Inference for variance risk premium

Shuang Zhang and Song Xi Chen
Guanghua School of Management, Peking University, Beijing, China, and
Lei Lu
Asper School of Business, University of Manitoba, Winnipeg, Canada

Received 30 April 2020
Revised 26 May 2020
Accepted 30 May 2020

Abstract

Purpose – With the presence of pricing errors, the authors consider statistical inference on the variance risk premium (VRP) and the associated implied variance, constructed from the option prices and the historic returns.
Design/methodology/approach – The authors propose a nonparametric kernel smoothing approach that removes the adverse effects of pricing errors and leads to consistent estimation for both the implied variance and the VRP. The asymptotic distributions of the proposed VRP estimator are developed under three asymptotic regimes regarding the relative sample sizes between the option data and historic return data.
Findings – This study reveals that existing methods for estimating the implied variance are adversely affected by pricing errors in the option prices, which causes the estimators for VRP statistically inconsistent. By analyzing the S&P 500 option and return data, it demonstrates that, compared with other implied variance and VRP estimators, the proposed implied variance and VRP estimators are more significant variables in explaining variations in the excess S&P 500 returns, and the proposed VRP estimates have the smallest out-of-sample forecasting root mean squared error.
Research limitations/implications – This study contributes to the estimation of the implied variance and the VRP and helps in the predictions of future realized variance and equity premium.
Originality/value – This study is the first to propose consistent estimations for the implied variance and the VRP with the presence of option pricing errors.

Keywords Implied variance, Variance risk premium, Pricing errors, Kernel estimation
Paper type Research paper

1. Introduction

Investors when investing in risky assets risk losing their money because of uncertainty of the future value of their investment. Risk premium is a compensation to the uncertainty or risk borne by the investors on their risky investments. It is a function of the asset return variance. However, the asset's return is stochastic and uncertain in its own right. There are considerable studies studying the time-varying return variance, such as the deterministic volatility model of Dumas *et al.* (2002), and the stochastic volatility model of Heston (1993). Variance risk premium (VRP), as a compensation for the random variation associated with the asset return variance, is quantified as the difference between the implied variance (IV) and the realized variance (RV) of the asset. The IV is the risk-neutral expectation of the total return variance, which can be empirically estimated from option prices on the asset. And the RV is the expectation of the total return variance under the physical measure.

The existence of the VRP has been studied in the literature. Using a sample of S&P 500 options and indices, Bakshi and Kapadia (2003) examined the statistical properties of the delta-hedged option portfolios to show existence of the VRP. Carr and Wu (2009) measured the VRP as a difference between the RV and the variance swap rate, which is identical to IV,

JEL Classification — C50, C58
Funding: This research is partially funded by China's National Key Research Special Program Grant 2016YFC0207701, National Key Basic Research Program Grant 2015CB856000 and National Natural Science Foundation of China grants 71532001 and 71371016.
The authors acknowledge support from LMEQF at Peking University.

China Finance Review International

2044-1398
DOI 10.1108/CFRI-04-2020-0044

and found that the VRP (RV-IV) was strongly negative for the S&P 500 and Dow Jones indices, that is, the counterpart of the swap who paid the RV and received IV actually got compensated for bearing the risk of uncertain variance. Bollerslev *et al.* (2009) analyzed the theoretical linkage between the volatility and the expected returns in financial markets via a general equilibrium model. They found that the variance of return variance contributed to the return premium as a compensation for bearing volatility risk via a general equilibrium model, and showed that the VRP was able to explain a nontrivial fraction of the variation in the post-1990 aggregate stock market returns.

Although the presence of the VRP is well established, its statistical inference requires further investigation. As the VRP is the difference between the IV and RV, the estimation of the VRP is naturally made by estimating the IV and RV, respectively. For the IV, Breeden and Litzenberger (1978) showed that the risk-neutral distribution of the underlying asset price can be extracted from option prices. Based on this finding, Britten-Jones and Neuberger (2000) derived a model-free expression of the IV by assuming that the asset price process follows a diffusion process. Jiang and Tian (2005) extended the model-free method to the asset price process with jumps and developed an IV estimator using observed option prices via the cubic spline method. For the estimation of the RV, Andersen *et al.* (2001) proposed an approach based on the sum of squared log-return of assets using high-frequency data. But there are market microstructure noises in the high-frequency data. Hansen and Lunde (2006) showed that the estimator of Andersen *et al.* (2001) was biased especially when the sampling frequency goes to infinity, the estimation of the RV is dominated by the microstructure noise. Zhang *et al.* (2005) proposed a bias-adjusted estimator to remove the adverse effects of the microstructure noise in the high-frequency data.

However, pricing errors are widely present in option prices, due to the bid-ask spread, the nonsynchronicity between the option and spot markets, and the measurement errors as revealed in Hentschel (2003) and Chen and Xu (2014). We show that the pricing errors can cause some existing IV estimators inconsistent, which include the estimator of Jiang and Tian (2005) and the VIX adopted by the Chicago Board Options Exchange (CBOE). In this work, we propose a consistent IV estimator by constructing a kernel estimator (Nadaraya, 1964; Watson, 1964) of the option price function to eliminate the adverse effects of the pricing errors. Combining our IV estimator and Zhang *et al.* (2005)'s RV estimator, we propose a consistent VRP estimator, whose asymptotic distributions are studied under three regimes of relative sample sizes between the option data and the historic return data. We theoretically and numerically show that the estimator is consistent despite the pricing errors, which is shown not the case for the other existing IV estimators. In empirical study, the behavior of our estimators for the IV and the VRP via S&P 500 index and option data is investigated. We studied the volatility forecast and found that the proposed IV estimator could explain more variation in RV than the estimator of Jiang and Tian (2005), the Black-Scholes (Black and Scholes, 1973) IV and the lagged daily RV. We also considered regression analysis of the excessive the S&P 500 returns on three VRP estimates, the RV and the other five commonly used covariates. It was found that the proposed VRP estimator was more significant among the three VRP estimates considered, contained the most information of the excess S&P 500 returns and had the smallest out-of-sample forecast root mean squared error when compared to the other two VRP estimates.

This chapter is organized as follows. Section 2 lays out the theoretical framework on the IV. Section 3 analyzes the effects of the option pricing errors on the IV estimation. A consistent IV estimator is proposed and its statistical properties are established in Section 4. Section 5 studied the asymptotic properties of the proposed VRP estimator. Simulation studies are reported in Section 6. Section 7 reports the empirical study on S&P 500 index and option data. Section 8 concludes.

2. Implied variance and existing estimators

Let us first outline the notion of IV. Let $\{S_t\}_{t=0}^{T}$ be the price of an underlying asset, and $\{\Omega_t\}_{t=0}^{T}$ be the accompanying sequence σ-fields such that $\{S_t, \Omega_t\}$ forms a martingale. The IV of S_t between the current date 0 and a future date T is the expectation of the integrated squared returns over $[0, T]$ under a risk neutral measure Q

$$\mathrm{IV}_T = E^Q\left[\int_0^T \left(\frac{dS_t}{S_t}\right)^2 dt\right].$$

The IV can be expressed via the European call option prices without relying on a specific option pricing model (Britten-Jones and Neuberger, 2000; Jiang and Tian, 2005). Let $C(S_0, K, T, r)$ be the price function of a European call option at the current date 0 with the underlying asset spot price S_0, the strike price K, the time to maturity T and the risk-free interest rate r. Without loss of generality, we assume the form of the call option price function $C(S_0, K, T, r)$ is unknown but to be "rational" as specified in Merton (1973) and it is homogeneous of degree 1, namely,

$$\frac{C(S_0, K, T, r)}{S_0} = C(1, M, T, r),$$

where $M = K/S_0$ is the moneyness. In the rest of the paper, we will abbreviate $C(1, M, T, r)$ as $C(M, T)$ and assume $S_0 \equiv 1$, and then the IV

$$\mathrm{IV}_T = 2e^{rT}\int_0^\infty \frac{C(M, T) - \max(0, 1 - Me^{-rT})}{M^2} dM. \tag{1}$$

There are two existing estimators of the IV. In 1993, CBOE introduced a stock market volatility index called VXO, which was developed by Whaley (1993) based on the implied volatility of the options on the S&P 100 index. It represents the market's aggregate expectation of volatility over the next 30 days. In 2003, CBOE made several significant changes and introduced a revamped market volatility index called VIX, which was based on the IV developed by Demeterfi *et al.* (1999), also called the fair value of future variance. The VIX is constructed on the S&P 500 index, and has become an important tracking index for investors.

In empirical applications, because the future dividend payments are difficult to determine, the forward price is calibrated via the put-call parity using liquid calls and puts with the moneyness closest to at-the-money. CBOE pairs the calls and puts with the same moneyness and time to maturity, and find the moneyness M^* at where the absolute difference between the call and put prices is the smallest, to be the moneyness closest to at-the-money. Let $P(M, T)$ and $C(M, T)$ be the put and call option prices with moneyness M and maturity T, respectively. Then, the forward index price FI is estimated by the put-call parity

$$\mathrm{FI} = M^* + e^{rT}\left[C(M^*, T) - P(M^*, T)\right].$$

If there are multiple pairs of call and put prices satisfying the above criteria, the average of the inferred forward prices is used. Let M_0 be the first moneyness below FI. The IV developed by Demeterfi *et al.* (1999) is

$$\mathrm{IV}_T^{\mathrm{DDKZ}} = 2e^{rT}\left[\int_0^{M_0} \frac{P(M, T)}{M^2} dM + \int_{M_0}^{\infty} \frac{C(M, T)}{M^2} dM\right] + 2\left[rT - \frac{e^{rT}}{M_0} + 1 - \ln(M_0)\right] \tag{2}$$

Assuming zero dividends payments, Jiang and Tian (2007) showed that IV_T^{DDKZ} was identical to (1). Applying Taylor expansion to the last term in (2) and noting $FI = e^{rT}$, (2) was approximately

$$IV_T^{DDKZ} \approx 2e^{rT}\left[\int_0^{M_0} \frac{P(M,T)}{M^2}dM + \int_{M_0}^{\infty} \frac{C(M,T)}{M^2}dM\right] - \left(\frac{FI}{M_0} - 1\right)^2. \quad (3)$$

Let $Q(M, T)$ denote the out-of-the-money option prices

$$Q(M,T) = \begin{cases} P(M,T) \text{if } M \in (0, M_0), \\ \frac{P(M,T)+C(M,T)}{2} \text{if } M = M_0, \\ C(M,T) \text{if } M \in (M_0, \infty). \end{cases}$$

Then (3) becomes

$$IV_T^{DDKZ} \approx 2e^{rT}\int_0^{\infty} \frac{Q(M,T)}{M^2}dM - \left(\frac{FI}{M_0} - 1\right)^2. \quad (4)$$

However, in the option market, only a finite number of strike prices are actually traded. Let $M_{\min}$ and $M_{\max}$ be the minimum and maximum moneyness of available options, respectively. Figure 1 gives the kernel density estimations of moneyness for SPX call options traded in years 2006 and 2007, respectively. It shows that the minimum (maximum) moneyness was around 0.2 (1.4).

Let N_l be the number of selected options used in the calculation, M_i be the moneyeness of the ith out-of-the-money option, $\Delta M_i = (M_{i+1} - M_{i-1})/2$ be the increment between moneyness. Employing the Riemann Sum to approximate the first term in (4), CBOE (2003) estimates the truncated IV as

$$IV_T^{CBOE} = 2e^{rT}\sum_{i=1}^{N_l} \frac{\Delta M_i}{M_i^2} Q(M_i, T) - \left(\frac{FI}{M_0} - 1\right)^2. \quad (5)$$

Jiang and Tian (2007) showed that the CBOE procedure might lead to several types of approximation errors, including truncation, discretization, expansion and interpolation errors. They proposed a set of solutions to fix the problems. They firstly inverted the prices of those listed call options to the Black-Scholes (BS) implied volatilities via the BS price formula

$$C_{BS}(S_0, K, T, r) = N(d_1)S_0 - N(d_2)Ke^{-rT},$$

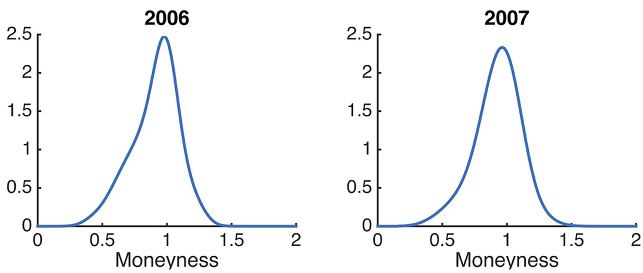

Figure 1. The kernel density estimates for the density of the moneyness of the SPX call options traded in 2006 and 2007, with the Gaussian kernel and the smooth bandwidth $h_{m1} = 0.073$ and $h_{m2} = 0.015$, respectively

where $d_1 = 1/\sigma\sqrt{T}[\ln(S_0/K) + (r + \sigma^2/2)T]$, and $d_2 = d_1 - \sigma\sqrt{T}$, σ is the underlying volatility, and S_0, K, T and r are as specified before.

