2(c;z)\}\tau}$ and gets larger as z moves away from the set. This may be viewed as a smile pattern with the depth of the smile occurred at the set. A similar analysis on the second derivative shows that

$$\frac{\partial^2 C_{BS}^{-1}(C;z)}{\partial C^2} \begin{cases} >0, & \text{if } e^{r\tau - \frac{\sigma^2(v;z)}{2}\tau} < m < e^{r\tau + \frac{\sigma^2(v;z)}{2}\tau}, \\ =0, & \text{if } m = e^{r\tau - \frac{\sigma^2(v;z)}{2}\tau} \text{ or } e^{r\tau + \frac{\sigma^2(v;z)}{2}\tau}, \\ <0, & \text{if } m < e^{r\tau + \frac{\sigma^2(v;z)}{2}\tau} \text{ or } m > e^{r\tau - \frac{\sigma^2(v;z)}{2}\tau}. \end{cases}$$

Hence, the second derivative has a "sad" shape. If $\tau \to 0$, the two roots of the second derivative converge to $m = 1$ at the money.

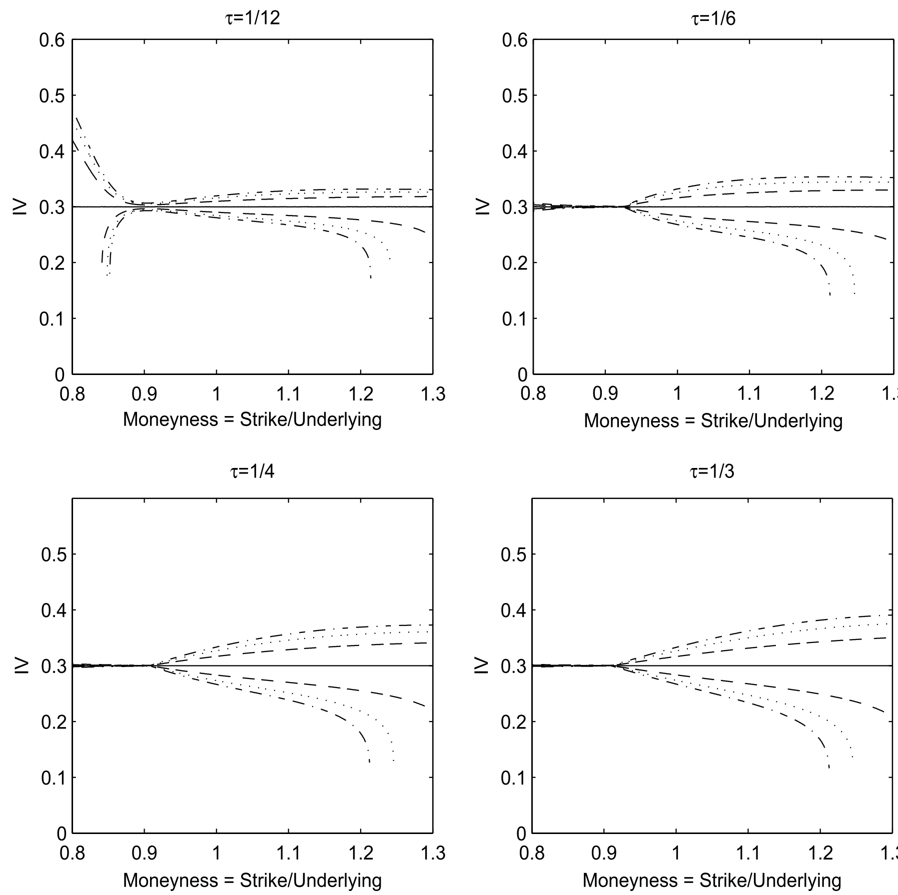


Fig. 4. Sensitivity of implied volatility to the relative price deviation. True volatility (solid line); implied volatility with relative pricing errors to OTM options, displayed in Fig. 3, multiplied by ± 0.5 (dashed lines), ± 0.8 (dotted lines) and ± 1 (dashed and dotted lines) at each maturity.

If the underlying price is Black–Scholes, the above diagnosis becomes more precise. In this case, $\sigma(c; z) = \sigma$ and $d_1 = \frac{-\ln(m)+(r+\frac{\sigma^2}{2})\tau}{\sigma\sqrt{\tau}}$. It can be checked that $\frac{\partial C_{BS}^{-1}(C;z)}{\partial C}$ is a decreasing function with respect to m until reaching the minimum $m^* = e^{-(r+\frac{\sigma^2}{2})\tau}$, and it increases for $m \geq m^*$. More specifically,

$$\frac{\partial^2\left(\frac{\partial C_{BS}^{-1}}{\partial C}\right)}{\partial m^2} = \sqrt{\frac{2\pi}{\tau}}e^{\frac{d_1^2}{2}}\frac{1}{m^2}\frac{1}{\sigma^2\tau}(1+d_1^2+\sigma\sqrt{\tau}d_1).$$

Hence the first derivative is convex (smile) with respect to the moneyness if $\sigma\sqrt{\tau} < 2$, which is easily satisfied for the commonly encountered situations.

On the second partial derivative, it can be shown that

$$\frac{\partial^2\left(\frac{\partial^2 C_{BS}^{-1}}{\partial C^2}\right)}{\partial m^2} = \frac{-2\pi}{\sigma^3\tau^2}\left(\frac{1}{m^2}\right)e^{d_1^2}\{4d_1^4 - 2\sigma\sqrt{\tau}d_1^3 + (10-2\sigma^2\tau)d_1^2 - 4\sigma\sqrt{\tau}d_1 + 2 - \sigma^2\tau\}.$$

A sufficient condition for $\frac{\partial^2 C_{BS}^{-1}}{\partial V^2}$ being concave (sad) with respect to the moneyness is $\sigma\sqrt{\tau} \leq \sqrt{37-\sqrt{649}}/3 \approx 1.1316$, which is easily satisfied in practice too.

Fig. 5 displays the first two derivatives of $C_{BS}^{-1}(C; z)$ with respect to C plotted against the moneyness, which confirms the above analysis. These features of the derivatives mean that, if the Black–Scholes price is valid, a smile is injected into the implied volatility if $E(\tilde{\epsilon}|X_i) \neq 0$. The smile will be offset by the second derivative which is concave. When the error is heteroscedastic, the picture will be mixed depending on the first two conditional moments $E(\tilde{\epsilon}_i^j|Z_i)$. We also note that the nature of the first derivative means that the variance of the implied volatility tends to be larger for in-the-money and out-of-money than at-the-money options for homoscedastic errors.

To confirm the above analyses, we conducted a simulation under the Black–Scholes model with pricing errors such that $Y_i = C_{BS}(X_i; \sigma) + \epsilon_i$ where $\varepsilon_i = \upsilon(M_i)C_{BS}(X_i)e_i$, $e_i \sim N(0, 1)$ and $\upsilon^2(M) = 0.002 + 0.152\{\max(1 - |M - 1.328|/0.319, 0)\}^{0.437}$. The errors were made relative to the underlying price $C_{BS}(X_i)$ with the conditional variance of $\epsilon_i/C_{BS}(X_i; \sigma)$ given X_i being modeled as a clipped quadratic function of the moneyness. The specific form of $\upsilon^2(M)$ was motivated by empirical analysis on the S&P 500 option data, which will be elaborated in Section 7. We chose the interest rate $r = 0.01$ and $\sigma = 0.3$ in the C_{BS} formula.

Fig. 6 displays the averaged implied volatility estimates based on 500 simulations on each chosen value of X_i for four maturity $\tau = 1/12, 1/6, 1/4$ and $1/3$ respectively. The bias in the implied volatility is clearly shown with a pronounced curving-up for $M < 1$. As the maturity increases, the curvature is reduced. However, the amount of spread-out (reflecting the variance of the implied volatility) is still significant. Although the implied volatility displayed little curvature for the out of the money ($M < 1$), it lays much below than the underlying $\sigma = 0.3$, which is especially severe for shorter maturities of two weeks and 2 months. We plot on the figure also the proposed implied kernel volatility estimates, whose details will be given in the next section. Fig. 6 shows that the proposed volatility estimates capture the underlying flat volatility quite accurately.

