International Journal of Theoretical and Applied Finance
Vol. 14, No. 4 (2011) 433–463

DOI: 10.1142/S0219024911006681

MODEL-FREE IMPLIED VOLATILITY: FROM SURFACE TO INDEX

M. FUKASAWA[*,¶], I. ISHIDA[*,‖], N. MAGHREBI[†,**], K. OYA[‡,††], M. UBUKATA[§,‡‡] and K. YAMAZAKI[*,§§]

[*]*Center for the Study of Finance and Insurance*
Osaka University, 1-3 Machikaneyama
Toyonaka, Osaka 560-8531, Japan

[†]*Graduate School of Economics*
Wakayama University, 930 Sakaedani
Wakayama 640-8510, Japan

[‡]*Graduate School of Economics*
Osaka University, 1-7, Machikaneyama
Toyonaka, Osaka 560-0043, Japan

[§]*Department of Economics*
Kushiro Public University of Economics
4-1-1 Ashino, Kushiro, Hokkaido 085-8585, Japan

[¶]*fukasawa@sigmath.es.osaka-u.ac.jp*
[‖]*i-ishida@sigmath.es.osaka-u.ac.jp*
[**]*nebilmg@eco.wakayama-u.ac.jp*
[††]*oya@econ.osaka-u.ac.jp*
[‡‡]*ubukata@kushiro-pu.ac.jp*
[§§]*k-yamazaki@sigmath.es.osaka-u.ac.jp*

Received 5 October 2010
Accepted 14 November 2010

We propose a new method for approximating the expected quadratic variation of an asset based on its option prices. The quadratic variation of an asset price is often regarded as a measure of its volatility, and its expected value under pricing measure can be understood as the market's expectation of future volatility. We utilize the relation between the asset variance and the Black-Scholes implied volatility surface, and discuss the merits of this new model-free approach compared to the CBOE procedure underlying the VIX index. The interpolation scheme for the volatility surface we introduce is designed to be consistent with arbitrage bounds. We show numerically under the Heston stochastic volatility model that this approach significantly reduces the approximation errors, and we further provide empirical evidence from the Nikkei 225 options that the new implied volatility index is more accurate in predicting future volatility.

Keywords: Model-free implied volatility index; volatility forecasting; volatility surface; variance swaps.

1. Introduction

Market volatility is one of the most important ingredients in financial decision-making. Yet it is not a quantity directly observable in the market. One measure of market volatility is the options-implied expectation of the short-term quadratic variation of the market index. The VIX index disseminated by the Chicago Board Options Exchange (CBOE) is one such measure calculated from the prices of the S&P 500 index options. This volatility index is *model-free* in the sense that it does not rely on a particular option pricing model such as the Black-Scholes model. Indeed, it approximates the risk-neutral expectations of the quadratic variations of the market index over a fixed expiration period based on their model-free link with the prices of the corresponding index options. The theory behind this link assumes, among other things, the existence of a continuum of options over an infinite range of strike prices. Thus, approximation errors would necessarily ensue in any attempt to quantify market volatility from the limited number of observable option prices.

In order to reduce approximation errors, this paper proposes a new method based on a formula induced by the model-free link introduced in the context of pricing variance swaps. Given a continuous semimartingale S standing for an asset price process, we have

$$\frac{1}{T}\mathbb{E}[\langle \log(S)\rangle_T] = \int \sigma(g(z))^2 \phi(z) \mathrm{d}z. \tag{1.1}$$

Here, σ expresses the Black-Scholes implied volatility of the asset S with maturity T as a function of log-moneyness $k = \log(K/F)$ with strike price K and forward price F, g is the inverse function of the mapping

$$k \mapsto d_2(k) := d_2(k, \sigma(k)) \tag{1.2}$$

where

$$d_2(k, \sigma) := -\frac{k}{\sigma\sqrt{T}} - \frac{\sigma\sqrt{T}}{2}, \tag{1.3}$$

and ϕ is the standard normal density. It should be noted that the use of the Black-Scholes implied volatility function σ does not imply the dependency of the new method on this particular option theory. The Black-Scholes implied volatility function is a *nonlinear transformation* of the option prices that reduces to a constant only in a special case where the underlying asset follows a log-normal distribution. The formula (1.1) states that an average of the Black-Scholes implied variance σ^2 is equal to an average of the quadratic variations. This issue will be discussed in further detail in Sec. 2.

The form of the integral (1.1) written with respect to the standard normal density ϕ, together with polynomial interpolation of the integrand, enables us to avoid numerical integration. Our approximation method therefore provides better accuracy than those which rely on discretization to compute integrals such as the CBOE method and the one proposed by Jiang and Tian [15]. We also propose a new C^1-interpolation method by cubic polynomials in the Black-Scholes implied volatility

scale to overcome the errors typically incurred by the usual C^2-interpolation in terms of cubic splines.