Suppose that at a given date, there are n_T option contracts of a underlying asset in a market, and the risk-free interest rate is fixed, such that the ith observed option price be Y_i at $Z_i = (M_i, T)$ with the moneyness M_i and the fixed time to maturity T. The BS implied volatility of the ith observed option is

$$\hat{\sigma}_I(Z_i) = C_{\mathrm{BS}}^{-1}(Y_i; Z_i).$$

Jiang and Tian (2005) then used the cubic spline method to smooth $\hat{\sigma}_I(Z_i)$ with respect to the moneyness M_i to obtain a smooth BS implied volatility surface $\widehat{\sigma_{\mathrm{JT}}}(Z)$. Specifically, let $v(\cdot)$ be a function constructed by the linear combination of piecewise cubic polynomials, such that

$$v(\cdot) = \operatorname*{argmin}_{v \in V} \sum_{i=1}^{n_T} [v(Z_i) - \hat{\sigma}_I(Z_i)]^2.$$

For more details on the asymptotic properties of the cubic spline, see Zhou *et al.* (1998). Then, $\widehat{\sigma_{\mathrm{JT}}}(Z) = v(Z)$ is the smoothed BS implied volatility function proposed by Jiang and Tian (2005). Their approach was similar to Ait-Sahalia and Lo (1998), except that Jiang and Tian (2005) used the cubic spline smoother, while Ait-Sahalia and Lo (1998) employed the kernel smoother. Then, the call price with the moneyness in $[M_{\min}, M_{\max}]$ can be obtained via

$$\widehat{C}_{\mathrm{JT}}(Z) = C_{\mathrm{BS}}[\widehat{\sigma_{\mathrm{JT}}}(Z); Z].$$

For options with moneyness in $[0, M_{\min}]$ and $[M_{\max}, \infty]$, Jiang and Tian (2005) used the estimated BS implied volatility at $M_{\min}$ and $M_{\max}$ to extrapolate option values, respectively,

$$\widehat{C}_{\mathrm{JT}}(Z) = C_{\mathrm{BS}}[\widehat{\sigma_{\mathrm{JT}}}(M_{\min}, T); M, T] \text{ for } M \in (0, M_{\min}),$$
$$\widehat{C}_{\mathrm{JT}}(Z) = C_{\mathrm{BS}}[\widehat{\sigma_{\mathrm{JT}}}(M_{\max}, T); M, T] \text{ for } M \in (M_{\max}, \infty).$$

And then they applied the trapezoidal rule to estimated

$$\widehat{\mathrm{IV}}_{\mathrm{JT}} = 2e^{rT} \int_0^\infty \frac{\widehat{C}_{\mathrm{JT}}(Z) - \max(0, 1 - Me^{-rT})}{M^2} dM. \quad (6)$$

3. Effects of pricing errors

We shall study the statistical properties of the IV estimators in the presence of option pricing errors. The pricing errors are widely present in the option data, due to bid-ask spread, nonsynchronicity between options and spot markets, discreteness in quoted prices and other random errors as elaborated in Hentschel (2003) and Chen and Xu (2014). Suppose there are n_T observed option prices with maturity T within an observation period. Recall that $Z_i = (M_i, T)$ and let $C(Z_i)$ be the underlying price function as the interest rate is assumed to be fixed. Then, the observed standardized (by division of the spot price) option prices Y_i admit

$$Y_i = C(Z_i) + \epsilon_i \text{ for } i = 1, \ldots, n_T, \quad (7)$$

where ϵ_i is the standardized pricing error with $E(\epsilon_i|Z_i) = 0$ and finite conditional variance $\sigma_\epsilon^2(Z_i)$.

There are obviously option prices at maturities other than T. Similar to Jiang and Tian (2005), we consider here a situation when the different maturities are well separated, which is especially the case when Model (7) is for a short observational period, say a day or a few days. We will discuss the prospect of combining options with different maturities in Section 4 when we propose our estimator for the IV. Model (7) was considered in Ait-Sahalia and Lo (1998). They constructed a nonparametric estimator for the state price density based on the relationship between the state-price density and the option prices. Ait-Sahalia and Lo (1998) estimated the implied volatility by first inverting the option price, and then doing the kernel smoothing of the implied volatility. Chen and Xu (2014) reversed the order by first obtaining the kernel smooth estimator $\widehat{C}(Z)$ to the option price function $C(Z)$, followed by inverting each price with $\widehat{C}(Z)$, which will make the estimation consistent.

Let $\widehat{\sigma_I}(Z_i)$ be the commonly used implied volatility by directly inverting the standardized option price Y_i such that $\widehat{\sigma_I}(Z_i) = C_{\mathrm{BS}}^{-1}(Y_i; Z_i)$, and $\sigma_I(Z_i) = C_{\mathrm{BS}}^{-1}(C(Z_i)_i; Z_i)$ be the underlying implied volatility. As $C_{\mathrm{BS}}^{-1}(C(Z_i)_i; Z_i)$ is infinitely differentiable with respected to C, if ϵ_i has finite third moment, then by Taylor expansion, we have,

$$\begin{aligned}\widehat{\sigma_I}(Z_i) &= C_{\mathrm{BS}}^{-1}[C(Z_i) + \epsilon_i; Z_i] \\ &= \sigma_I(Z_i) + \frac{\partial C_{\mathrm{BS}}^{-1}[C(Z_i); Z_i]}{\partial C}\epsilon_i + \frac{1}{2}\frac{\partial^2 C_{\mathrm{BS}}^{-1}[C(Z_i); Z_i]}{\partial C^2}\epsilon_i^2 + \frac{1}{3!}\frac{\partial^3 C_{\mathrm{BS}}^{-1}[C(Z_i) + a\epsilon_i; Z_i]}{\partial C^3}\epsilon_i^3 \quad (8)\end{aligned}$$

for some $a \in (0,1)$. If the error is unbiased so that $E(\epsilon_i|Z_i) = 0$, the conditional bias of $\widehat{\sigma_I}(Z_i)$

$$\begin{aligned}b(Z_i) &=: E[\widehat{\sigma_I}(Z_i)|Z_i] - \sigma_I(Z_i) \\ &= \frac{1}{2}\frac{\partial^2 C_{\mathrm{BS}}^{-1}[C(Z_i); Z_i]}{\partial C^2}E(\epsilon_i^2|Z_i) + \frac{1}{3!}E\left\{\frac{\partial^3 C_{\mathrm{BS}}^{-1}[C(Z_i) + a\epsilon_i; Z_i]}{\partial C^3}\epsilon_i^3|Z_i\right\},\end{aligned} \quad (9)$$

which is impacted by the second and the third conditional moments of the pricing error as well as the second and third derivatives of the inverse of the pricing function. This means that the commonly used implied volatility $\widehat{\sigma_I}(Z_i)$ is biased in the presence of pricing errors in the option. Taking conditional variance on (8)

$$\begin{aligned}\mathrm{Var}[\widehat{\sigma_I}(Z_i)|Z_i] = &\left\{\frac{\partial C^{-1}[C(Z_i); Z_i]}{\partial C}\right\}^2 E(\epsilon_i^2|Z_i) + \frac{1}{4}\left\{\frac{\partial C_{\mathrm{BS}}^{-1}[C(Z_i); Z_i]}{\partial C}\right\}^2 \mathrm{Var}(\epsilon_i^2|Z_i) \\ &+ \frac{1}{36}\mathrm{Var}\left\{\frac{\partial^3 C_{\mathrm{BS}}^{-1}[C(Z_i) + a\epsilon_i; Z_i]}{\partial C^3}\epsilon_i^3\middle|Z_i\right\} \\ &+ \frac{1}{2}\left\{\frac{\partial C_{\mathrm{BS}}^{-1}[C(Z_i); Z_i]}{\partial C}\frac{\partial^2 C_{\mathrm{BS}}^{-1}[C(Z_i); Z_i]}{\partial C^2}\right\}E(\epsilon_i^3|Z_i) \\ &+ \frac{1}{6}\frac{\partial C_{\mathrm{BS}}^{-1}[C(Z_i); Z_i]}{\partial C}E\left\{\frac{\partial^3 C_{\mathrm{BS}}^{-1}[C(Z_i) + a\epsilon_i; Z_i]}{\partial C^3}\epsilon_i^4\middle|Z_i\right\} \\ &+ \frac{1}{12}\frac{\partial^2 C_{\mathrm{BS}}^{-1}[C(Z_i); Z_i]}{\partial C^2}\mathrm{Cov}\left\{\epsilon_i^2, \frac{\partial^3 C_{\mathrm{BS}}^{-1}[C(Z_i) + a\epsilon_i; Z_i]}{\partial C^3}\epsilon_i^3\middle|Z_i\right\},\end{aligned}$$

which suggests that the variance does not necessarily get smaller as the sample size is increased, since $\widehat{\sigma_I}(Z_i)$ is attained based on a single price. Both the conditional bias and variance indicate that the commonly used BS implied volatility $\widehat{\sigma_I}(Z_i)$ is not statistically consistent.

Recall that Jiang and Tian (2005) employed the cubic spline method to estimate an implied volatility function from $\widehat{\sigma_I}(Z_i)$. Assuming the interest rate is fixed, let M be the predictor variable, $\widehat{\sigma_I}(Z)$ be the response variable. Then, Jiang and Tian (2005) estimated the implied volatilities $\widehat{\sigma_{JT}}(Z)$ from (8) via cubic spline method,

$$E[\widehat{\sigma_{\mathrm{JT}}}(Z_i)] - \sigma_I(Z_i) = b(Z_i) + o\left(h_{n_T}^4\right),$$

where $b(Z_i)$ is defined in (9). Because $b(Z_i)$ does not get smaller as the sample size is increased, the estimator $\widehat{\sigma_{\mathrm{JT}}}(Z_i)$ is asymptotically biased to $\sigma_I(Z_i)$, which indicates that the estimator $\widehat{\sigma_{\mathrm{JT}}}(Z_i)$ is not consistent to $\sigma_I(Z_i)$.

Then Jiang and Tian (2005) used the BS model to transform the extracted implied volatilities $\widehat{\sigma_{\mathrm{JT}}}(Z_i)$ into call option prices $\widehat{C}_{\mathrm{JT}}(Z) = C_{\mathrm{BS}}[\widehat{\sigma_{\mathrm{JT}}}(Z); Z]$. By Taylor expansion, for $\alpha \in (0, 1)$,

$$\begin{aligned}\widehat{C}_{\mathrm{JT}}(Z_i) &= C_{\mathrm{BS}}[\sigma(Z_i)] + \frac{\partial C_{\mathrm{BS}}[\sigma(Z_i)]}{\partial\sigma}[\widehat{\sigma_{\mathrm{JT}}}(Z_i) - \sigma(Z_i)] + \frac{1}{2}\frac{\partial^2 C_{\mathrm{BS}}[\sigma(Z_i)]}{\partial\sigma^2}[\widehat{\sigma_{\mathrm{JT}}}(Z_i) - \sigma(Z_i)]^2 \\ &\quad + \frac{1}{3!}\frac{\partial^3 C_{\mathrm{BS}}\{\sigma(Z_i) + \alpha[\widehat{\sigma_{\mathrm{JT}}}(Z_i) - \sigma(Z_i)]\}}{\partial\sigma^3}[\widehat{\sigma_{\mathrm{JT}}}(Z_i) - \sigma(Z_i)]^3\end{aligned}$$

The conditional bias of $\widehat{C}_{\mathrm{JT}}(Z_i)$ given Z_i is

$$\begin{aligned}E[\widehat{C}_{\mathrm{JT}}(Z_i) - C(Z_i) &= \frac{\partial C_{\mathrm{BS}}[\sigma(Z_i)]}{\partial\sigma}b(Z_i) + \frac{1}{2}\frac{\partial^2 C_{\mathrm{BS}}[\sigma(Z_i)]}{\partial\sigma^2}b^2(Z_i) \\ &\quad + \frac{1}{2}\frac{\partial^2 C_{\mathrm{BS}}[\sigma(Z_i)]}{\partial\sigma^2}\,\mathrm{Var}\,[\widehat{\sigma_{\mathrm{JT}}}(Z_i)|Z_i] \\ &\quad + \frac{1}{3!}E\left\{\frac{\partial^3 C_{\mathrm{BS}}[\sigma(Z_i) + \alpha(\widehat{\sigma_{\mathrm{JT}}}(Z_i) - \sigma(Z_i))]}{\partial\sigma^3}[\widehat{\sigma_{\mathrm{JT}}}(Z_i) - \sigma(Z_i)]^3\right\},\end{aligned}$$

where $b(Z_i)$ is defined in (9). The bias is related to $b(Z_i)$ and the first three conditional moments of $\widehat{\sigma_{\mathrm{JT}}}(Z_i)$, and does not diminish to zero as the number of options n_T is increased.