In summary, the implied volatility is vulnerable to pricing errors, so vulnerable that it is not consistent due to the bias conveyed

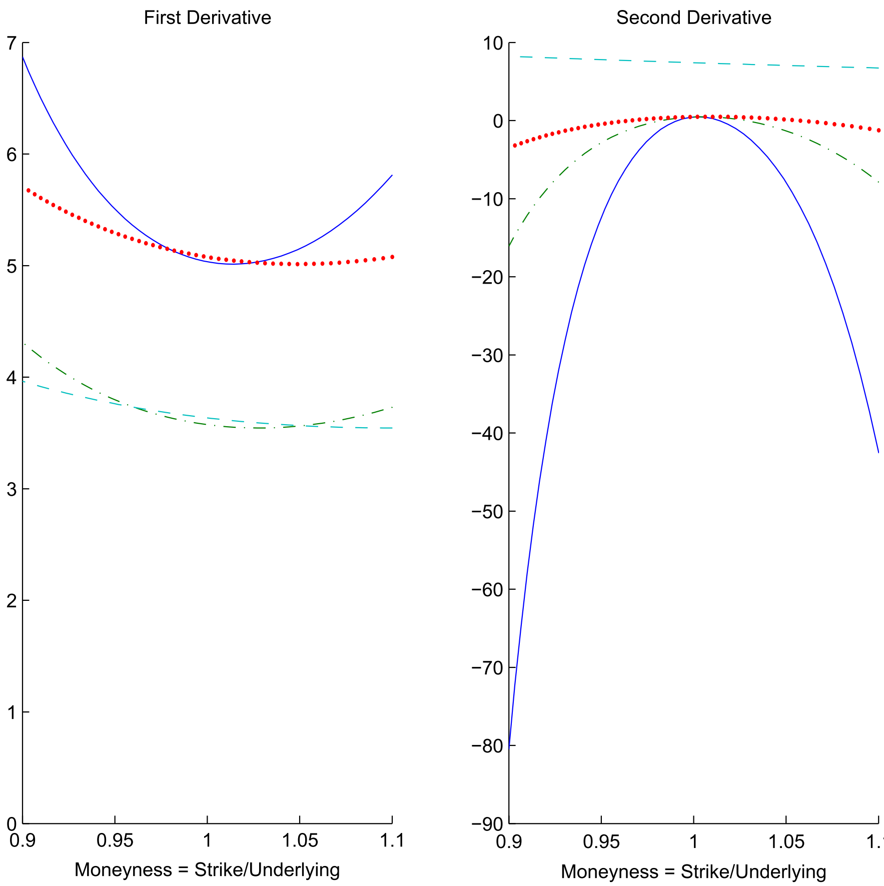


Fig. 5. The two derivatives of $C_{BS}^{-1}(C; z)$ with respect to C as functions of the moneyness for selected values of (τ, σ) and $r \equiv 0.01$. Left panel: the solid, dash-dotted, dotted and dashed line represent (τ, σ) being (0.25, 0.3), (0.5, 0.3), (1, 0.3) and (2, 1.5), respectively. Right panel: the solid, dash-dotted, dotted and dashed line represent (τ, σ) as (0.25, 0.3), (0.5, 0.3), (0.25, 0.6) and (0.5, 0.6), respectively.

in (3.6). The patterns of $\frac{\partial C_{BS}^{-1}(C;z)}{\partial C}$ and $\frac{\partial^2 C_{BS}^{-1}(C;z)}{\partial C^2}$ can be easily injected to the implied volatility. The approach of inverting one price at a time is a poor method in recovering the underlying volatility inherited in the option prices.

5. Kernel estimation

We propose a consistent volatility estimator which is free of the afore-mentioned problems with the conventional implied volatility. Our proposal is based on the nonparametric regression method (Härdle, 1990; Fan and Gijbels, 1996) that provides a kernel smoothed estimation of the underlying price function $C(X)$ in the presence of the pricing errors. The kernel smoothing method has been employed in Aït-Sahalia and Lo (1998) and Fan and Mancini (2009) for estimating the state price density/survival functions.

Let $K(\cdot)$ be a univariate symmetric probability density function, which is called the kernel. Suppose there are d components in the covariate X_i, which is often 3 or 4 depending on if the interest rate varies enough to be treated as a covariate. Let $\mathcal{K}(x_1, \ldots, x_d) = K(x_1) \cdots K(x_d)$ be a d-variate product kernel, and $h_1, \ldots, h_d$ be d smoothing bandwidths used to smooth at each components of X_i respectively. The bandwidths are required to converge to zero while maintaining $nh_1 \cdots h_d \to \infty$. Define for $x = (x_1, \ldots, x_d)^T$, $\mathcal{K}_h(x) = (h_1 \cdots h_d)^{-1} \mathcal{K}(x_1/h, \ldots, x_d/h_d)$.

As Model (3.1) implies that $C(X)$ is the regression function of the option price given X, the Nadaraya–Watson (NW) kernel estimator (Nadaraya, 1964; Watson, 1964) of the option price $C(X)$ is

$$\hat{C}(x) = \frac{\sum_{i=1}^{n} \mathcal{K}_h(X_i - x) Y_i}{\sum_{i=1}^{n} \mathcal{K}_h(X_i - x)}, \tag{5.1}$$

which is a locally weighted least square estimator at each x. The local linear kernel estimator (Fan and Gijbels, 1996) is a more advanced version with better properties (free of boundary bias, with much simpler form for the bias) than the NW estimation. For easy expedition, we will present our approach in terms of the NW estimator.

If $C(X)$ is homogeneous of degree 1, we consider Model (3.4) with covariate $Z_i = (M_i, \tau_i)$ and $E(\tilde{\epsilon}_i | Z_i) = 0$. The kernel estimator of $C(Z)$ at $z = (m, \tau)$ is

$$\hat{C}_k(z) = \frac{\sum_{i=1}^{n} K_{h_m}(M_i - m) K_{h_\tau}(\tau_i - \tau) \tilde{Y}_i}{\sum_{i=1}^{n} K_{h_m}(M_i - m) K_{h_\tau}(\tau_i - \tau)}, \tag{5.2}$$

where $K_h(x) = h^{-1} K(x/h)$, and h_m and h_τ are smoothing bandwidths which control the amount of local averaging and smoothness in a neighborhood of $z = (m, \tau)$.

The following conditions are assumed in the kernel estimation of $C(z)$.

(A1) The price function C and f_Z, the probability density function of Z, have continuous second partial derivatives with respect to m and τ, respectively.

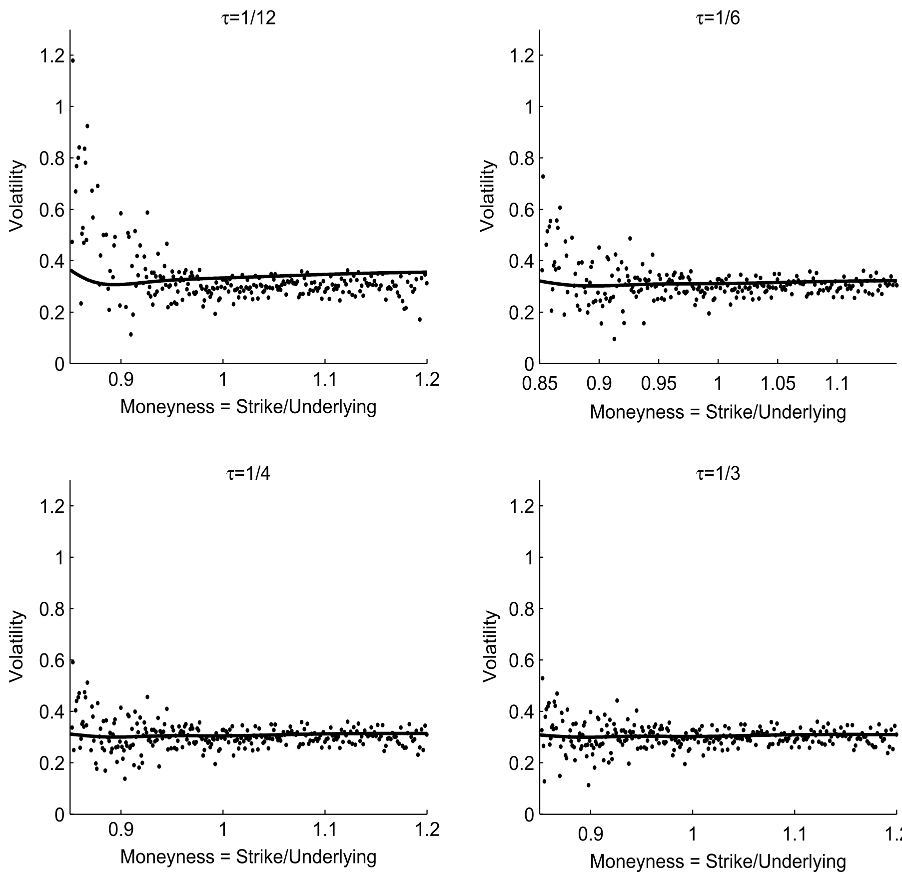


Fig. 6. Average implied volatility (dots) and average kernel implied kernel volatility (solid lines) based on the calibrated relative pricing error with $v^2(m) = 0.002 + 0.152\{\min(1 - |m - 1.328|/0.319, 0)\}^{0.437}$.