We develop a new volatility index using our approximation method and evaluate its efficiency numerically based on the Nikkei 225 index options traded on the Osaka Securities Exchange (OSE). Before using the actual options price data, we first consider artificial data generated under the Heston stochastic volatility model (Heston [14]). The result shows that the new index values are significantly closer to the true values than those produced by the CBOE procedure. We then use the actual options data and evaluate empirically its ability to forecast market volatility, using as many as 15 realized variance estimators of the Nikkei 225 stock average. Our new index is associated with higher R^2's than its CBOE counterpart for all cases, implying that it forecasts market volatility more accurately. These numerical results provide further evidence on the effectiveness of our approximation method and the usefulness of the new volatility index.

The remainder of the paper is organized as follows. We first review in Sec. 2 the theory underlying the development of the model-free implied volatility, and describe in Sec. 3 the new algorithm for computing the volatility index. We then discuss the advantages of using our approximation method over the CBOE procedure in Sec. 4, along with numerical results within the setting of the Heston model. Section 5 provides some empirical evidence based on the market prices of the Nikkei 225 options, and shows that the new volatility index has the potential to predict future volatility more accurately than its comparative index based on the CBOE procedure. Section 6 concludes the paper. For readers' convenience, Appendix A gives the proofs of the theorems in Sec. 2 that are extracted from Fukasawa [11]. Appendix B provides a detailed description of the various estimators of realized variance used in Sec. 5.

2. Model-Free Formulas and their Approximations

This section reviews the model-free formulas for the quadratic variation of the underlying asset and the approximation methods. Let S and S^0 stand for risky and risk-free asset price processes, respectively. If S is a continuous semimartingale and S^0 is a deterministic process of locally bounded variation, then we have, by Itô's formula,

$$\log\left(\frac{S_T}{S_0}\right) = \int_0^T \frac{\mathrm{d}S_t}{S_t} - \frac{1}{2}\int_0^T \frac{\mathrm{d}\langle S\rangle_t}{S_t^2} = \int_0^T \frac{\mathrm{d}S_t^*}{S_t^*} + \log\left(\frac{S_T^0}{S_0^0}\right) - \frac{1}{2}\langle \log(S)\rangle_T$$

where $S_t^* := S_t/S_t^0$. This implies, under any risk-neutral expectation $\mathbb{E}$, that

$$\begin{aligned}\mathbb{E}\left[\langle \log(S)\rangle_T\right] &= -2\mathbb{E}\left[\log\left(\frac{S_T}{S_0}\right)\right] + 2\mathbb{E}\left[\log\left(\frac{S_T^0}{S_0^0}\right)\right] \\ &= -2\mathbb{E}\left[\log\left(\frac{S_T}{\mathbb{E}[S_T]}\right)\right]. \qquad (2.1)\end{aligned}$$

The expected quadratic variation is therefore determined by the distribution of $S_T/\mathbb{E}[S_T]$. It should be remarked here that this relation between the expected quadratic variation and the expected log price was extended to the time-changed Lévy model by Carr *et al.* [8], where the multiplier in the above equality changes from -2 in general.

Now, denote by F the T-expiry forward price of the asset S, and by $C(K)$ and $P(K)$, respectively, the undiscounted call and put option prices of the same asset with strike K and maturity T. As usual, we suppose that there exists a probability measure $\mathbb{E}$ such that

$$F = \mathbb{E}[S_T], \quad C(K) = \mathbb{E}[(S_T - K)_+], \quad P(K) = \mathbb{E}[(K - S_T)_+]$$

for all $K > 0$. In addition, we suppose that S_T admits a density under $\mathbb{E}$ and that there exists $p > 0$ such that

$$\mathbb{E}[S_T^{-p}] < \infty.$$

Then, the density apparently coincides with $C''(K) = P''(K)$, and after integration-by-parts we have for any $K_0 > 0$ that

$$-2\mathbb{E}[\log(S_T/F)] = 2\int_0^{K_0} \frac{P(K)}{K^2}\mathrm{d}K + 2\int_{K_0}^{\infty} \frac{C(K)}{K^2}\mathrm{d}K + 2\int_{K_0}^{F} \frac{K-F}{K^2}\mathrm{d}K. \quad (2.2)$$