The conditional variance of the estimator $\widehat{C}_{\mathrm{JT}}(Z_i)$ is

$$\mathrm{Var}[\widehat{C}_{\mathrm{JT}}(Z_i)|Z_i] \approx \left\{\frac{\partial C_{\mathrm{BS}}[\sigma(Z_i)]}{\partial\sigma}\right\}^2 \mathrm{Var}[\widehat{\sigma_{\mathrm{JT}}}(Z_i)|Z_i],$$

which does not get reduced as n_T is increased. Thus, this together with the analysis on the bias suggest that $\widehat{C}_{\mathrm{JT}}(Z_i)$ is inconsistent to $C(Z)$. This inconsistency will lead to inconsistency of $\widehat{\mathrm{IV}}_{\mathrm{JT}}$ as shown in the following.

Substituting $\widehat{C}_{\mathrm{JT}}(Z)$ into the IV formula (1), the conditional bias of the estimated IV is

$$E(\widehat{\mathrm{IV}}_{\mathrm{JT}}|Z_i) = 2e^{rT}\int_0^\infty M^{-2}\{E[\widehat{C}_{\mathrm{JT}}(Z)|Z_i] - C(Z)\}dM,$$

which does not converge to zero as the sample size is increased, since the integrand does not due to bias of $\widehat{C}_{\mathrm{JT}}(Z)$. In fact, the bias of $\widehat{\mathrm{IV}}_{\mathrm{JT}}$ is related to $b(Z_i)$ and the first three conditional moments of $\widehat{\sigma_{\mathrm{JT}}}(Z_i)$ given Z_i. Thus, the estimator $\widehat{\mathrm{IV}}_{\mathrm{JT}}$ is not consistent to IV_T.

As VIX has become an important volatility tracking index for investors, we analyze the effects of the pricing error on VIX as well. Under Model (7) with the pricing errors, the IV estimator by CBOE as given in (5) becomes

$$\widehat{\mathrm{IV}}_{\mathrm{CBOE}} = 2e^{rT}\sum_{i=1}^{N_l}\frac{\Delta M_i}{M_i^2}\left[Q(M_i,T)+\epsilon_i\right] - \left(\frac{\mathrm{FI}}{M_0}-1\right)^2$$
$$= 2e^{rT}\sum_{i=1}^{N_l}\frac{\Delta M_i}{M_i^2}Q(M_i,T) - \left(\frac{\mathrm{FI}}{M_0}-1\right)^2 + 2e^{rT}\sum_{i=1}^{N_l}\frac{\Delta M_i}{M_i^2}\epsilon_i. \quad (10)$$

Thus, the pricing errors have little effect on the bias as the last term in (10) is unbiased. However, the term generates a variance that does not diminish as the sample size gets large. These are well reflected in a simulation study reported in Section 6.

4. Error-corrected IV estimator

We propose a new method to obtain a consistent estimation for the model-free IV in the presence of the price errors in the options. Our proposal is based on the kernel smooth estimation of the option price function to remove the adverse effects of the option pricing errors.

Let $K(\cdot)$ be a univariate symmetric probability density function, which is called the kernel. For some integer $r \geq 2$ and $\kappa \neq 0$, the rth order kernel satisfies $\int |K(u)|du < \infty$ and

$$\mu_j(k) = \int u^j K(u)du = \begin{cases} 1 \text{ if } j = 0, \\ 0 \text{ if } 1 \leq j \leq r-1, \\ k \text{ if } j = r. \end{cases}$$

The standardized observed option prices Y_i follow Model (7). Assuming fixed interest rate, the NW kernel estimator of the option price function $C(M,\ T)$ with respect to the moneyness M is

$$\widehat{C}_{\mathrm{NW}}(M,T) = \frac{\sum_{i=1}^{n_T} K_{h_m}(M_i - M)Y_i}{\sum_{i=1}^{n_T} K_{h_m}(M_i - M)}, \quad (11)$$

where $K_h(\cdot) = \frac{1}{h}K(\cdot/h)$, h_m is the bandwidths for smoothing with respected to the moneyness M.

To avoid the boundary bias, the local linear least square kernel estimator (Fan, 1992) is attained by first acquiring

$$\left(\widehat{\theta}_0, \widehat{\theta}_1\right) = \underset{\theta_0,\theta_1}{\operatorname{argmin}} \sum_{i=1}^{n_T}\left[Y_i - \theta_0 - (M_i - M)\theta_1\right]^2 K_{h_m}(M_i - M), \quad (12)$$

and then local linear estimator $\widehat{C}_{\mathrm{LL}}(M,T) = \widehat{\theta}_0$.

The proposed IV estimator based on the kernel smoothing is

$$\widehat{\mathrm{IV}}_T = 2e^{rT}\int_0^\infty \frac{\widehat{C}(M,T) - \max(0, Me^{-rT})}{M^2}dM, \quad (13)$$

where $\widehat{C}(M,T)$ is either the NW kernel estimator (11) or the local linear estimator (12).

As we have mentioned in Section 2, the above approach that conducts the kernel smoothing with respect to moneyness M while fixing the maturity at T is designed for the situation where there are well separations in the observed maturity times in the option data when we select a short observational window like a day or so. If we are willing to extend the observational window to, say, a month, the observed maturities τ_i will be more densely distributed. In such cases, the bivariate kernel smoothing in (11) and (12) may be administrated to achieve more information combining. In this paper, we will concentrate on the case where the observed time to maturities are well separated (line one week, one month or three months) at a particular cross-section of time, and the bivariate kernel smoothing is not needed. Nevertheless, our study will provide insight to the more densely distributed maturity case.

The following conditions are assumed in the theoretical analysis of the proposed IV estimation.

(A1) The smoothing bandwidths satisfy $h_m \to 0$ and $nh_m \to \infty$.

(A2) The option price function $C(M, T)$ and $f(M, T)$, the probability density function of M at T, both have continuous second derivatives with respect to M. The density function $f(M, T)$ is compactly supported with respect to M on $[M_{\min}, M_{\max}]$ for two real numbers $M_{\min} < 1 < M_{\max}$.

(A3) The conditional variance of the pricing error $\mathrm{Var}(\epsilon_i|M_i, T) = \sigma^2(M_i, T)$ is finite, and have continuous second derivatives with respect to M.

Let $K^{(2)}$ be the convolution of K and $MK^{(2)}(t) = \int\int uK(u)K(t+u)du$. The following theorem provides general results on the bias and variance of the proposed $\widehat{\mathrm{IV}}_T$.

Theorem 1. Under assumptions (A1)–(A3), if we employ the NW kernel estimator of the option prices (11), the bias of the $\widehat{\mathrm{IV}}_T$.

$$b(\widehat{\mathrm{IV}}_T) = e^{rT}\mu_r(k)h_m^r \int_{M_{min}}^{M_{\max}} \frac{1}{M^2}\left\{2f^{-1}(M)\frac{\partial C(M,T)}{\partial M}\frac{\partial f(M,T)}{\partial M} + \frac{\partial^2 C(M,T)}{\partial M^2}\right\}dM + o\left(h_m^r\right).$$

and the variance of $\widehat{\mathrm{IV}}_T$

$$\mathrm{Var}(\widehat{\mathrm{IV}}_T) = \frac{4e^{rT}}{n_T}\int_{M_{min}}^{M_{\max}} \frac{\sigma_\epsilon^2(M,T)}{M^4 f(M,T)}dM + O\left(n_T^{-1}h_m\right). \tag{14}$$

If we employ the local linear estimator of the option prices (12), the bias and the variance of the $\widehat{\mathrm{IV}}_T$ are, respectively,

$$b(\widehat{\mathrm{IV}}_T) = e^{rT}\mu_r(k)h_m^r \int_{M_{min}}^{M_{\max}} \frac{1}{M^2}\frac{\partial^2 C(M,T)}{\partial M^2}dM + o\left(h_m^r\right) \text{ and } \mathrm{Var}(\widehat{\mathrm{IV}}_T) = \frac{4e^{rT}}{n_T}\int_{M_{\min}}^{M_{\max}} \frac{\sigma_\epsilon^2(M,T)}{M^4 f(M,T)}dM + O\left(n_T^{-1}h_m\right).$$

It is noted that the variance of the IV estimator is of order n_T^{-1} rather than $(n_T h_m)^{-1}$ as in the univariate kernel regression estimation (Hardle, 1990). This is because the integration of the moneyness used in the definition of the IV leads to improvement of the convergence rate. Theorem 1 suggests that the bias and variance of $\widehat{\mathrm{IV}}_T$ diminish to zero as $n_T \to \infty$ and $h_m \to 0$. These indicate that $\widehat{\mathrm{IV}}_T$ is a consistent estimator of IV_T.

Let

$$I_1(T) = \int_{M_{min}}^{M_{\max}} \frac{\sigma_\epsilon^2(M,T)}{M^4 f(M,T)} dM \text{ and}$$

$$I_2(T) = \int_{M_{min}}^{M_{\max}} \frac{1}{M^2}\left\{2f^{-1}(M,T)\frac{\partial C(M,T)}{\partial M}\frac{\partial f(M,T)}{\partial M} + \frac{\partial^2 C(M,T)}{\partial M^2}\right\} dM.$$

The mean squared error (MSE) of $\widehat{\mathrm{IV}}_T$ is

$$\mathrm{MSE}(\widehat{\mathrm{IV}}_T) = e^{2rT}\left\{\frac{4}{n_T}I_1(T) + \left[h_m^r \mu_2(k) I_2(T)\right]^2\right\}[1 + o(1)],$$

which reveals that the smaller h_m, the smaller MSE of $\widehat{\mathrm{IV}}_T$. If we select the bandwidth h_m such that $n_T h_m^{2r} = o(1)$, the leading order of MSE $(\widehat{\mathrm{IV}}_T)$ is n_T^{-1} and the bias due to the kernel smoothing is negligible.

To study the asymptotic distribution of $\widehat{\mathrm{IV}}_T$, we need to consider how to remove the impact of the bias $b(\widehat{\mathrm{IV}}_T)$ in the kernel smoothing estimators for the IV. There are in general two approaches for the bias correction. One is to assign smaller h_m to make the squared bias a smaller order of the variance so that $n_T h_m^{2r} = o(1)$ (Chen, 1996), and the other is to correct for the bias explicitly by estimating $b(\widehat{\mathrm{IV}}_T)$. We consider the first approach due to its simplicity. It can be implemented by using a kernel of the second order in the bandwidth selection that will result to a $h_m = O(n_T^{-1/5})$, attained for instance by the cross-validate method. Then, we employ a forth order kernel ($r = 4$) in the kernel estimators for $C(Z)$ with the above selected h_m so that the square bias $b^2(\widehat{\mathrm{IV}}_T)$ is at the order of $h_m^8 = O(n_T^{-8/5})$, which is a smaller order of n_T^{-1}.

To gain a quick empirical evidence for the theoretical results, as part of the full simulation study reported in Section 6, we conducted a simulation study to compare the proposed estimator with Jiang and Tian (2005)'s estimator based on 2000 simulations. We assumed that the underlying asset price process followed (22), the observed option price was (24), with the volatility of the relative error specified in (25). Figure 2 plots the average biases of the two estimated option prices $\widehat{C}(Z)$ and $\widehat{C}_{\mathrm{JT}}(Z)$ and the bias of the two estimated integrands $[\widehat{C}(Z) - \max(0, 1 - Me^{-rT})]/M^2$ and $[\widehat{C}_{\mathrm{JT}}(Z) - \max(0, 1 - Me^{-rT})]/M^2$ when the time to maturity $T = 1/12$. Figure 2 indicates that the biases of $\widehat{C}(Z)$ and $[\widehat{C}(Z) - \max(0, 1 - Me^{-rT})]/M^2$ were decreased as the sample size got larger, while neither of the biases of $\widehat{C}_{\mathrm{JT}}(Z)$ nor $[\widehat{C}_{\mathrm{JT}}(Z) - \max(0, 1 - Me^{-rT})]/M^2$ got smaller, which confirmed the theoretical analysis made in this and last sections. When the moneyness was decreased, the price of call option was increased, and the volatility of the pricing error as specified in (25), which was highly correlated with the option price, was increased as well. Thus, when the moneyness was smaller than 0.95, we found that the bias of Jiang and Tian's estimated option price got larger as the moneyness was decreased.

Theorem 2. Under the assumptions (A1)–(A3), if we select a second order kernel and conduct the undersmoothing with the bandwidth $h_m = o(n_T^{-1/4})$, or a higher-order kernel in the NW or LL estimator and the bandwidth at the order of $n_T^{-1/5}$, then as $n_T \to \infty$.