(A2) The conditional variance $V(z) = Var(\epsilon|Z = z)$ is finite and has continuous first partial derivatives with respect to m and τ, respectively.

(A3) The kernel K satisfies $\int |K(\mu)|d\mu < \infty$, $\int \mu K(\mu)d\mu = 0$, $\int \mu^2 K(\mu)d\mu < \infty$. The smoothing bandwidths satisfy $h_m \to 0$, $h_\tau \to 0$, $nh_m h_\tau \to \infty$ and $\frac{h_m}{h_\tau} = O(1)$ as $n \to \infty$.

From the literature of nonparametric regression (Härdle, 1990; Fan and Gijbels, 1996), it can be confirmed that $\hat{C}_h(z)$ is consistent to $C(z;\theta)$ under Conditions (A1)–(A3). In particular,

$$\begin{aligned} E\{\hat{C}_k(z)\} \;=\; & C(z;\theta) + \frac{1}{2}\mu_2(K)\left(\frac{\partial^2 C}{\partial m^2} + \frac{2}{f_z(z)}\frac{\partial C}{\partial m}\frac{\partial f_z}{\partial m}\right)h_m^2 \\ & + \frac{1}{2}\mu_2(K)\left(\frac{\partial^2 C}{\partial \tau^2} + \frac{2}{f_z(z)}\frac{\partial C}{\partial \tau}\frac{\partial f_z}{\partial \tau}\right)h_\tau^2 \\ & + \mu_2(K)\frac{1}{f_z(z)}\left(\frac{\partial C}{\partial m}\frac{\partial f_z}{\partial \tau} + \frac{\partial C}{\partial \tau}\frac{\partial f_z}{\partial m}\right)h_m h_\tau \\ & + o(h_m^2 + h_\tau^2) \quad \text{and} \end{aligned} \tag{5.3}$$

$$Var\{\hat{C}_k(z)\} = \frac{R^2(K)v^2(z)}{nh_m h_\tau f_z(z)} + o\{(nh_m h_\tau)^{-1}\} \tag{5.4}$$

where $\mu_2(K) = \int u^2 K(u)du$ and $R(K) = \int K^2(u)du$. The corresponding result to (5.3) for the local linear estimator $\hat{C}_{ll}(z)$ (Ruppert and Wand, 1994) is

$$\begin{aligned} E\{\hat{C}_{ll}(z)\} = & C(z;\theta) + \frac{1}{2}\mu_2(K)\left(\frac{\partial^2 C}{\partial m^2}h_m^2 + \frac{\partial^2 C}{\partial \tau^2}h_\tau^2\right) \\ & + o(h_m^2 + h_\tau^2) \end{aligned} \tag{5.5}$$

while the form of the variance is the same with (5.4).

In summary, we have

$$\hat{C}(z) - C(z) = O_p\{h_m^2 + h_\tau^2 + (nh_m h_\tau)^{-1/2}\}, \tag{5.6}$$

for $\hat{C}$ being either the NW estimator $\hat{C}_k$ or the local linear estimator $\hat{C}_{ll}$. Hence, both kernel estimators are consistent to the underlying price $C(z)$ under Condition (A1)–(A3).

Our purpose here is to recover the underlying volatility by inverting $\hat{C}_k(Z_i)$ rather than Y_i or $\tilde{Y}_i$. The proposed implied kernel (IK) volatility estimator at Z_i is

$$\hat{\sigma}_{IK}(Z_i) = C_{BS}^{-1}\{\hat{C}_k(Z_i); Z_i\}. \tag{5.7}$$

The IK volatility estimator differs from

$$\hat{\sigma}_{KI}(z) = \frac{\sum_{i=1}^{n} K_{h_m}(M_i - m)K_{h_\tau}(\tau_i - \tau)\hat{\sigma}_I(Z_i)}{\sum_{i=1}^{n} K_{h_m}(M_i - m)K_{h_\tau}(\tau_i - \tau)}, \tag{5.8}$$

considered in Aït-Sahalia and Lo (1998) which replaces $\tilde{Y}_i$ by $\hat{\sigma}_I(Z_i)$ in (5.2). It does the implied volatility inversion $\hat{\sigma}_I(Z_i)$ first followed by kernel smoothing. The proposed $\hat{\sigma}_{IK}$ reverse the order of the smoothing and volatility inversion.

To appreciate the merits of $\hat{\sigma}_{IK}$, a Taylor expansion as (3.5) for $\hat{\sigma}_{IK}(Z_i)$ yields

$$\begin{aligned} \hat{\sigma}_{IK}(Z_i) = & \sigma(Z_i) + \frac{\partial C_{BS}^{-1}\{C(Z_i); Z_i\}}{\partial C}\{\hat{C}_h(Z_i) - C(Z_i)\} \\ & + \frac{1}{2}\frac{\partial^2 C_{BS}^{-1}\{C(Z_i); Z_i\}}{\partial C^2}\{\hat{C}_h(Z_i) - C(Z_i)\}^2 \\ & + O_p[\{h_m^2 + h_\tau^2 + (nh_m h_\tau)^{-1/2}\}^{3/2}]. \end{aligned} \tag{5.9}$$

Taking conditional expectation on both sides of (5.9), we have from (5.3) and (5.4),

$$E\{\hat{\sigma}_{IK}(Z_i)|Z_i\} - \sigma(Z_i) = \frac{\partial C_{BS}^{-1}\{C(Z_i); Z_i\}}{\partial C} E\{\hat{C}_h(Z_i) - C(Z_i)|Z_i\} + o(h_m^2 + h_\tau^2) + O\{(nh_m h_\tau)^{-1}\}. \quad (5.10)$$

From (5.3), the leading bias approaches to zero as both h_m and h_τ converges to zero as the sample size n increases. Hence the implied kernel volatility estimator is asymptotically unbiased. Also, the conditional variance

$$Var\{\hat{\sigma}_{IK}(Z_i)|Z_i\} = \left[\frac{\partial C_{BS}^{-1}\{C(Z_i); Z_i\}}{\partial C}\right]^2 \frac{R^2(K)V^2(K)}{nh_m h_\tau f_Z(z)} + o\{(nh_m h_\tau)^{-1}\} \quad (5.11)$$

which diminishes to zero at the rate of $(nh_m h_\tau)^{-1}$. This together with the asymptotically unbiasedness means that $\hat{\sigma}_{IK}$ is consistent.

We note that despite $\frac{\partial C_{BS}^{-1}\{C(Z_i);Z_i\}}{\partial C}$ and $\frac{\partial^2 C_{BS}^{-1}\{C(Z_i);Z_i\}}{\partial C^2}$ still appear at the same locations in (5.10) and (5.11) as they did in (3.7) and (3.9), they do not matter much now as both $E\{\hat{\sigma}_k(Z_i) - C(Z_i)|Z_i\}$ and $Var\{\hat{\sigma}_k(Z_i)|Z_i\}$ converge to zero. This is the reason why $\hat{\sigma}_{IK}$ is consistent to $\sigma(Z_i)$ whereas the conventional $\hat{\sigma}_I$ is not.

A similar analysis can be made on $\hat{\sigma}_{KI}(Z_i)$ of Aït-Sahalia and Lo (1998) given in (5.8). Since $\hat{\sigma}_I(Z_i)$ is biased to $\sigma(Z_i)$ as revealed in Section 3, $\hat{\sigma}_{KI}(Z_i)$ has the same issue of bias as $\hat{\sigma}_I(Z_i)$. Thus, $\hat{\sigma}_{KI}$ is also not consistent, although the smoothing can reduce the variability.

In practice, the smoothing bandwidths can be obtained by the method of cross-validation, namely we obtain (h_m, h_τ) by minimizing

$$CV(h) = \frac{1}{n}\sum_{i=1}^{n}\{\tilde{Y}_i - \hat{C}_{h,-i}(Z_i)\}^2 \quad \text{where} \quad (5.12)$$

$$\hat{C}_{h,-i}(Z_i) = \frac{\sum_{j\neq i} K_{h_m}(M_j - M_i)K_{h_\tau}(\tau_j - \tau_i)\tilde{Y}_j}{\sum_{j\neq i} K_{h_m}(M_j - M_i)K_{h_\tau}(\tau_j - \tau_i)} \quad (5.13)$$

is the leave-out-one estimator of $C(Z_i)$; see Härdle (1990) and Fan and Gijbels (1996) for more details.