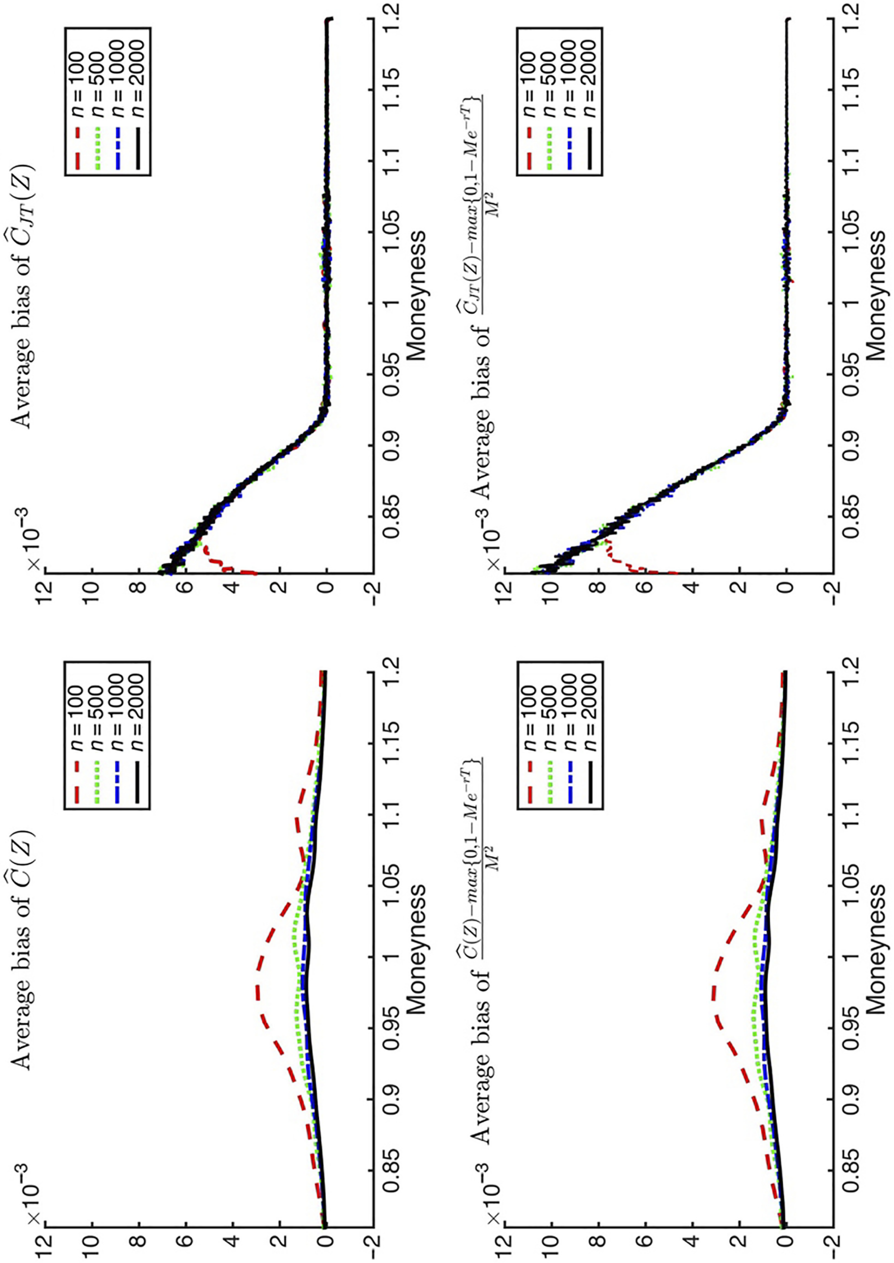

Figure 2.
Average biases of the estimated option prices $\widehat{C}(Z)$ and $\widehat{C_{JT}}(Z)$, and the estimated integrands $[\widehat{C}(Z) - \max(0, 1 - Me^{-rT})]/M^2$ and $[\widehat{C_{JT}}(Z) - \max(0, 1 - Me^{-rT})]/M^2$ at $T = 1/12$. Based on 2000 simulations

$$\sqrt{n_T}(\widehat{\mathrm{IV}}_T - \mathrm{IV}_T) \xrightarrow{d} N\left(0, v_{\mathrm{NW}}^2\right), \tag{15}$$

where

$$v_{\mathrm{NW}}^2 = 4e^{2rT} \int_{M_{min}}^{M_{\max}} \frac{\sigma_\epsilon^2(M, T)}{M^4 f(M, T)} dM.$$

To employ this theorem for various statistical inference, we need to estimate v_{NW}^2 in (15), which requires estimation of $\sigma_\epsilon^2(M, T)$ and $f(M, T)$ separately. The probability density function $f(M, T)$ can be estimated by a univariate kernel density estimator (Silverman, 1986),

$$\widehat{f}(M, T) = \frac{1}{n_T} \sum_{i=1}^{n_T} K_{h_m}(M_i - M).$$

To estimate $\sigma_\epsilon^2(M, T)$, we consider kernel smoothing of the estimated squared residuals $\widehat{r}_i^{\,2} = [Y_i - \widehat{C}(M_i, T)]^2$, where $\widehat{C}(M_i, T)$ is the N–W estimator defined in (11). Then $\sigma_\epsilon^2(M, T)$ can be estimated by applying the local linear estimator with $\widehat{r}_i^2$ as the response variable,

$$\left(\widehat{\delta}_0, \widehat{\delta}_1\right) = \underset{\delta_0, \delta_1}{\operatorname{argmin}} \sum_{i=1}^{n_T} \left[\widehat{r}_i^2 - \delta_0 - (M_i - M)\delta_1\right]^2 K_{h_m}(M_i - M),$$

and $\widehat{\sigma}_\epsilon^{\,2}(M, T) = \widehat{\delta}_0$. Thus, an estimation to the variance of v_{NW}^2 is

$$\widehat{v}_{\mathrm{NW}}^2 = 4e^{2rT} \int_{M_{\min}}^{M_{\max}} \frac{\widehat{\sigma}_\varepsilon^{\,2}(M, T)}{M^4 \widehat{f}(M, T)} dM. \tag{16}$$

It can be shown that (Silverman, 1986; Fan and Yao, 1998) under certain regularity conditions, $\widehat{f}(M, T) \xrightarrow{p} f(M, T)$ and $\widehat{\sigma}_\epsilon^{\,2}(M, T) \xrightarrow{p} \sigma_\epsilon^2(M, T)$ as $n_T \to \infty$. Applying the mapping theorem and the Slutsky lemma, as $n_T \to \infty$, we have

$$\sqrt{n_T}\widehat{v}_{\mathrm{NW}}^{-1}(\widehat{\mathrm{IV}}_T - \mathrm{IV}_T) \xrightarrow{d} N(0, 1). \tag{17}$$

5. Variance risk premium

Following Bollerslev *et al.* (2009), we measure the VRP as the difference between the model-free implied variance IV_T over the $[0, T]$ time interval inferred from the option prices and the realized variance RV_{-T} over the $[-T, 0]$ time interval based on historic high frequency return data.

Before discussing our proposed VRP estimator, let us first review the existing RV estimators. Suppose the asset log-price process follows an Ito process

$$d\log S_t = \mu_t d_t + \sigma_t dW_t, \tag{18}$$

where the drift component μ_t and the instantaneous volatility σ_t are continuous stochastic processes, W_t is a standard Brownian motion. The RV between a past date $-T$ and the current date 0 is the expected sum of squared returns under the physical measure $\mathbb{P}$,

$$\mathrm{RV}_{-T} \equiv E^{\mathbb{P}}\left[\int_{-T}^{0} \left(\frac{dS_t}{S_t}\right)^2 dt\right].$$

Let Δ be the sampling frequency, $m_T = T/\Delta$ be the sample size of return, and $\{p_t\}_{t=0}^{m_T}$ be the observed asset prices between the dates $-T$ and 0. The log-return of the asset over a small time interval is $r_t = \log p_{t+1} - \log p_t$. Andersen *et al.* (2001) proposed an estimator of the RV,

$$\widehat{\mathrm{RV}}_{-T} = \sum_{j=1,\cdots,m_T} r_j^2. \tag{19}$$

When the sampling frequency $\Delta \to 0$, the estimator $\widehat{\mathrm{RV}}_{-T}$ is consistent for RV_{-T} if there is no asset pricing error. However, high-frequency data encounter the market microstructure noise in which is caused by the price discreteness, infrequent trading and bid-ask bounce effects. Suppose that the observed asset price $\{p_t\}_{t\geq 0}$ at the sampling times is

$$\log p_t = \log S_t + u_t,$$

where $\log S_t$ is the theoretical log-price, u_t is the independent microstructure noise with $E(u_t|S_t) = 0$ and $\mathrm{Var}(u_t|S_t) = \sigma_u^2$. Hansen and Lunde (2006) showed that the estimator in (19) is biased,

$$E\left(\widehat{\mathrm{RV}}_{-T}\right) - \mathrm{RV}_{-T} = E\left[2\sum_{t=1}^{m_T}(\log S_{t+1} - \log S_t)u_t + \sum_{t=1}^{m_T} u_t^2\right] = 2m_T\sigma_u^2.$$

When the sampling frequency of the returns $\Delta \to 0$, the estimation of the RV is dominated by the microstructure noise, which makes $\widehat{\mathrm{RV}}_{-T}$ inconsistent to RV_{-T}. To reduce the adverse effect of microstructure noise on the RV estimation, Zhang *et al.* (2005) proposed a bias-corrected estimator, denoted as $\widehat{\mathrm{RV}}_{-T}^{\mathrm{ZMA}}$, for RV_{-T} by combining a lower frequency based RV estimator $\widehat{\mathrm{RV}}_{-T}^{\mathrm{avg}}$ over subsets of return data with the RV_{-T} specified in (19) using all the data. Zhang *et al.* (2005) partition the full grids into K nonoverlapping subsets of lower frequency $K\Delta$. Specifically, for $k = 1, \cdots, K$, the k-th subset

$$\mathrm{G}^{(k)} = \{k, k+K, k+2K, \cdots, k+LK\},$$

where is the integer part of m_k/K. Then the average subsampled RV estimator is

$$\widehat{\mathrm{RV}}_{-T}^{\mathrm{avg}} = \frac{1}{K}\sum_{k=1}^{k}\sum_{j,j+\in \mathrm{G}^{(k)}}\left(\log p_{j+} - \log p_j\right)^2,$$

where $j+ = j+K$. The bias-adjusted estimator proposed by Zhang *et al.* (2005) was

$$\widehat{\mathrm{RV}}_{-T}^{\mathrm{ZMA}} = \widehat{\mathrm{RV}}_{-T}^{\mathrm{avg}} - \frac{L}{m_T}\widehat{\mathrm{RV}}_{-T}. \tag{20}$$

Suppose $K = \mathrm{cm}_T^{2/3}$ and c is a constant, Zhang *et al.* (2005) showed that

$$m_T^{\frac{1}{6}}\left(\widehat{\mathrm{RV}}_{-T}^{\mathrm{ZMA}} - \mathrm{RV}_{-T}\right) \xrightarrow{d} \left(8c^{-2}\sigma_u^4 + c\eta^2 T\right)^{\frac{1}{2}} Z \text{ as } m_T \to \infty, \tag{21}$$

where c is a constant, $\eta^2 = \frac{4}{3}\int_{-T}^{0}\sigma_t^4 dt$ and Z is the standard normal random variable independent of η^2. In other words, $m_T^{1/6}(\widehat{\mathrm{RV}}_{-T}^{\mathrm{ZMA}} - \mathrm{RV}_{-T})$ is asymptotically mixed normally distributed.

To obtain a consistent estimator of the VRP in the presence of the pricing error, we consider using our proposed IV estimator in conjunction with the RV estimator $\widehat{\mathrm{RV}}_{-T}^{\mathrm{ZMA}}$ proposed in Zhang *et al.* (2005), which gives rise to the proposed VRP estimator,

$$\widehat{\mathrm{VRP}}_T = \widehat{\mathrm{IV}}_T - \widehat{\mathrm{RV}}_{-T}^{\mathrm{ZMA}}.$$

Before discussing the asymptotic distribution of $\widehat{\mathrm{VRP}}_T$, we define a notation $a_n \asymp b_n$, which means that $a_n = O(b_n)$ and $b_n = O(a_n)$. Suppose there are n_T options with time to maturity T being observed, the asymptotic variance of $\widehat{\mathrm{IV}}_T$ is discussed in Theorem 1, which diminishes to zero at the rate of $\sqrt{n_T}$. As discussed in Section 4, we can use a high-order kernel in the estimation of $\widehat{\mathrm{IV}}_T$ and choose $h_m \asymp n_T^{-1/5}$. Then, the rate of convergence of $\widehat{\mathrm{IV}}_T$ is of order $\sqrt{n_T}$ while the bias is of a smaller order. Suppose there are m_T high-frequency underlying asset returns within the time horizon T observed for the RV estimation, the asymptotic distribution of $\widehat{\mathrm{RV}}_{-T}^{\mathrm{ZMA}}$ is established in (21) with rate of convergence being $m_T^{1/6}$. When deriving the asymptotic distribution of $\widehat{\mathrm{VRP}}_T$, we need to compare the rates of convergence of $\widehat{\mathrm{IV}}_T$ and $\widehat{\mathrm{RV}}_{-T}^{\mathrm{ZMA}}$, which comes to the relative orders of magnitude of n_T and m_T. We first clarify the relationships between the time horizon T and the sample sizes. The sample size m_T is number of returns between the dates $-T$ and 0, which is related with the horizon T and the data frequency. Figure 3 plots $(m_T)^{1/6}$ (left) with different data frequency (1 min, 5 min and 15 min) and $\sqrt{n_T}$ (right) against the horizon $-T$ and T, respectively. The larger T and the higher frequency, the larger m_T is. And the sample size n_T is the number of observed option prices with maturity T, and dose not matter with the maturity, but related with the variety of option strike prices. We assumed that $\sqrt{n_T}$ to be in the interval $[\sqrt{30}, 10]$, which was the gray box in Figure 3, according to the S&P 500 option data from January 2009 to December 2015.

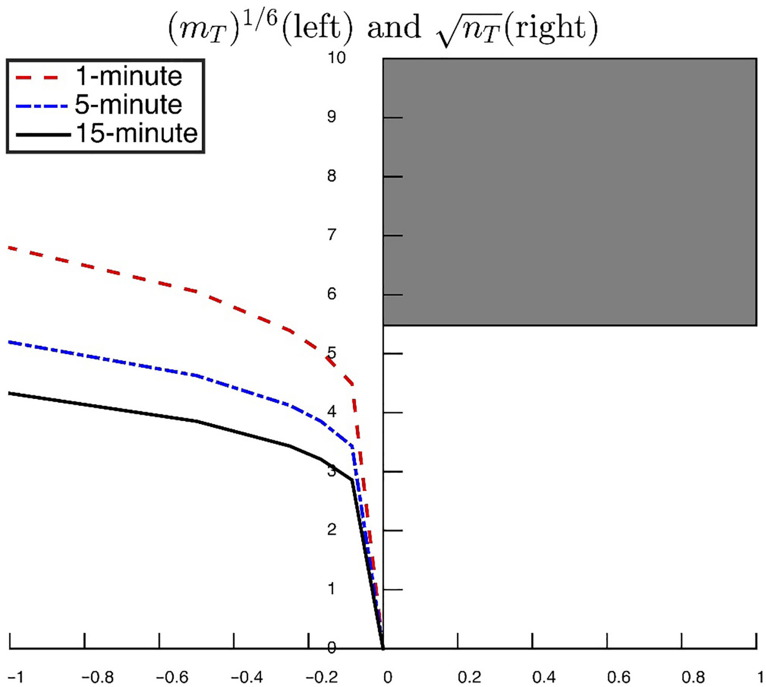

Figure 3.
Plot of $(m_T)^{1/6}$ (left) with different data frequency (1-min, 5-min and 15-min) and $\sqrt{n_T}$ (right) against the horizon $[t-T,\ t]$ and $[t,\ t+T]$, with fixing $t=0$ respectively. The $\sqrt{n_T}$ dose not depend on T, but was assumed to be in the interval $[\sqrt{30},\ 10]$ according to the S&P 500 option data from January 2009 to December 2015

The asymptotic variance of $\widehat{\mathrm{VRP}}_T$ is dominated by the one with the slower order of rate of convergence. By assuming the conditions of Theorem 1 in Section 4 and the conditions of Theorem 4 in Zhang *et al.* (2005), the asymptotic distribution of $\widehat{\mathrm{VRP}}_T$ are considered under three cases.

(1) Case 1: If there are more option prices such that $n_T^{-1/2} m_T^{1/6} \to 0$, then $\widehat{\mathrm{IV}}_T - \mathrm{IV}_T \asymp n_T^{-1/2} = o_p(m_T^{-1/6})$ and $\widehat{\mathrm{VRP}}_T - \mathrm{VRP}_T$ is stochastically dominated by $\widehat{\mathrm{RV}}_{-T}^{\mathrm{ZMA}} - \mathrm{RV}_{-T}$. By applying Slutsky lemma, it can be readily shown that, as $n_T, m_T \to \infty$,

$$m_T^{\frac{1}{6}}\left(\widehat{\mathrm{VRP}}_T - \mathrm{VRP}_T\right) = m_T^{\frac{1}{6}}\left(\widehat{\mathrm{IV}}_T - \widehat{\mathrm{RV}}_{-T}^{\mathrm{ZMA}} - \mathrm{IV}_T + \mathrm{RV}_{-T}\right)$$
$$= m_T^{\frac{1}{6}}\left(\mathrm{RV}_{-T} - \widehat{\mathrm{RV}}_{-T}^{\mathrm{ZMA}}\right) + o_p(1) \xrightarrow{d} \left(8c^{-2}\sigma_u^4 + c\eta^2 T\right)^{\frac{1}{2}} Z,$$

where $Z \sim N(0,1)$ and the other parameters are specified in (21).

(2) Case 2: If $n_T^{-1/2} m_T^{1/6} \to C > 0$, then $\widehat{\mathrm{IV}}_T - \mathrm{IV}_T \asymp n_T^{-1/2}$ and $\widehat{\mathrm{RV}}_{-T}^{\mathrm{ZMA}} - \mathrm{RV}_{-T} \asymp m_T^{1/6}$ are at the same order. Hence, $\widehat{\mathrm{VRP}}_T - \mathrm{VRP}_T$ is stochastically dominated by both $\widehat{\mathrm{IV}}_T - \mathrm{IV}_T$ and $\widehat{\mathrm{RV}}_{-T}^{\mathrm{ZMA}} - \mathrm{RV}_{-T}$ so that as $n_T, m_T \to \infty$,

$$m_T^{\frac{1}{6}}\left(\widehat{\mathrm{VRP}}_T - \mathrm{VRP}_T\right) = m_T^{\frac{1}{6}}\left(\widehat{\mathrm{IV}}_T - \widehat{\mathrm{RV}}_{-T}^{\mathrm{ZMA}} - \mathrm{IV}_T + \mathrm{RV}_{-T}\right) \xrightarrow{d} \gamma_1 Z_1 + \gamma_2 Z_2,$$

Where $\gamma_1 = \left(8c^{-2}\sigma_u^4 + c\eta^2 T\right)^{1/2}$,

$$\gamma_2 = 2e^{rT}\left[\int_{M_{min}}^{M_{\max}} \frac{\sigma_\epsilon^2(M,T)}{M^4 f(M,T)} dM\right]^{1/2}$$

and Z_1 and Z_2 are two independent standard normal random variables.

(3) Case 3: If there are more historic return data such that $n_T^{1/2} m_T^{-1/6} \to 0$, then $\widehat{\mathrm{RV}}_{-T}^{\mathrm{ZMA}} - \mathrm{RV}_{-T} \asymp m_T^{1/6} = o_p(n_T^{-1/2})$ while $\widehat{\mathrm{IV}}_T - \mathrm{IV}_T - b(\widehat{\mathrm{IV}}_T) = O_p(n_T^{-1/2})$. Hence, $\widehat{\mathrm{VRP}}_T - \mathrm{VRP}_T$ is stochastically dominated by $\widehat{\mathrm{IV}}_T - \mathrm{IV}_T$ and as $n_T, m_T \to \infty$

$$\sqrt{n_T}\left(\widehat{\mathrm{VRP}}_T - \mathrm{VRP}_T\right) = \cdot\sqrt{n_T}\left(\widehat{\mathrm{IV}}_T - \widehat{\mathrm{RV}}_{-T}^{\mathrm{ZMA}} - \mathrm{IV}_T + \mathrm{RV}_{-T}\right)$$
$$= \sqrt{n_T}(\widehat{\mathrm{IV}}_T - \mathrm{IV}_T)$$
$$+ o_p(1) \cdot \xrightarrow{d} 2e^{rT}\left[\int_{M_{min}}^{M_{\max}} \frac{\sigma_\epsilon^2(M,T)}{M^4 f(M,T)} dM\right]^{\frac{1}{2}} Z$$

Thus, in all three cases, $\widehat{\mathrm{VRP}}_T$ is a consistent estimator to VRP_T and is asymptotically normally or mixed normally distributed. The variance of $\widehat{\mathrm{VRP}}_T - \mathrm{VRP}_T$ diminishes to zero at the rate of $m_T^{1/6}$ or $\sqrt{n_T}$ depending on the relative order of $m_T^{1/6}$ and $\sqrt{n_T}$.

6. Simulation studies

We report results of simulation experiments which were designed to provide numerical confirmation to our theoretical findings in the previous sections. Besides experimenting our proposed estimators of IV and VRP, we also consider the estimators of Jiang and Tian (2005) and the CBOE for comparison purpose. We compared the relative bias, the standard deviation and the root mean squared error of the three IV estimators, which included the proposed $\widehat{\mathrm{IV}}_T$, $\widehat{\mathrm{IV}}_{JT}$ the estimator in Jiang and Tian (2005), and $\widehat{\mathrm{IV}}_{\mathrm{CBOE}}$ the estimator used by the CBOE.

In the simulation, we consider two models for the underlying asset price process S_t.

(1) Model 1: The two components μ_t and σ_t in (18) were constants so that

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t, \tag{22}$$

with $\mu = 0.05$ and $\sigma = 0.3$.

(2) Model 2:

$$\frac{dS_t}{S_t} = \mu dt + \sigma(S_t, t) dW_t, \tag{23}$$

which was the deterministic volatility model (Dumas *et al.*, 2002) where $\sigma(S_t, t) = 0.3 + 4e^{-t}(S_0/S_t - 1)^2$ was a deterministic function of the asset price and time, and $\mu = 0.05$.

The option prices were generated according to

$$Y_i = C(Z_i) + C(Z_i)\sigma_e(M)e_i, \tag{24}$$

where $C(Z_i)$ was the option price function implied by Models (22) and (23), respectively, $e_i \sim N(0,1)$, and

$$\sigma_e(M) = \sqrt{\left[0.002 + 0.152 \max\left(1 - \frac{|M - 1.328|}{0.319}, 0\right)\right]^{0.437}}, \tag{25}$$

The option maturities were one week, one month and three months, respectively. To motivate the model for generating the moneyness, we have plotted in Figure 1 the kernel density estimation of the moneyness of S&P 500 options, which led us to design the simulated moneyness followed a Beta distribution $M \sim \frac{1}{2}[\mathrm{Beta}(4,\ 4) + 3/2]$. We set the risk-free rates at the three maturities were zero and the underlying asset did not pay any dividend. The sample size n_T for the number of option prices was 100, 300, 500, 700, 1000 and 2000, respectively. For each n_T, the simulation was replicated 2000 times.

Figure 4 plots the bias, the standard deviation (SD) and the root mean squared error (RMSE) of the three IV estimators with the underlying asset price process (23) when $T = 1/12$ and $1/4$. A clear feature conveyed from the figures was that, the bias, the SD and the RMSE of the proposed $\widehat{\mathrm{IV}}_T$ all got smaller as the number of options (sample size) n_T got larger. This confirmed the consistency of the proposed IV estimator. In contrast, the relative bias of $\widehat{\mathrm{IV}}_{\mathrm{JT}}$ did not get smaller as the sample size was increased, while the SD of $\widehat{\mathrm{IV}}_{\mathrm{JT}}$ did got reduced as n_T increased. These were agreeable with the diagnostics we have made in Section 3 due to the adverse effect of the pricing error. The performance of $\widehat{\mathrm{IV}}_{\mathrm{CBOE}}$ was strange in two aspects. One was that the SD and the RMSE could not be reduced by increasing the number of option data at each maturity under both spot price Models (22) and (23). Another was that the bias was quite erratic between the two spot price models and the time to maturities. Under Model (23), the relative biases of $\widehat{\mathrm{IV}}_{\mathrm{CBOE}}$ were all negative, which revealed that $\widehat{\mathrm{IV}}_{\mathrm{CBOE}}$ underestimated IV_T. In all cases, the bias of $\widehat{\mathrm{IV}}_{\mathrm{CBOE}}$ did not admit a diminishing trend as the number of options increases. These rendered the consistency of $\widehat{\mathrm{IV}}_{\mathrm{CBOE}}$.