An approach (Dumas et al., 1998; Hentschel, 2003), which is quite widely employed by practitioners, is the regression modeling of the volatility

$$\hat{\sigma}_I(Z_i) = g(Z_i; \theta) + e_i, \quad (5.14)$$

where $g(z; \theta)$ is a parametric function with unknown parameters θ and e_i's are the residuals. The function g prescribes a pattern of the implied volatility with respect to the moneyness and maturity. However, as $\hat{\sigma}_I(Z_i)$ is adversely affected by the pricing errors, regressing $\hat{\sigma}_I(Z_i)$ does not produce valid estimation of the underlying volatility. We propose replacing $\hat{\sigma}_I(Z_i)$ by $\hat{\sigma}_{IK}(Z_i)$ in the regression. Then, the regression modeling of the implied volatility works if model $g(z; \theta)$ is correctly specified.

The implied kernel volatility estimator can be used to test for the specification of the parametric $g(Z; \theta)$ in (5.14). Specifically, we want to test

$H_0 : \sigma(z) = g(z; \theta_0)$ for all z and some θ_0 within the parameter space.

The specification test statistic is $D_n = \frac{1}{n}\sum_{i=1}^{n}\{\hat{\sigma}_{IK}(Z_i) - g(Z_i; \hat{\theta})\}^2$ where $\hat{\theta}$ is estimated by minimizing $\sum_{i=1}^{n}\{\hat{\sigma}_{IK}(Z_i) - \sigma_{DV}(Z_i; \theta)\}^2$ with respect to θ. The statistic D_n is the average deviation between the assumed parametric $g(Z; \theta)$ under the best condition (as $\hat{\theta}$ is used) and the model-free estimate of $\sigma(z)$ by $\hat{\sigma}_{IK}(Z_i)$. This formulation is in the same spirit as Härdle and Mammen (1993) and Chen and Gao (2007).

To perform testing on H_0, we need to find the null distribution of D_n to obtain a critical value c_α at a significant level α so that H_0 is rejected if $D_n \geq c_\alpha$. We use the wild bootstrap method (Härdle and Mammen, 1993). It first generates re-sampled option prices Y_i^* at Z_i by adding simulated errors ϵ^* such that

$$Y_i^* = C(Z_i; \hat{\theta}) + \epsilon_i^* \quad \text{for } i = 1, 2, \ldots, n, \quad (5.15)$$

where each ϵ_i^* is generated from a distribution $\hat{F}_i$ based on $\hat{\epsilon}_i = Y_i - C(Z_i; \hat{\theta})$ such that

$$E_{\hat{F}_i}(\epsilon_i^*) = 0, E_{\hat{F}_i}(\epsilon_i^{*2}) = \hat{\epsilon}_i^2 \quad \text{and} \quad E_{\hat{F}_i}(\epsilon_i^{*3}) = \hat{\epsilon}_i^3. \quad (5.16)$$

Here $E_{\hat{F}_i}$ denotes expectation with respect to $\hat{F}_i$. There are several ways to determine $\hat{F}_i$. One is to make $\hat{F}_i$ discrete on two points so that (5.16) is satisfied. Given $\hat{\epsilon}_i$, $\hat{F}_i$ has the form $P\{(\frac{1+\sqrt{5}}{2})\hat{\epsilon}_i\} = \frac{5-\sqrt{5}}{10}$ and $P\{(\frac{1-\sqrt{5}}{2})\hat{\epsilon}_i\} = \frac{5+\sqrt{5}}{10}$. The name "wild" reflects that $\hat{F}_i$ is constructed based on a single residual $\hat{\epsilon}_i$.

We compute the implied kernel volatility $\hat{\sigma}_{IK}^*(Z_i)$ based on the re-sampled $\{(Z_i, Y_i^*)\}_{i=1}^n$ using the original bandwidths but re-estimating θ to obtain $\hat{\theta}^*$. The bootstrap version of D_n is $D_n^* = \frac{1}{n}\sum_{i=1}^{n}\{\hat{\sigma}_{IK}^*(Z_i) - g(Z_i; \hat{\theta}^*)\}^2$. Repeating the above procedure B times for a large integer B, we obtain $D_n^{*1} \leq D_n^{*2} \leq \cdots \leq D_n^{*B}$ after sorting them in the ascending order. The bootstrap estimate of the critical value c_α is $\hat{c}_\alpha = D_n^{*[\alpha B]+1}$ where $[\cdot]$ denotes the integer truncation. The H_0 is rejected if $D_n \geq \hat{c}_\alpha$. In Section 7, we will demonstrate this procedure to test for the specification of a set of parametric models for the S&P 500 option data.

6. Simulation studies

We report the results from simulation studies which were designed to provide numerical confirmation to our theoretical findings in the previous sections. The performance of the conventional implied volatility $\hat{\sigma}_I$, the smoothed implied volatility $\hat{\sigma}_{KI}$ of Aït-Sahalia and Lo (1998) and the proposed implied kernel volatility $\hat{\sigma}_{IK}$ were evaluated by conducting simulations with different pricing models and error distributions.

Three option price models were experimented for $C(X)$: Black–Scholes (BS) $C_{BS}(\sigma; X_i)$ with $\sigma = 0.3$, the ad-hoc deterministic volatility (Dumas et al., 1998) $C_{BS}\{\sigma(Z_i); X_i\}$ with $\sigma(Z_i) = 0.3 + 4(M_i^{-1} - 1)^2$ (ad-hoc DV-1) and $\sigma(Z_i) = 0.3 + 4e^{-\tau_i}(M_i^{-1} - 1)^2$ (ad-hoc DV-2), and Merton-jump pricing formula (Merton, 1976) $C_{MJ}(X_i, \theta)$ with $\theta \equiv (\lambda, m, \upsilon, \sigma) = (5, 1, 0.12, 0.06\sqrt{5})$ (MJ-1) and $(5, 1, 0.06\sqrt{2}, 0.06\sqrt{15})$ (MJ-2). In the Merton-jump model, the parameter λ is the unit intensity of the Poisson jump, and (m, υ) defines the distribution of the jump size η such that $\ln(1+\eta)$ follows $N\{\ln(m), \upsilon\}$.

The distribution of the covariable $X = (S, K, \tau)$ was generated according to the empirical distributions of these three covariables from S&P 500 option data of one year from March 15, 2008. Once these X_i were generated, we created call prices Y_i according to

$$Y_i = C(Z_i) + \epsilon_i, \quad i = 1, \ldots, n$$

for each of the three pricing models. The errors $\varepsilon_i = \upsilon(Z_i)C(Z_i)e_i$, where $e_i \sim N(0, 1)$ and $\upsilon^2(Z_i) = E\{\varepsilon_i/C(Z_i)|Z_i\}$ was made a function of Z_i representing heteroscedasticity in the relative error. The form of $\upsilon^2(Z_i)$ was acquired from the S&P 500 option data by first conducting the nonparametric estimation of the underlying option price function $C(Z)$, and then obtaining the estimated errors by subtracting the nonparametrically estimated prices from the

Table 1
Average squared bias, variance and MSE of estimators $\hat{\sigma}_I$, $\hat{\sigma}_{KI}$ and $\hat{\sigma}_{IK}$ with pricing errors $\epsilon_i = C(X_i)e_i$, where $e_i \sim N(0, \upsilon^2(m, \tau))$ with pricing formulae: the Black–Scholes (BS) with $\sigma = 0.3$, the ad-hoc deterministic volatility (ad-hoc DV) and the Merton Jump (MJ), and $\upsilon^2(m) = 0.002 + 0.152\{\min(1 - |m - 1.328|/0.319, 0)\}^{0.437}$.