To gain insight on the erratic performance of $\widehat{\mathrm{IV}}_{\mathrm{CBOE}}$, we decomposed

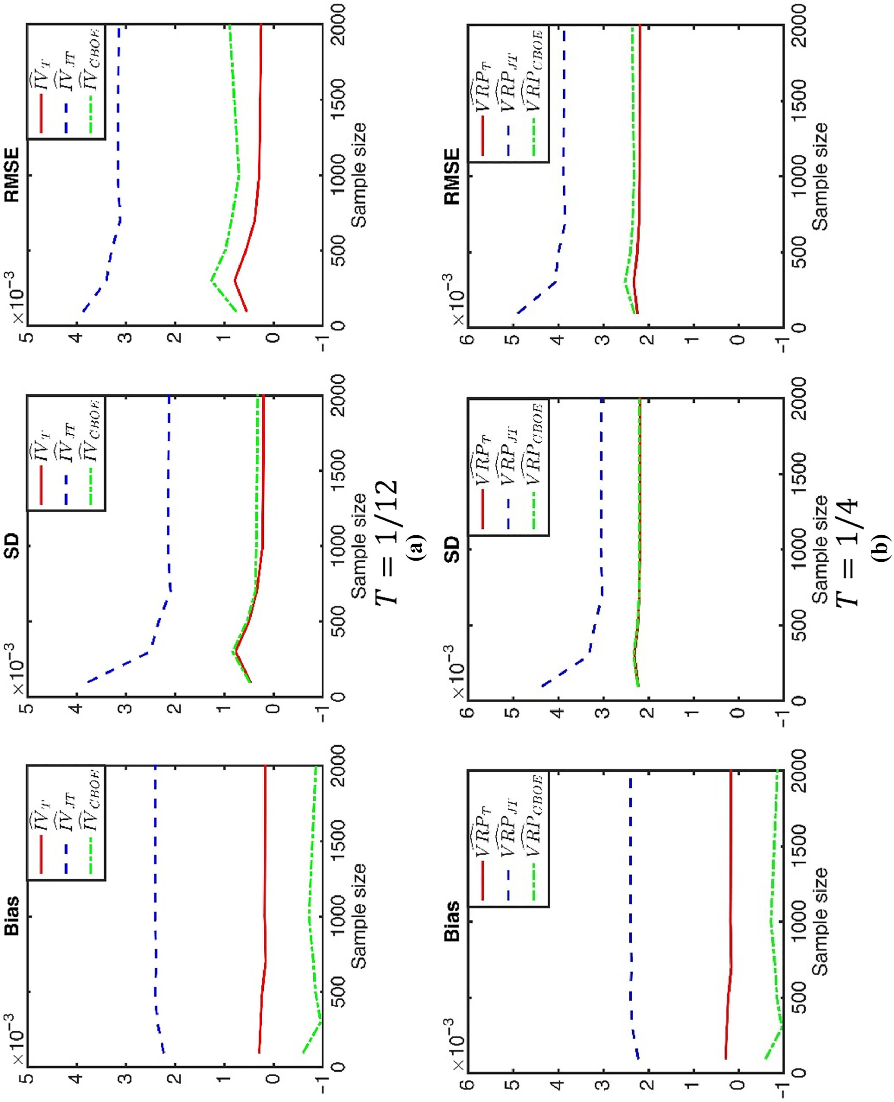

Figure 4.
The bias, the standard deviation (SD) and the root mean squared error (RMSE) for the estimation $\widehat{\mathrm{IV}}_T$, $\widehat{\mathrm{IV}}_{\mathrm{JT}}$ and $\widehat{\mathrm{IV}}_{\mathrm{CBOE}}$ with the underlying asset pricing process (22)

$$\widehat{\mathrm{IV}}_{\mathrm{CBOE}} = \left[2e^{rT}\sum_{i=1}^{N_l}\frac{\Delta M_i}{M_i^2}Q(M_i,T) - \left(\frac{\mathrm{FI}}{M_0}-1\right)^2\right] + 2e^{rT}\sum_{i=1}^{N_l}\frac{\Delta M_i}{M_i^2}\epsilon_i \\ =: \widehat{\mathrm{IV}}_{\mathrm{CBOE1}} + \widehat{\mathrm{IV}}_{\mathrm{CBOE2}},\ \text{say}. \tag{26}$$

Clearly, $\widehat{\mathrm{IV}}_{\mathrm{CBOE2}}$ is influenced by the pricing errors, and is uncorrelated with $\widehat{\mathrm{IV}}_{\mathrm{CBOE1}}$. We conducted simulations to evaluate $\widehat{\mathrm{IV}}_{\mathrm{CBOE1}}$ and $\widehat{\mathrm{IV}}_{\mathrm{CBOE2}}$ separately. Table 1 reports the bias and SD of $\widehat{\mathrm{IV}}_{\mathrm{CBOE1}}$ and $\widehat{\mathrm{IV}}_{\mathrm{CBOE2}}$, respectively, when the time to maturity $T = 1/52, 1/12$ and $1/4$. The SD of $\widehat{\mathrm{IV}}_{\mathrm{CBOE1}}$ was negligible, relative to that of $\widehat{\mathrm{IV}}_{\mathrm{CBOE2}}$. Thus, it was the pricing errors that much elevated the SD of $\widehat{\mathrm{IV}}_{\mathrm{CBOE}}$ although it is largely unbiased.

For the deterministic volatility Model (23), we conducted simulation experiments on the VRP by estimating the IV and the RV separately. Figure 5 provides the results of the three VRP estimators with 1-min asset price data for the RV estimation. The results with 15 min and 5 min asset returns showed the same patterns with Figure 5. Comparing the three VRP estimates, the RMSEs of $\widehat{\mathrm{VRP}}_T$ were the smallest among the three estimators in all cases. With the sample size of option prices used in the IV estimation increased, the bias, the SD and the RMSE of $\widehat{\mathrm{VRP}}_T$ all got smaller, while the bias of $\widehat{\mathrm{VRP}}_{\mathrm{JT}}$ and $\widehat{\mathrm{VRP}}_{\mathrm{CBOE}}$ did not, which agreed with the theoretical consistency analysis reported early. We also noticed that, the biases of $\widehat{\mathrm{VRP}}_{\mathrm{CBOE}}$ were all negative, because $\widehat{\mathrm{IV}}_{\mathrm{CBOE}}$ underestimated IV_T as revealed in Table 1. It is noted that the variance of $\widehat{\mathrm{VRP}}$ can be decomposed to the variances of $\widehat{\mathrm{IV}}$ and $\widehat{\mathrm{RV}}$. As the time horizon increased, the variance of $\widehat{\mathrm{RV}}_{-T}^{\mathrm{ZMA}}$ got larger, and started to dominate the variance of $\widehat{\mathrm{VRP}}$.

7. Empirical results

To demonstrate the application of the proposed IV and VRP estimators, we analyzed the S&P 500 index and option data from January 2009 to December 2015. The S&P 500 option

Table 1.
Bias and standard deviation (SD) of the two components of $\widehat{\mathrm{VI}}_{\mathrm{CBOE}}$

		$\widehat{\mathrm{IV}}_{\mathrm{CBOE1}}$		$\widehat{\mathrm{IV}}_{\mathrm{CBOE2}}$	
T	n_T	Bias ($\times10^{-6}$)	SD ($\times10^{-16}$)	Bias ($\times10^{-6}$)	SD ($\times10^{-2}$)
1/52	100	−2.78	0.23	−31.10	0.26
	300	0.21	0.54	252.36	0.28
	500	0.09	0.43	71.02	0.37
	700	0.01	0.27	−38.39	0.24
	1,000	0.02	0.39	−18.81	0.16
	2,000	0.01	0.23	1.73	0.22
1/12	100	−4.31	1.91	−32.51	0.26
	300	3.05	2.74	251.19	0.58
	500	1.01	1.79	73.97	0.37
	700	0.14	2.73	−38.23	0.24
	1,000	0.23	2.62	−17.25	0.16
	2,000	0.07	2.30	1.61	0.22
1/4	100	−153.65	9.79	−32.68	0.17
	300	−94.59	6.49	250.11	0.58
	500	−50.24	9.96	79.18	0.37
	700	−46.47	2.57	−38.46	0.24
	1,000	−47.15	4.23	−14.41	0.16
	2,000	−25.91	7.84	1.82	0.22

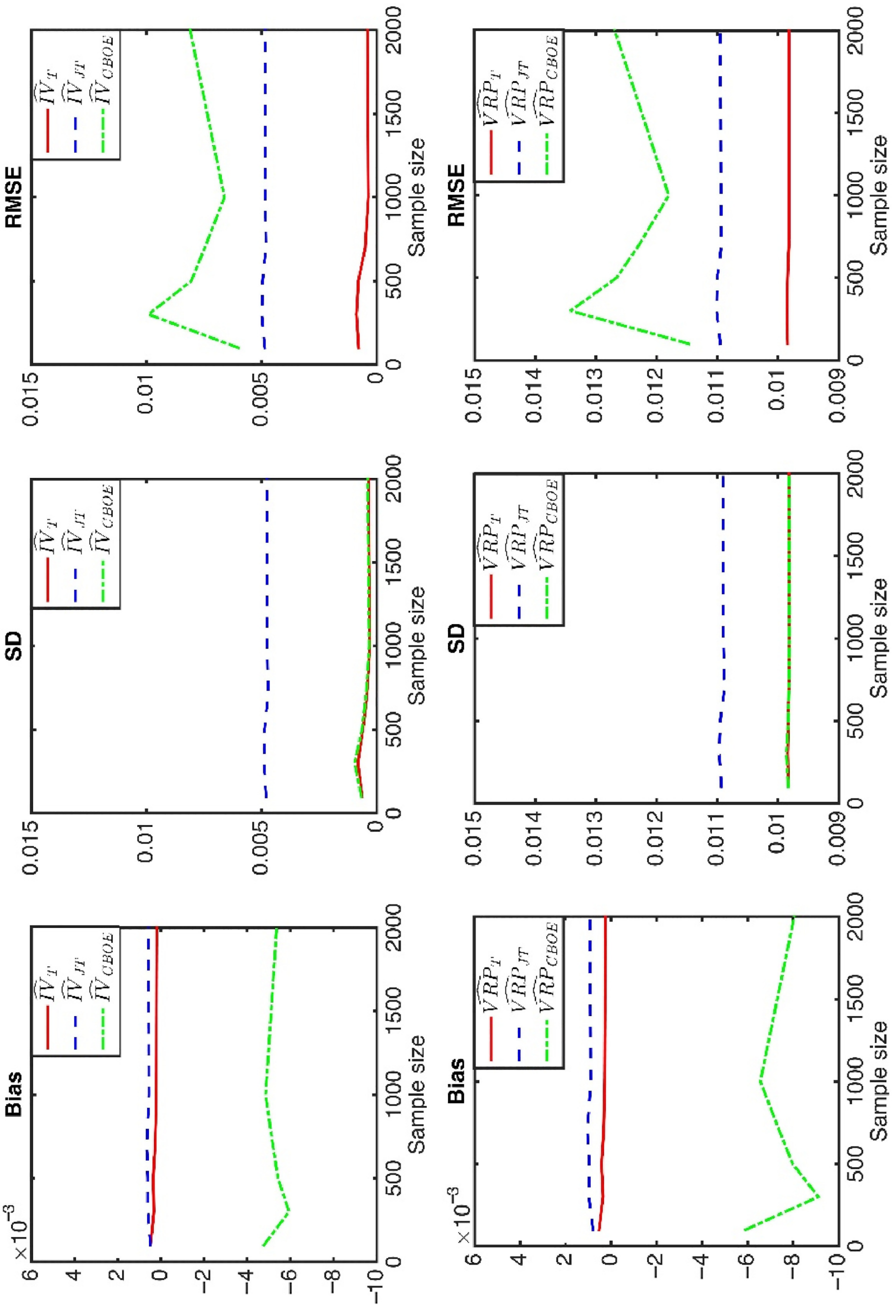

Figure 5.
The bias, the standard deviation (SD) and the root mean squared error (RMSE) for the estimation $\widehat{\mathrm{VRP}}_T$, $\widehat{\mathrm{VRP}}_{\mathrm{JT}}$ and $\widehat{\mathrm{VRP}}_{\mathrm{CBOE}}$ with the underlying asset pricing process (23)

data were obtained from the CBOE, and the S&P 500 index data at 5 min intervals were obtained from the Bloomberg. Daily Treasury bill yields at various maturities were obtained from the US Department of Treasury as proxies to the risk-free rates. We adopted the following rules to filter the option data. The minimum tick was 1/16 for options with prices less than three dollars, and 1/8 for others. Options with prices less than 1/8 were excluded. We excluded the illiquid options with the expiration either less than one week or more than eight months. All selected option prices should satisfy the nonarbitrage principle

$$C(S_0, T, K, r) \geq \max(0, S_0 - K) \text{ and } K \geq P(S_0, T, K, r) \geq \max\ (0, K - S).$$

As mentioned before, we estimated the VRP via estimating the IV and the RV separately. The three IV estimators were $\widehat{\mathrm{IV}}_T$, $\widehat{\mathrm{IV}}_{\mathrm{JT}}$ and $\widehat{\mathrm{IV}}_{\mathrm{CBOE}}$, which were matched with three VRP estimators $\widehat{\mathrm{VRP}}_T$, $\widehat{\mathrm{VRP}}_{\mathrm{JT}}$ and $\widehat{\mathrm{VRP}}_{\mathrm{CBOE}}$, respectively. In addition to those variance measures, we also considered the commonly used Black-Scholes implied volatility. In particular, we estimated the Black-Scholes at-the-money implied volatility using the method of Chen and Xu (2014) by $\widehat{\sigma_{\mathrm{BS}}} = C_{\mathrm{BS}}^{-1}[\widehat{C}(Z)]$ with the moneyness $M = 1$, where $\widehat{C}(Z)$ is the kernel estimator of the option price function given in (11). Using the high-frequency S&P 500 index data, the RV estimator is $\widehat{\mathrm{RV}}_{-T}^{\mathrm{ZMA}}$. All the IV estimates and BS implied volatility estimation were made monthly from January 2009 to November 2015, and the RV estimates were also monthly from February 2009 to December 2015, which resulted in 83 monthly estimates.