Model	n	Squared Bias			Variance			MSE		
		$\hat{\sigma}_I$	$\hat{\sigma}_{KI}$	$\hat{\sigma}_{IK}$	$\hat{\sigma}_I$	$\hat{\sigma}_{KI}$	$\hat{\sigma}_{IK}$	$\hat{\sigma}_I$	$\hat{\sigma}_{KI}$	$\hat{\sigma}_{IK}$
	300	0.0192	0.0155	0.0131	0.0028	0.0013	0.0013	0.0220	0.0167	0.0144
	500	0.0199	0.0167	0.0100	0.0028	0.0008	0.0010	0.0227	0.0175	0.0110
BS	700	0.0203	0.0172	0.0107	0.0027	0.0008	0.0010	0.0230	0.0180	0.0117
	1000	0.0210	0.0181	0.0097	0.0029	0.0006	0.0008	0.0239	0.0187	0.0105
	2000	0.0221	0.0195	0.0084	0.0029	0.0004	0.0007	0.0250	0.0199	0.0091
	300	0.0205	0.0162	0.0112	0.0032	0.0016	0.0013	0.0237	0.0178	0.0125
	500	0.0195	0.0161	0.0102	0.0030	0.0010	0.0012	0.0226	0.0171	0.0114
ad-hoc DV-1	700	0.0226	0.0187	0.0118	0.0033	0.0010	0.0011	0.0258	0.0197	0.0129
	1000	0.0201	0.0171	0.0102	0.0030	0.0006	0.0009	0.0231	0.0177	0.0111
	2000	0.0185	0.0163	0.0068	0.0029	0.0004	0.0007	0.0215	0.0167	0.0075
	300	0.0080	0.0061	0.0054	0.0019	0.0006	0.0009	0.0099	0.0067	0.0062
	500	0.0252	0.0219	0.0135	0.0036	0.0009	0.0013	0.0287	0.0228	0.0148
ad-hoc DV-2	700	0.0169	0.0141	0.0087	0.0027	0.0007	0.0009	0.0196	0.0148	0.0097
	1000	0.0253	0.0223	0.0125	0.0032	0.0007	0.0010	0.0285	0.0229	0.0135
	2000	0.0192	0.0171	0.0069	0.0029	0.0003	0.0006	0.0221	0.0174	0.0075
	300	0.0069	0.0041	0.0042	0.0036	0.0014	0.0015	0.0105	0.0055	0.0057
	500	0.0046	0.0031	0.0020	0.0031	0.0008	0.0009	0.0077	0.0039	0.0028
MJ-1	700	0.0080	0.0060	0.0027	0.0037	0.0010	0.0010	0.0116	0.0070	0.0036
	1000	0.0061	0.0044	0.0020	0.0036	0.0006	0.0008	0.0097	0.0050	0.0028
	2000	0.0077	0.0061	0.0015	0.0038	0.0004	0.0007	0.0115	0.0065	0.0021
	300	0.0083	0.0058	0.0046	0.0033	0.0011	0.0015	0.0116	0.0069	0.0061
	500	0.0080	0.0059	0.0038	0.0034	0.0008	0.0010	0.0114	0.0067	0.0048
MJ-2	700	0.0102	0.0079	0.0037	0.0037	0.0007	0.0009	0.0139	0.0087	0.0046
	1000	0.0093	0.0072	0.0029	0.0037	0.0005	0.0009	0.0130	0.0077	0.0037
	2000	0.0088	0.0070	0.0022	0.0036	0.0003	0.0006	0.0124	0.0073	0.0028

observed option prices. After inspecting the estimated relative errors against the moneyness, we regressed the square of the estimated relative errors on the moneyness. The least square fit was

$$\upsilon^2(z) = 0.002 + 0.152\{\min(1 - |m - 1.328|/0.319, 0)\}^{0.437} \quad (6.1)$$

which we used in the simulation. The sample size was $n = 300, 500, 700, 1000$ and 2000. For each n, the simulation was replicated 2000 times.

As the above simulation models are all homogeneous of degree one, we use the covariate $Z_i = (M_i, \tau_i)$ and the standardized prices $\tilde{Y}_i$. The measures of the performance were the averaged empirical bias, variance and the mean square errors (MSE).

Table 1 reports the bias, variance and MSE of the three volatility estimators for each of the pricing models. There is a consistent trend in the empirical performance of the three estimators. As predicted by our theoretical analysis in Sections 3–5, the bias of the conventional implied volatility $\hat{\sigma}_I$ and the kernel smoothed volatility $\hat{\sigma}_{KI}$ did not get smaller as the sample size was increased. This happened for all the three pricing formulae. The conventional implied volatility $\hat{\sigma}_I$'s variance was not reduced neither as sample size got larger. These agreed with the diagnostics we have made early that $\hat{\sigma}_I$ is not a consistent estimator. Although the smoothing after implied volatility estimator $\hat{\sigma}_{KI}$ utilized more data information as reflected in its variance being reduced as n was increased, it was still severely biased since the kernel smoothing was conducted on the biased $\hat{\sigma}_I$. We observe that the bias of the proposed $\hat{\sigma}_{IK}$ decreased as the sample size n was increased and was much smaller than those of $\hat{\sigma}_I$ and $\hat{\sigma}_{KI}$. The MSE of $\hat{\sigma}_{IK}$ was noticeably less than those of the other two volatility estimators, which confirmed the consistency of the approach.

7. An empirical study

We report the results from an empirical analysis on the S&P 500 index option prices in Chicago Board Options Exchange. The options are European style and are the most actively traded in the world. We consider a one year period that includes the six months prior to and the six months after September 15, 2008, the date of Lehman Brothers' bankruptcy.

The data acquired from Market Express (http://www.marketdataexpress.com) consisted of the average last bid and ask prices for both calls and puts for each option product with a specific maturity date at various strikes. The last price is the most recently traded price which moves between the bid and ask as selling and buying orders are filled. Since the minimum tick for options traded below three dollars is $\frac{1}{16}$ and $\frac{1}{8}$ for all others, options with prices less than $\frac{1}{8}$ are discarded. These treatments are consistent with the existing practices in the literature (Aït-Sahalia and Lo, 1998; Bakshi et al., 1997).

There are issues of asynchrony and dividend payout for the option data. Asynchrony refers to that the underlying prices and option prices are not recorded at the same time. Although the S&P provides dividend payment information, the future dividend payments in the life of option contracts are difficult to determine. To account for these issues, as Aït-Sahalia and Lo (1998) and Fan and Mancini (2009), we used the forward prices based on the put–call parity. That is, for each option series with the same time-to-maturity τ at time t, we inferred their forward price $F_{t,\tau}$ by

$$F_{t,\tau} = (C_{tj} - P_{tj})e^{r\tau} + K_j, \quad (7.1)$$

where (C_{tj}, P_{tj}) is the call–put pair that has the same time to maturity τ and strike price K_j that is the nearest to being at-the-money in the option series. If there were multiple pairs of (C_{tj}, P_{tj}) satisfying the above criteria, the average of the inferred forward prices was used.

We included options with expiration at least 1 week and less than 4 months since these options are most actively traded as reflected by the trading volume and open interest. After the filtering, the final data set had 84777 call–put pairs with a daily average of 337 pairs.

To capture some time dynamics in the underlying volatility, we divided the whole data set to 12 monthly subsets from March 15th 2008. For each monthly data set, we conducted the volatility

Table 2
Cross-validation bandwidths for the 12 months from March 15th 2008.

Month	1	2	3	4	5	6	7	8	9	10	11	12
(a) Calls												
h_m	0.015	0.014	0.014	0.015	0.015	0.015	0.023	0.033	0.038	0.032	0.034	0.037
h_τ	0.019	0.018	0.021	0.018	0.017	0.021	0.008	0.007	0.009	0.008	0.013	0.021
(b) Puts												
h_m	0.015	0.014	0.014	0.015	0.015	0.015	0.023	0.033	0.038	0.032	0.034	0.037
h_τ	0.019	0.018	0.021	0.018	0.017	0.021	0.008	0.007	0.009	0.008	0.013	0.021

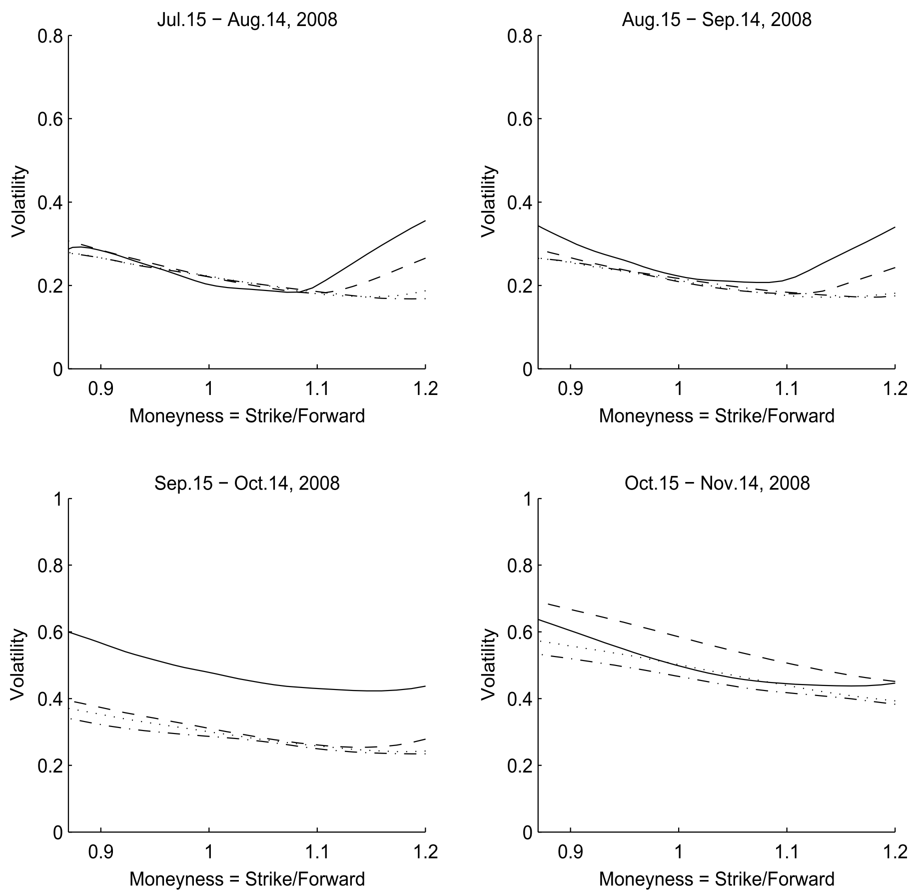


Fig. 7. Implied kernel volatilities for calls for the four months from July 15 and from October 15, 2008. Time to maturity was 2 weeks (solid line), 1 month (dashed line), 2 months (dotted line) and 3 months (dash-dotted line).

estimation using the proposed implied kernel volatility estimator, for both the calls and puts, with maturities at 2 weeks, and one, two and three months, respectively. For each of the four maturities, we pooled data of the week before and the week after together with data in the week of maturity.

We first estimated the call and put price functions for each monthly data by smoothing with respect to the moneyness and the time to maturity according to (5.2). The two smoothing bandwidths (h_m, h_τ) were obtained by the cross-validation (CV) described at the end of Section 5, whose values for the 12 months are reported in Table 2. Table 2 shows that the bandwidths for the calls were quite close to those of the puts. However, a substantial change occurred at the 7th month with the bandwidths for the moneyness much increased and that for the maturity much reduced. The 7th month started on the September 15th.

The use of kernel smoothed option prices reduced the incidences of truncation by an average 14.55% and 6.16% for the S&P 500 calls and puts respectively. This is because $\hat{C}_h(Z_i)$ is closer to the true price (which is above the lower bound) than $\tilde{Y}_i$, since the pricing errors are filtered out in the kernel smoothing.

The implied kernel volatilities based on the calls for the two months before and two months after the September 15th are shown in Fig. 7. We observe that the levels of the volatility for the two months after the September 15th were substantially higher than those two months before, reflecting the excessive volatility in the market in those turbulence months. These were consistent with the estimated pricing errors conveyed in Figs. 1 and 2. The implied kernel volatility estimates at the different maturities for the two months before the crisis displayed some "regular" patterns: (i) converged to a similar level at the money, (ii) displayed a clear descending order for OTM and ITM as the maturity was increased. These patterns disappeared after the out-burst of the crisis with the volatility level at the shortest maturity (2 weeks) jumped to a higher level in Month 7 and then receded back in Month 8. However, that for the one month maturity jumped in Month 8 and was higher than that for the two weeks maturity for almost the entire range of moneyness considered. It is observed that the volatility curves were largely smiling with the degree of the smiles reduced as the maturity was increased.

Figs. 8 and 9 display the conventional IV, the proposed IK and the KI volatility estimates for two three-day periods from January

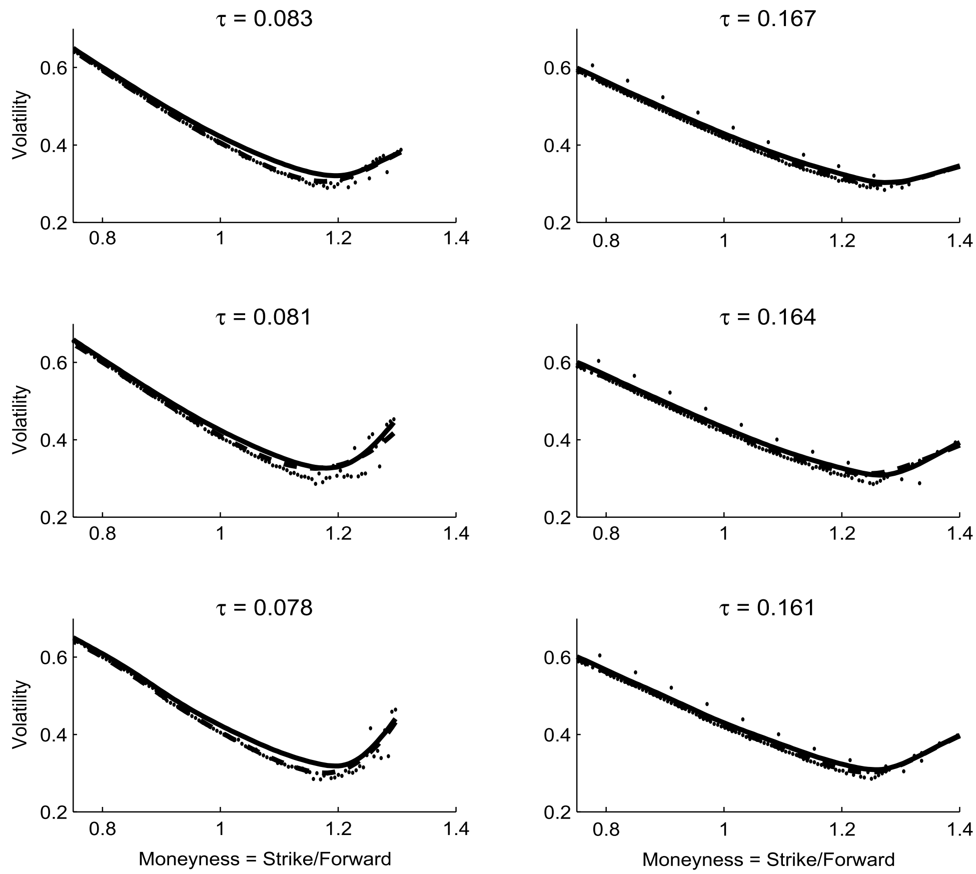


Fig. 8. Volatility estimates on January 21 (top panels), January 22 (middle panels) and January 23, 2009 (bottom panels), matured on February 21 (left panels) and March 21 (right panels). The implied volatility (dots), the kernel implied volatility (dashed lines) and the implied kernel volatility (solid line).

22–24 and March 2–4 2009 based on the daily data at two shorter maturities, respectively. The three day volatility estimates plotted in each figure provide the information on the time evolution of the volatility. We observe that the difference among the three volatility estimates were the greatest at the shorter maturity on each day, which echoes the finding that the pricing errors tend to be the most severe for shorter maturity as depicted in Figs. 1 and 2 (which plot the raw relative pricing errors) and in Fig. 7 for the kernel volatility estimates. There was not a significant change in the volatility within each of the three day period, indicating that the volatility in those periods was quite stable. The biggest difference between the IK and KI volatility estimates appeared for the moneyness (K/F) between 1 and 1.2, and the difference for further deep OTM was quite small. These could be understood by Figs. 1 and 2 which showed larger relative pricing errors for the moneyness between 1 and 1.2, and the size of the error declined for deep OTM.

We carried out the specification test based on the wild-bootstrap procedure outlined in Section 5 for the following volatility models $g(m, \tau; \theta)$ in (5.14):

(i) $\sigma(m, \tau) = \sigma$,

(ii) $\sigma(m, \tau) = \theta_0 + \theta_1 m + \theta_2 m^2$,

(iii) $\sigma(m, \tau) = \beta_0 + \beta_1 m + \beta_2 m^2 + \beta_3 m^3 + \beta_4 m^4$,

(iv) $\sigma(m, \tau) = \theta_0 + \theta_1 m + \beta_2 m^2 + \beta_3 \tau + \beta_4 \tau m + \beta_5 \tau^2$,

(v) $\sigma(m, \tau) = \theta_0 + \theta_1 m + \theta_2 m^2 + \theta_3 \tau + \theta_4 \tau m$,

(vi) $\sigma(m, \tau) = \beta_0 + \beta_1 \tilde{m} + \beta_2 \tilde{m}^2$.