Following Canina and Figlewski (1993) and Christensen and Prabhala (1998), to analyze the information content of the volatility forecasts, we considered a set of regression models which regressed the subsequent realized variance $\widehat{\mathrm{RV}}_{-T}^{\mathrm{ZMA}}$ over the time horizon $[0, T]$ on combinations of the BS IV $\widehat{\sigma}_{\mathrm{BS}}^2$ with maturity T, the IV estimates $\widehat{\mathrm{IV}}_T$ and $\widehat{\mathrm{IV}}_{\mathrm{JT}}$ over the time horizon $[0,\ T]$, and the historic variance. Following Jiang and Tian (2005), we used the lagged RV on the latest trading day, denoted as LRV, as a proxy for the historic variance, since it contains the most relevant information for the subsequent RV. The LRV was estimated from the 5-min S&P 500 index returns, say $\{R_i\}_{i=1}^n$, using a formula given by Hansen and Lunde (2006).

$$\widehat{\mathrm{LRV}} = \sum_{i=1}^n R_i^2 + 2\left(\frac{n}{n-1}\right)\sum_{i=1}^n R_i R_{i-1}.$$

Since the univariate regression is the restricted form of the multivariate regression, we consider the following two multivariate regression models:

$$\begin{aligned} \widehat{\mathrm{RV}}_T^{\mathrm{ZMA}} &= \beta_0 + \beta^{\mathrm{IV}}\widehat{\mathrm{IV}}_T + \beta^{\mathrm{BS}}\widehat{\sigma}_{BS}^2 + \beta^{\mathrm{LRV}}\widehat{\mathrm{LRV}} + \epsilon, \widehat{\mathrm{RV}}_T^{\mathrm{ZMA}} \\ &= \beta_0 + \beta_{\mathrm{JT}}^{\mathrm{IV}}\widehat{\mathrm{IV}}_{\mathrm{JT}} + \beta^{\mathrm{BS}}\widehat{\sigma}_{BS}^2 + \beta^{\mathrm{LRV}}\widehat{\mathrm{LRV}} + \epsilon. \end{aligned} \quad (27)$$

Table 2 reports the fitting results of the three sets of the regression models in (27). The results showed that the coefficients of the proposed $\widehat{\mathrm{IV}}_T$ and $\widehat{\mathrm{IV}}_{\mathrm{JT}}$ were mostly significant at 5% or 1% levels, and were all significant at either 10% level. In contrast, the coefficients of the Black-Scholes IV $\widehat{\sigma}_{\mathrm{BS}}^2$ and the lagged realized variance $\widehat{\mathrm{LRV}}$ were largely insignificant in explaining the $\widehat{\mathrm{RV}}_T^{\mathrm{ZMA}}$, especially when IV estimates were included as covariates. Although the IV simply reflects the market's expectation of the future variance of the underlying asset's return, from the univariate regressions for $\widehat{\sigma}_{\mathrm{BS}}^2$ and $\widehat{\mathrm{LRV}}$ (labeled as (1) and (3) in Table 2), we found that the BS IV $\widehat{\sigma}_{BS}^2$ was a poor predictor of the subsequent RV compared with the historic variance $\widehat{\text{LRV}}$. All three panels show that having the proposed $\widehat{\text{IV}}_T$ led to higher R^2 than the corresponding models with $\widehat{\text{IV}}_{\text{JT}}$ as covariate, which suggested that using a consistent IV estimate led to better volatility fitting and forecasting performance.

Bollerslev *et al.* (2009) constructed a general equilibrium model for the stock market, and found that the return variance and the return variance and the volatility of volatility contributed to the return premium as a compensation for bearing risk, while the VRP was generated due to the volatility of volatility. Consequently, the VRP should serve as a useful predictor for the returns over the same horizon. Using S&P 500 index data and the VIX, Bollerslev *et al.* (2009) showed that the VRP was able to explain a nontrivial fraction of the variation in the post-1990 aggregate stock market returns. Following Bollerslev *et al.* (2009), let R_T be the excess return beyond the risk-free interest rates over a time horizon T, namely

$$R_T = \frac{1}{h}\sum_{j=1}^{h} r_j - r,$$

where r was the risk-free interest rate, the unit time interval was one month, h was the number of months over the horizon $[0,\ T]$, and r_j was the return of the j-th month.

We examined the explanatory power of different VRP estimations in predicting the S&P 500 index returns using the following model developed by Bollerslev *et al.* (2009)

$$R_T = \beta_0 + \beta_1 \text{VRP}_T + u, \tag{28}$$

where u was an error term, and $\text{VRP}_T = \text{IV}_T - \text{RV}_{-T}$ following the definition in Bollerslev *et al.* (2009). Figure 6 plots the time series of the three VRP estimations, $\widehat{\text{VRP}}_T$, $\widehat{\text{VRP}}_{\text{JT}}$ and $\widehat{\text{VRP}}_{\text{CBOE}}$, from January 2009 to December 2015. And Table 3 reports the summary statistics of them. The VRP estimator $\widehat{\text{VRP}}_{\text{CBOE}}$ was larger than the other two VRP estimators overall. The standard deviation of $\widehat{\text{VRP}}_T$ was the smallest. And the three VRP estimators were all negatively skewed. In addition to the covariates in (28), we also considered other commonly used variables (see, for example, Lettau; Ludvigson, 2001) to predict the excess returns of the aggregate stock market as measured by S&P 500, which included the price-earnings ratio $\log(P_t/E_t)$, the default spread DFSP_t defined as the difference between Moody's BAA and AAA bond yields, the term spread TMSP_t defined as the difference between the 10-years and 3-month Treasury yields, the stochastically detrended risk-free rate RREL_t defined as the

	β_0	$\beta_{\text{JT}}^{\text{IV}}$	β^{IV}	β^{BS}	β^{LRV}	Adjusted R^2
(1)	0.012** (0.005)	–	–	0.563*** (0.102)	–	0.266
(2)	0.014** (0.004)	–	–	0.087 (0.144)	0.866*** (0.207)	0.395
(3)	0.015** (0.004)	–	–	–	0.983*** (0.132)	0.400
(4)	0.010** (0.004)	0.840*** (0.107)	–	–	–	0.424
(5)	0.011** (0.004)	0.559* (0.254)	–	–	0.373 (0.306)	0.427
(6)	0.009* (0.004)	0.795*** (0.169)	–	0.050 (0.142)	–	0.417
(7)	0.011* (0.004)	0.556* (0.266)	–	0.005 (0.147)	0.371 (0.319)	0.420
(8)	0.006 (0.004)	–	0.719*** (0.089)	–	–	0.439
(9)	0.008 (0.005)	–	0.534* (0.208)	–	0.293 (0.298)	0.439
(10)	0.006 (0.004)	–	0.739** (0.149)	−0.025 (0.148)	–	0.432
(11)	0.008 (0.005)	–	0.564* (0.226)	−0.054 (0.151)	0.314 (0.305)	0.432

Note(s): The numbers inside the parentheses beside the parameter estimates were their standard errors, and *, ** and *** indicate that the coefficients of the regressions were significantly different from zero at the 10%, 5% and 1% level, respectively

Table 2. Ordinary least square estimates of the regression Models (27)

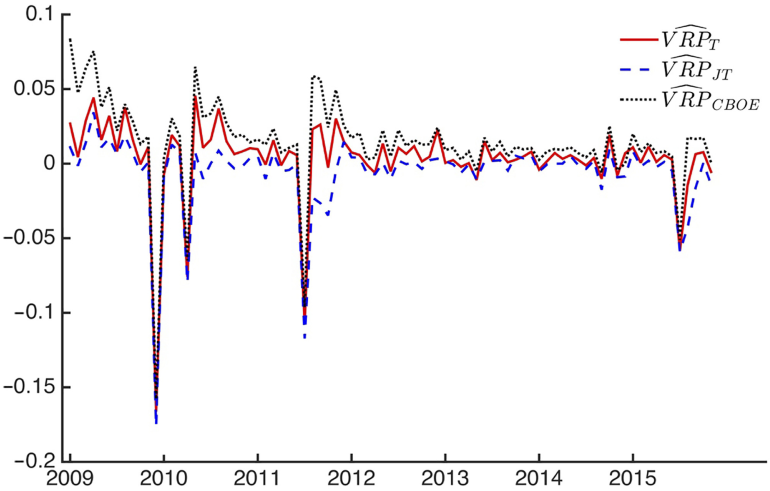

Figure 6. The time series of the three VRP estimators $\widehat{\text{VRP}}_T$, $\widehat{\text{VRP}}_{\text{JT}}$ and $\widehat{\text{VRP}}_{\text{CBOE}}$ from January 2009 to December 2015 based on S&P 500 option data

Table 3. Summary statistics of $\widehat{\text{VRP}}_T$, $\widehat{\text{VRP}}_{\text{JT}}$ and $\widehat{\text{VRP}}_{\text{CBOE}}$ for S&P 500 option from January 2009 to December 2015

	Mean	SD	Min	Max	Skewness	Kurtosis
$\widehat{\text{VRP}}_T$	0.0033	0.0033	−0.1663	0.0457	−3.6653	20.6249
$\widehat{\text{VRP}}_{\text{JT}}$	−0.0059	0.0273	−0.1759	0.0344	−4.0976	23.1002
$\widehat{\text{VRP}}_{\text{CBOE}}$	0.0142	0.0314	−0.1579	0.0831	−2.4012	14.5679

one-month T-bill rate minus the past 12-month moving averages, and the consumption-wealth ratio CAY_t defined in Lettau and Ludvigson (2001), which led to the model

$$R_T = \beta_0 + \beta_1 \text{VRP}_T + \beta_2 \text{RV}_{-T} + \beta_3 \log\left(\frac{P}{E}\right) + \beta_4 \text{DFSP} + \beta_5 \text{TMSP} + \beta_6 \text{RREL} + \beta_7 \text{CAY} + u. \quad (29)$$

Table 4 reports the monthly results of a family of nested regression models based on (29). Comparing the three univariate regressions with the VRP as the covariate (labeled as (1), (2) and (3)), the coefficient of the proposed estimator $\widehat{\text{VRP}}_T$ was the most significant with the adjusted R^2 being the highest (14.79%). This suggested that, among the three VRP estimators, the proposed VRP estimator explained the most of the variations in the excess returns. For the regressions involving the other predictors in (29) (labeled as (4), (5) and (6)), the adjusted R^2's were larger than those of the univariate regressions overall. The coefficient of the proposed $\widehat{\text{VRP}}_T$ in (4) was the most significant, with the highest adjusted R^2 (22.08%) among the three regression models. We conducted variable selection based on the Akaike Information Criterion (AIC). We first calculated the AIC of all possible regressions models, and then selected the model with the smallest AIC, which selected VRP, log (P/E) and RREL, and avoided $\widehat{\text{RV}}_{-T}^{\text{ZMA}}$, DFSP, TMSP and CAY. Regression models (labeled as (7), (8) and (9)) show the results after variable selection. Among the three models after the variable selection, the coefficients of the VRP estimators and the covariate RREL were all significant at the 0.1% level. In particular, Model (7) which included the proposed VRP estimator, had the highest

Table 4. Regression results of Model (29) with three different VRP estimates ($\widehat{\text{VRP}}_T$, $\widehat{\text{VRP}}_{\text{JT}}$ and $\widehat{\text{VRP}}_{\text{CBOE}}$)

Regressors	Models (1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Constant	0.03 (0.60)	0.09 (1.76)	−0.02 (−0.37)	−0.63 (−1.83)	−0.43 (−1.0)	−0.71* (−2.03)	−0.55. (−1.90)	−0.48 (−1.61)	−0.50 (−1.64)
$\widehat{\text{VRP}}$	0.07*** (3.90)			0.07*** (2.58)			0.08*** (4.59)		
$\widehat{\text{VRP}}_{\text{JT}}$		0.06** (3.35)			0.09. (1.93)			0.07*** (4.06)	
$\widehat{\text{VRP}}_{\text{CBOE}}$			0.05** (3.21)			0.05. (1.94)			0.06*** (3.88)
$\widehat{\text{RV}}_{-T}^{\text{ZMA}}$				−0.58 (0.19)	1.08 (0.24)	−3.05 (−1.14)			
$\log(P/E)$				−0.16* (1.07)	0.10 (0.62)	0.16 (1.07)	0.19* (2.04)	0.19* (1.99)	0.15 (1.55)
DFSP				0.10 (0.36)	0.14 (0.47)	0.18 (0.65)			
TMSP				0.03 (0.31)	0.02 (0.18)	0.04 (0.36)			
RREL				−3.2*** (−3.0)	−3.51*** (−3.3)	−3.36*** (−3.09)	−3.00*** (−3.72)	−3.04*** (−3.67)	−3.03*** (−3.62)
CAY				−1.5 (−0.40)	−1.5 (−0.36)	−1.27 (−0.32)			
Adj.R^2(%)	14.79	11.08	10.18	22.08	19.17	19.19	25.77	22.21	21.02

Note(s): The figures inside the parentheses underneath the parameter estimates were the *t*-statistics, and ·, *, ** and *** indicate that the coefficients of the regression were significantly different from zero at the 10%, 5%, 1% and 0.1% level, respectively

Regressors	Models (1)	(2)	(3)	(4)	(5)	(6)
Constant	0.04 (0.61)	0.10 (1.50)	−0.01 (−0.07)	−0.62. (−1.77)	−0.52 (−1.47)	−0.61 (−1.67)
$\widehat{\text{VRP}}$	0.06** (2.88)			0.07** (3.47)		
$\widehat{\text{VRP}}_{\text{JT}}$		0.06** (2.59)			0.07** (3.21)	
$\widehat{\text{VRP}}_{\text{CBOE}}$			0.04** (2.18)			0.06** (2.80)
$\log(P/E)$				0.22. (1.92)	0.21. (1.84)	0.19 (1.62)
RREL				−3.22** (−3.02)	−3.28** (−3.01)	−3.30** (−2.94)
Adj.R^2(%)	11.68	9.42	6.42	22.25	20.13	16.84
Forecasting RMSE	0.282	0.302	0.286	0.262	0.283	0.268
OOS R^2(%)	28.88	18.44	26.85	38.61	28.38	35.77

Note(s): Regression estimates based on the training data set and the prediction RMSE and out-of-sample (OOS) R^2 based on the testing data for six models. The numbers inside the parentheses beside the parameter estimates were their standard errors, and *, ** and *** indicate that the coefficients of the regressions were significantly different from zero at the 10%, 5% and 1% level, respectively

Table 5.
Out-of-sample regression and prediction

adjusted R^2 (25.77%). In conclusion, the proposed $\widehat{\text{VRP}}_T$ is an effective predictor and contained the most information of the excess returns of S&P 500 index.

Then, we conducted the out-of-sample prediction on R_T using the three univariate regression models using the three VRP estimates and the three multivariate regression models with a VRP estimate and the other two selected covariates after the AIC variable selection as specified in Table 4. We split the data into two parts, the training set with the first 56 (monthly) observations and the testing set that consists of the last 27 observations. We first used the training set to estimate the parameters, and then predicted the excessive returns of the test set with the estimated parameters from the training data. Table 5 reports the out-of-sample prediction RMSE. Among the three univariate models, the one with the proposed VRP estimator was the most significant as compared with the other two models with $\widehat{\text{VRP}}_{\text{JT}}$ and $\widehat{\text{VRP}}_{\text{CBOE}}$ as the covariate, as shown by having the largest t-statistics, the smallest forecasting RMSE (0.282) and the largest out-of-sample (OOS) R^2 (0.289). Among the three Models (4), (5) and (6) after the AIC variable selection, the coefficient of the proposed VRP estimator $\widehat{\text{VRP}}_T$ was the most significant when compared with $\widehat{\text{VRP}}_{\text{JT}}$ and $\widehat{\text{VRP}}_{\text{CBOE}}$, and the Model (4) had the smallest forecasting RMSE (0.262) and the largest OOS R^2 (0.386).

In this paper, we have used the RV estimator $\widehat{\text{RV}}_{-T}^{\text{ZMA}}$ proposed by Zhang *et al.* (2005). Other consistent estimators for RV can be also employed to work together with the proposed IV estimator for a consistent VRP estimator. However, Bollerslev *et al.* (2009) used $\widehat{\text{RV}}_{-T}$ defined in (19) as the RV in the estimation of VRP without removing the adverse effect microstructure noise, and showed less significant coefficients comparing with Table 4 with a smaller R^2. The consistent RV estimators help to improve the predictive power of VRP in the return forecasting with the presence of microstructure noise in high-frequency data.

Bollerslev *et al.* (2009) showed that the VRP had the most significant predictive power in the mid-term. We also studied Model (29) in longer horizon. Table 6 reports the quarterly regression results for Model (29), which showed the similar results with Table 4. The t-statistics of the three VRP estimations in quarterly regression results were larger than those in monthly regression results, which implied the more significant predictive power in the mid-term horizon. And the R^2's in Table 6 were also larger than those in Table 4.

Table 6. Quarterly regression results of Model (29) with three different VRP estimates ($\widehat{\text{VRP}}_T$, $\widehat{\text{VRP}}_{\text{JT}}$ and $\widehat{\text{VRP}}_{\text{CBOE}}$)

Regressors	Models (1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Constant	−0.12. (−1.75)	−0.11 (−1.65)	−0.14. (−1.80)	−0.07 (−0.40)	−0.11 (−0.57)	−0.16 (−0.79)	−0.11 (−0.77)	−0.07 (−0.42)	−0.11 (−0.68)
$\widehat{\text{VRP}}$	0.05*** (3.99)			0.06* (2.61)			0.07*** (3.51)		
$\widehat{\text{VRP}}_{\text{JT}}$		0.06*** (3.81)			0.08* (2.46)			0.07* (2.51)	
$\widehat{\text{VRP}}_{\text{CBOE}}$			0.07*** (3.73)			0.05* (2.27)			0.06* (2.33)
$\widehat{\text{RV}}_{-\text{T}}^{\text{ZMA}}$				−0.37 (−0.30)	−0.52 (−0.42)	−0.49 (−0.39)			
$\log(P/E)$				0.20* (2.61)	0.21** (2.77)	0.21** (2.81)	0.17* (2.41)	0.21** (2.88)	0.21** (2.91)
DFSP				−0.27. (−1.96)	−0.27. (−1.95)	−0.26. (−1.92)	−0.25* (−2.30)	−0.29* (−2.49)	−0.29* (−2.41)
TMSP				−0.15** (−2.96)	−0.15** (−2.97)	−0.14** (−2.85)	−0.13** (−2.89)	−0.15** (−3.15)	−0.15** (−3.04)
RREL				0.74 (1.43)	0.71 (1.35)	0.72 (1.37)	0.69 (1.38)	0.75 (1.47)	0.76 (1.48)
CAY				2.71 (1.26)	2.87 (1.33)	2.88 (1.31)			
Adj.R^2(%)	15.37	14.16	13.60	23.55	22.81	21.98	23.83	23.65	22.85

Note(s): The figures inside the parentheses underneath the parameter estimates were the *t*-statistics, and ·, *, ** and *** indicate that the coefficients of the regression were significantly different from zero at the 10%, 5%, 1% and 0.1% level, respectively

8. Conclusion

We have proposed a consistent estimation approach for the IV and the VRP, which can remove the adverse effects of the pricing errors in the option prices, by first conducting nonparametric regression estimation of the option price function. This is an improvement over the existing estimators for both IV and VRP proposed by Jiang and Tian (2005) and CBOE (2003). Our theoretical study and numerical simulations show that in the presence of option pricing errors, Jiang and Tian (2005) and CBOE's IV estimators and subsequently the VRP estimators are adversely affected by the observational errors in the option prices. We also present an asymptotic analysis on the proposed VRP estimation under three asymptotic regimes regarding the relative order of sample sizes in the option data for IV estimation to the historic return data for estimating the RV.

In the empirical study of the S&P 500 data, we found that the IV was a more significant predictor for the subsequent RV than the Black-Scholes IV and the lagged RV, which was consistent with the findings in Jiang and Tian (2005). Besides, the proposed consistent estimators for the IV and VRP provide more accurate fitting and forecasting performance than the other IV and VRP estimates, as well as the BS IV and the lagged RV. In the forecasting of S&P 500 returns, we verified that the VRP as a predictor is more efficient than other commonly used predictors, such as the price-earnings ratio and the stochastically detrend risk-free rate. And our proposed estimation for the VRP contains more information of the stock market returns, as compared to the existing model-free approaches.

References

Ait-Sahalia, Y. and Lo, A.W. (1998), "Nonparametric estimation of state-price densities implicit in financial asset prices", *Journal of Finance*, Vol. 53 No. 2, pp. 499-547.

Andersen, T.G., Bollerslev, T., Diebold, F.X. and Labys, P. (2001), "The distribution of realized exchange rate volatility", *Journal of Financial Economics*, Vol. 61, pp. 43-76.

Bakshi, G. and Kapadia, N. (2003), "Delta-hedged gains and the negative market volatility risk premium", *Review of Financial Studies*, Vol. 16 No. 2, pp. 527-566.

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

Bollerslev, T., Tauchen, G. and Zhou, H. (2009), "Expected stock returns and variance risk premia", *Review of Financial Studies*, Vol. 22 No. 11, pp. 4463-4492.

Breeden, D.T. and Litzenberger, R.H. (1978), "Prices of state-contingent claims implicit in option prices", *Journal of Business*, Vol. 51 No. 4, pp. 621-651.

Britten-Jones, M. and Neuberger, A. (2000), "Option prices, implied price processes, and stochastic volatility", *Journal of Finance*, Vol. 2, pp. 839-866.

Canina, L. and Figlewski, S. (1993), "The informational content of implied volatility", *Review of Financial Studies*, Vol. 6 No. 3, pp. 659-681.

Carr, P. and Wu, L. (2009), "Variance risk premiums", *Review of Financial Studies*, Vol. 22 No. 3, pp. 1311-1341.

Chen, S.X. (1996), "Empirical likelihood confidence intervals for nonparametric density estimation", *Biometrika*, Vol. 83, pp. 329-341.

Chen, S.X. and Xu, Z. (2014), "On implied volatility for options, some reasons to smile and more to correct", *Journal of Econometrics*, Vol. 179 No. 1, pp. 1-15.

Chicago Board Options Exchange. (2003), "VIX – CBOE volatility index", available at: http://www.cboe.com/micro/vix/vixwhite.pdf.

Christensen, B.J. and Prabhala, N.R. (1998), "The relation between implied and realized volatility", *Journal of Financial Economics*, Vol. 50, pp. 125-150.

Demeterfi, K., Derman, E., Kamal, M. and Zou, J. (1999), *More than You Ever Wanted to Know about Volatility Swaps*, Quantitative Strategies Research Notes, Goldman Sachs.

Dumas, B., Fleming, J. and Whaley, R.E. (2002), "Implied volatility functions: empirical tests", *Journal of Finance*, Vol. 53 No. 6, pp. 2059-2106.

Fan, J. (1992), "Design-adaptive nonparametric regression", *Journal of the American Statistical Association*, Vol. 87, pp. 998-1004.

Fan, J. and Yao, Q. (1998), "Efficient estimation of conditional variance functions in stochastic regression", *Biometrika*, Vol. 85 No. 3, pp. 645-660.

Hansen, P.R. and Lunde, A. (2006), "Realized variance and market microstructure noise", *Journal of Business and Economic Statistics*, Vol. 24 No. 2, pp. 127-161.

Hardle, W. (1990), *Applied Nonparametric Regression: References*, Cambridge University Press, Cambridge.

Hentschel, L. (2003), "Errors in implied volatility estimation", *Journal of Financial and Quantitative Analysis,* Vol. 38, pp. 779-810.

Heston, S.L. (1993), "A closed-form solution for options with stochastic volatility with applications to bond and currency options", *Review of Financial Studies*, Vol. 6 No. 2, pp. 327-343.

Jiang, G.J. and Tian, Y.S. (2005), "The model-free implied volatility and its information content", *Review of Financial Studies*, Vol. 18 No. 4, pp. 1305-1342.

Jiang, G.J. and Tian, Y.S. (2007), "Extracting model-free volatility from option prices: an examination of the vix index", Social Science Electronic Publishing, Vol. 14 No. 3, pp. 35-60.

Lettau, M. and Ludvigson, S. (2001), "Consumption, aggregate wealth, and expected stock returns", *Journal of Finance*, Vol. 56 No. 3, pp. 815-849.

Merton, R.C. (1973), "Theory of rational option pricing", *Bell Journal of Economics*, Vol. 4 No. 1, pp. 141-183.

Nadaraya, E.A. (1964), "On estimating regression", *Theory of Probability and Its Applications*, Vol. 9 No. 1, pp. 157-159.

Silverman, B. (1986), *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London.

Watson, G.S. (1964), "Smooth regression analysis", *The Indian Journal of Statistics, Series A*, Vol. 26 No. 4, Sankhya, pp. 359-372.

Whaley, R. (1993), "Derivatives on market volatility: hedging tools long overdue", *Journal of Derivatives*, Vol. 1 No. 1, pp. 71-84.

Zhang, L., Mykland, P.A. and Ait-Sahalia, Y. (2005), "A tale of two time scales: determining integrated volatility with noisy high-frequency data", *Journal of the American Statistical Association*, Vol. 199 No. 472, pp. 1394-1411.

Zhou, S., Shen, X. and Wolfe, D.A. (1998), "Local asymptotics for regression splines and confidence regions", *The Annals of Statistics*, Vol. 26 No. 5, pp. 1760-1782.

Corresponding author
Song Xi Chen can be contacted at: csx@gsm.pku.edu.cn