While Model (i) is that of Black–Scholes, Models (ii) and (iii) prescribe the second and fourth order polynomials models with respect to the moneyness, and Models (iv) and (vi) include both the moneyness and the time to maturity. Model (vi) is a quadratic model in terms of the standardized moneyness $\tilde{m} = \ln(m)/\sqrt{\tau}$. These models are commonly used by practitioners as shown in Dumas et al. (1998) and Hentschel (2003).

The specification test was performed for the six models based on the 12 monthly data sets with the number of bootstrap resamples $B = 500$. The six models were all rejected for all 12 months with very small p-values for each month. The test statistics D_n were all way above the empirical bootstrap distributions of D_n, which were the histograms of the bootstrap statistics $\{D_n^{*b}\}_{b=1}^{B}$. This indicates that in reality the form of the volatility function $\sigma(m, \tau)$ is more complicated to model. We may consider other models, in particularly the stochastic volatility (SV) model. However, the current approach cannot be applied to the SV model due to the latent aspect of the volatility. How to extend the approach to such setting is a topic for future research.

Finally, we evaluate the performance of three pricing approaches in terms of in-sample fit, out-of-sample prediction and hedging performance. The three pricing methods are (i) the Black–Scholes, (ii) the ad-hoc Deterministic Volatility (ad-hoc DV) of Dumas et al. (1998), and (iii) the nonparametric pricing by the local linear smoothing $\hat{C}_{ll}(z)$.

The Black–Scholes price at a z is $C_{BS}(z; \hat{\sigma})$ where $\hat{\sigma}$ is obtained by the parametric least square method based on each monthly data. The ad-hoc DV price is $C_{BS}\{\sigma(\hat{\theta}); z\}$ with a quadratic $\sigma(\theta) = \theta_1 m^2 + \theta_2 m + \theta_3$ where the parameter $\theta = (\theta_1, \theta_2, \theta_3)$ were estimated by minimizing

$$\sum_{i=1}^{n} \{C_{BS}^{-1}(\tilde{Y}_i; Z_i) - (\theta_1 M_i^2 + \theta_2 M_i + \theta_3)\}^2 \tag{7.2}$$

as proposed by Dumas et al. (1998).

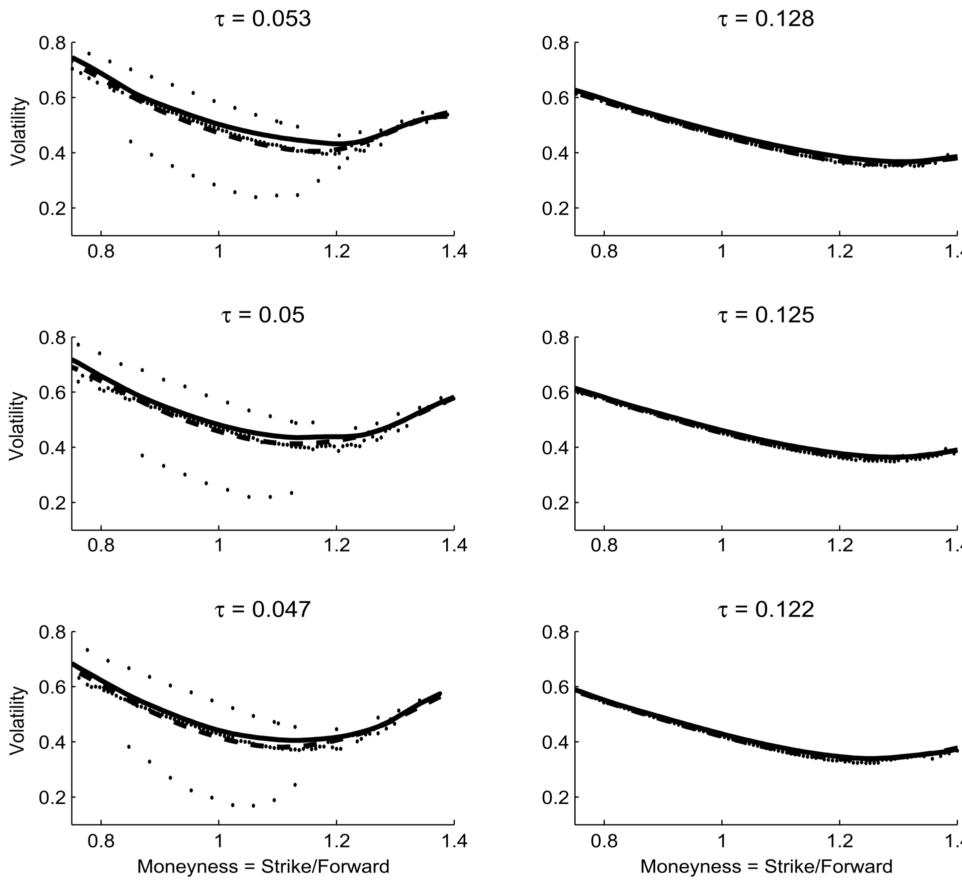


Fig. 9. Volatility estimates on March 2 (top panels), March 3 (middle panels) and March 4, 2009 (bottom panels), matured on March 21 (left panels) and April 18 (right panels). The implied volatility (dots), the kernel implied volatility (dashed lines) and the implied kernel volatility (solid line).

The in-sample fitting errors were the average squared differences between the observed prices and the estimated prices based on each monthly data. The prediction errors were the average squared differences between the observed option prices and the predicted priced based on the previous month's data. Table 3 reports the performance of in-sample fitting and out-of-sample prediction of the three pricing approaches. Regarding the in-sample performance, for both calls and puts, the nonparametric pricing was the best. The ad-hoc DV was better than the Black–Scholes for 6 of the 12 months for calls and 11 of the 12 months for puts.

Table 3 shows that, for the out-of-sample prediction for calls, prediction based on the nonparametric pricing was the best for 8 of the 11 months. For the out-of-sample prediction of puts, there is no 'clear' winner. Prediction based on the ad-hoc DV was the best for 6 of the 11 months, the nonparametric was the best for 4 out of the 11 months, and prediction based on the Black–Scholes was the best for 1 of the 11 months. Note that there were only 11 months available for prediction. The ad-hoc DV had better performance than the Black–Scholes in 7 out of 11 months for calls and 10 out of 11 months for puts.

The hedging performance was measured by the average squared hedging errors, which was used in Dumas et al. (1998) and Fan and Mancini (2009). The hedging error is the difference between changes of option prices under a model and changes of the observed prices. On a day, a hedger has a long position of $h_t = \frac{\partial C_t}{\partial S_t}$ shares of the underlying asset to neutralize the risk associated with a short position of one unit of call option at certain strike price and time to maturity. Assuming the hedging is continuously adjusted, the changes from the underlying asset in a week of 5 business days is $\int_t^{t+5} h_t dS_t = C_{t+5} - C_t$, holding other factors constant. Thus the hedging error for a week starting from t is $(Y_{t+5} - Y_t) - (C_{t+5} - C_t)$, where Y_t and Y_{t+5} are observed option prices and C_t and C_{t+5} are calibrated prices by a model or an approach. We aggregated the squared hedging errors into each of the 12 months. The average hedging errors over a monthly period $[T_1, T_2]$ is

$$\frac{1}{\sum_{t=T_1}^{T_2} n_t} \sum_{t=T_1}^{T_2} \sum_{j=1}^{n_t} [(Y_{(t+5),j} - Y_{t,j}) - \{C(X_{(t+5),j}) - C(X_{t,j})\}]^2, \quad (7.3)$$

where n_t is the number of option pairs at day t.

The hedging performance of the three methods are reported in Table 4. We note that the four months from the August 15th had the highest hedging errors, which was peaked in the 7th month from the September 15th. It is observed that for 9 out of the 12 months, for both calls and puts, the nonparametric approach was the best among the three, and for the remaining 3 months, it was the second best. The nonparametric approach had much less hedging errors than the other two (for some months less than 50%). The Black–Scholes formula had better hedging performance than the ad-hoc DV for all the 12 months for the calls, and 11 out of the 12 months for puts. This replicated the finding of Dumas et al. (1998)'s empirical study on S&P 500 data. Hence, regarding the hedging performance, the nonparametric pricing approach was better than the Black–Scholes, which was better than the ad-hoc DV pricing approach.

8. Conclusion

We analyze the properties of the conventional implied volatility based on a single price inversion and find that it can be subject

Table 3
Average squared in-sample fitting errors and prediction errors multiplied by 10^5 for the Black–Scholes (BS), ad-hoc deterministic volatility (ad-hoc DV), and the nonparametric kernel estimator using cross-validation bandwidth.

Month	In-sample-fit			Out-of-sample-prediction		
	BS	ad-hoc DV	Kernel	BS	ad-hoc DV	Kernel
(a) Calls						
1	1.167	0.528	0.235	NA	NA	NA
2	0.710	0.278	0.096	2.582	1.920	2.068
3	0.983	1.233	0.222	1.010	0.597	0.327
4	0.893	1.079	0.179	1.264	0.982	0.710
5	0.691	0.840	0.181	0.691	0.459	0.390
6	0.761	1.391	0.183	0.768	2.749	0.332
7	9.485	12.183	2.070	22.613	28.567	21.653
8	7.178	5.847	1.290	21.714	25.287	19.787
9	6.419	8.931	1.377	7.707	3.902	4.079
10	3.903	3.736	0.486	18.207	38.595	23.310
11	3.450	3.379	0.668	3.568	3.390	1.141
12	2.669	2.630	0.661	2.942	1.750	1.271
(b) Puts						
1	1.269	0.463	0.242	NA	NA	NA
2	0.798	0.183	0.102	2.677	1.627	2.077
3	1.067	0.363	0.223	1.090	0.402	0.321
4	0.956	0.263	0.166	1.338	0.834	0.714
5	0.750	0.371	0.187	0.750	0.295	0.397
6	0.816	0.290	0.173	0.825	0.471	0.366
7	9.971	11.363	2.078	22.704	21.067	21.978
8	8.205	4.560	1.326	22.730	12.194	18.689
9	7.030	3.498	1.342	8.290	3.330	31.824
10	4.198	1.702	0.505	17.765	25.565	23.094
11	3.693	1.600	0.672	3.805	1.200	1.151
12	2.986	1.922	0.685	3.243	0.991	1.781

Table 4
Average squared hedging errors for the Black–Scholes (BS), ad-hoc deterministic volatility (ad-hoc DV), and the nonparametric kernel estimator using cross-validation bandwidth.

Month	BS	ad-hoc DV	Kernel
(a) Calls			
1	10.392	10.355	6.787
2	2.859	3.425	2.729
3	9.314	10.158	9.777
4	4.816	5.710	5.312
5	3.541	4.398	3.874
6	39.092	83.474	18.502
7	83.372	96.556	42.232
8	36.415	48.843	20.566
9	35.535	53.846	21.294
10	16.990	19.290	7.003
11	9.183	9.795	6.553
12	8.202	8.783	7.279
(b) Puts			
1	10.522	10.778	6.933
2	2.993	3.667	2.925
3	9.141	11.223	9.842
4	4.338	5.382	5.064
5	3.769	4.280	4.135
6	37.941	80.171	18.041
7	76.820	92.772	41.278
8	34.631	55.790	21.349
9	33.956	48.024	20.912
10	16.846	19.709	7.246
11	8.678	8.458	6.239
12	8.516	8.838	7.609

to a systematic bias in the presence of pricing errors. As the pricing errors are encountered in option data, this makes the implied volatility adversely affected. The proposed implied kernel (IK) estimation is robust against the pricing errors in the options. When the error is small, the IK volatility estimates may be close to the conventional IV or the KI estimates. However, when the errors are substantial, the IK can filter out the errors and produces more accurate volatility estimates. Hence, we view our proposal as an effort to make the volatility estimation robust. Our analysis also shows no theoretical support for the practice of implied volatility inversion for short maturity at the money options, namely inverting those prices will not exempt from the adverse impacts of pricing errors.

Although our study was motivated by the work of Hentschel (2003), we emphasis in this paper on the statistical properties of the implied volatility and propose an alternative volatility estimator via nonparametric kernel estimation of the option price function. Given the importance of the implied volatility in decoding the underlying volatility among the option prices, it is crucial to use a reliable and robust volatility estimator. Our proposed implied kernel volatility is such an estimator which can capture the underlying volatility for a wide range of underlying price dynamics and option formulae. As demonstrated in the study of the S&P 500 option data, the proposed volatility estimator can be used to calibrate the underlying volatility function from option data via the specification test procedure.

Acknowledgments

We are grateful to two referees and the Editor Professor Yacine Aït-Sahalia for constructive suggestions which have improved the content and the presentation of the paper. Chen acknowledges the support of National Science Foundation of China Key Grant 11131002 and Grant 71371016.

References

Aït-Sahalia, Y., Duarte, J., 2003. Nonparametric option pricing under shape restrictions. Journal of Econometrics 116, 9–47.
Aït-Sahalia, Y., Lo, A., 1998. Nonparametric estimation of state-price densities implicit in financial asset prices. Journal of Finance 53, 499–547.
Bakshi, G., Cao, C., Chen, Z., 1997. Empirical performance of alternative option pricing models. Journal of Finance 52, 2003–2049.
Bates, D.S., 1996. Testing option pricing models. In: Maddala, G.S., Rao, C.R. (Eds.), Statistical Methods in Finance. Elsevier, Amsterdam, pp. 567–611.
Bates, D.S., 2000. Post-87 crash fears in S&P 500 futures options market. Journal of Econometrics 94, 181–238.
Black, F., Scholes, M., 1973. The pricing of options and corporate liabilities. Journal of Political Economy 81, 637–654.
Chen, S.X., Gao, J., 2007. An adaptive empirical likelihood test for parametric time series regression models. Journal of Econometrics 141, 950–972.
Chen, S.X., Xu, Z., 2013. On implied volatility for options—some reasons for smile and more to correct. Technical Report, Center for Statistical Science, Peking University.
Derman, E., Kani, I., 1994. Riding on the smile. Risk 7, 32–39.
Dumas, B., Fleming, J., Whaley, R.E., 1998. Implied volatility functions: empirical tests. Journal of Finance 53, 2059–2106.
Dupire, B., 1994. Pricing with a smile. Risk 7, 18–20.
Fan, J.Q, Gijbels, I., 1996. Local Polynomial Modelling and its Applications. Chapman and Hall, London.
Fan, J.Q., Mancini, L., 2009. Option pricing with model-guided nonparametric methods. Journal of the American Statistical Association 104, 1351–1372.
Härdle, W., 1990. Applied Nonparametric Regression. Cambridge University Press, Cambridge.
Härdle, W., Mammen, E., 1993. Comparing nonparametric versus parametric regression fits. Annals of Statistics 21, 1926–1947.
Hentschel, L., 2003. Errors in implied volatility estimation. Journal of Financial and Quantitative Analysis 38, 779–810.
Heston, S., 1993. A closed-form solution of options with stochastic volatility with applications to bond and currency options. The Review of Financial Studies 6, 327–343.
Hull, J.C., 1999. Options, Futures and Other Derivatives, fourth ed. Prentice Hall, Upper Saddle River, New Jersey.
Hull, J., White, A., 1987. The pricing of options on assets with stochastic volatilities. Journal of Finance 42, 281–300.
Jorion, P., 1995. Predicting volatility in the foreign exchange market. Journal of Finance 50, 507–528.
Macbeth, J., Merville, L., 1979. An empirical examination of the Black–Scholes call option pricing model. Journal of Finance 34, 1173–1186.
Mayhew, S., 1995. Implied volatility. Financial Analysts Journal 51, 8–20.

Merton, R.C., 1973. Theory of rational option pricing. The Bell Journal of Economics and Management Science 4, 141–183.
Merton, R.C., 1976. Option pricing when underlying stock returns are discontinuous. Journal of Financial Economics 3, 125–144.
Nadaraya, E.A., 1964. On estimating regression. Theory of Probability and Its Applications 10, 157–159.
Pan, J., 2002. The Jump-risk premia implicit in options: evidence from an integrated time-series study. Journal of Financial Economics 63, 3–50.
Rubinstein, M., 1985. Nonparametric tests of alternative option pricing models using all reported trades and quotes on the 30 most active CBOE option classes from August 23, 1976 through August 31, 1978. Journal of Finance 40, 455–480.
Rubinstein, M., 1994. Implied binomial trees. Journal of Finance 49, 771–818.
Ruppert, D., Wand, M.P., 1994. Multivariate locally weighted least squares regression. Annals of Statistics 22, 1346–1370.
Watson, G.S., 1964. Smoothing regression analysis. Sankhya: The Indian Journal of Statistics: Series A 26, 359–372.