Combining this and (2.1), we obtain a model-free relation between the expected quadratic variations and the vanilla option prices, which enables us to calculate a model-free measure of the volatility implicit in the market prices of the options contracts. In practice, however, the number of available option prices is finite and we need to approximate the integrals on the right-hand side of (2.2). The CBOE procedure, which underlies the computation of the current VIX index, refers to the approximation formula:

$$\begin{aligned} -2\mathbb{E}[\log(S_T/F)] \approx{}& 2\sum_{K=K_{\min}}^{K_0-1} \frac{P(K)}{K^2}\Delta K + \frac{P(K_0)+C(K_0)}{K_0^2}\Delta K_0 \\ &+ 2\sum_{K=K_0+1}^{K_{\max}} \frac{C(K)}{K^2}\Delta K - \left(\frac{F}{K_0} - 1\right)^2. \end{aligned} \quad (2.3)$$

Here, $K_{\min}$ and $K_{\max}$ represent, respectively, the lowest and highest strikes among the listed options. The strike price K_0 is taken to be closest to the forward price F among the set of strike prices smaller than or equal to F. This approximation method will be revisited in Sec. 4, in comparison to our own approximation method described in the next section. For a comprehensive account of the CBOE procedure, reference can be made to, for example, Carr and Wu [9].

The formula (1.1) is given in Gatheral [12] as a reformulation of (2.2). See also Carr and Lee [7]. The essentially same formula can be found in Morokoff *et al.* [17] and Chriss and Morokoff [10]. Here we state a result from Fukasawa [11] where the formula is proved under less restrictive conditions. We give the proof in Appendix A.

Theorem 2.1. *Define* $P_{\mathrm{BS}} : \mathbb{R} \times (0,\infty) \to (0,\infty)$ *by*

$$P_{\mathrm{BS}}(k,\sigma) := Fe^k\Phi(-d_2(k,\sigma)) - F\Phi(-d_1(k,\sigma)),$$

where Φ *is the standard normal distribution function,* $d_2(\cdot,\cdot)$ *is as defined in* (1.3)*, and* $d_1(k,\sigma) := d_2(k,\sigma) + \sigma\sqrt{T}$*. Then the Black-Scholes implied volatility* $\sigma : \mathbb{R} \to [0,\infty)$ *is well-defined by*

$$\sigma(k) := P_{\mathrm{BS}}(k,\cdot)^{-1}(P(Fe^k)), \tag{2.4}$$

or, equivalently,

$$P_{\mathrm{BS}}(k,\sigma(k)) = P(Fe^k).$$

Moreover, the mapping $d_2 : k \mapsto d_2(k,\sigma(k))$ *is a decreasing function and it holds that*

$$-\frac{2}{T}\mathbb{E}[\log(S_T/F)] = \int_{-\infty}^{\infty} \sigma(g(z))^2\phi(z)\mathrm{d}z,$$

where g *is the inverse function of* d_2*.*

As we shall shortly see in the next section, we successfully avoid numerical integrations by taking advantage of the form of the integral with respect to ϕ and the polynomial interpolation of the integrand. As can be easily seen in the analysis in Appendix A, the condition that S_T admits a density also implies that σ is continuously differentiable and its derivative is absolutely continuous. It is therefore natural to take a C^1-interpolation/extrapolation by a piecewise polynomial. In addition, we have the following result that should be taken into consideration. For its proof, see Appendix A.

Theorem 2.2. *For* $z, z_0 \in \mathbb{R}$*, put*

$$\hat{\sigma}(z) := \sigma(g(z)),$$

$$\alpha_{\pm}(z;z_0) := -z \pm \sqrt{\hat{\sigma}(z_0)^2 + 2z_0\hat{\sigma}(z_0) + z^2}.$$

Then we have the following.

(i) *For every* $z \in \mathbb{R}$*, we have*

$$\frac{\mathrm{d}\hat{\sigma}}{\mathrm{d}z}(z) > -1.$$

(ii) *For every* $z, z_0 \in \mathbb{R}$ *with* $z > z_0 \geq 0$*, we have*

$$\hat{\sigma}(z) > \alpha_+(z,z_0) > (\hat{\sigma}(z_0) + z_0 - z)_+.$$

(iii) *For every* $z, z_0 \in \mathbb{R}$ *with* $z_0 > z \geq 0$*, we have*

$$\hat{\sigma}(z) < \alpha_+(z,z_0) < \hat{\sigma}(z_0) + z_0 - z.$$

(iv) *For every* $z, z_0 \in \mathbb{R}$ *with* $z < z_0 \leq 0$, *we have*

$$\hat{\sigma}(z) > \alpha_-(z, z_0).$$

(v) *There exists* $z^* < 0$ *such that* $\hat{\sigma}(z^*) = -z^*$ *and it holds for every* $z, z_0 \in \mathbb{R}$ *with* $z < z_0 \leq z^*$ *that*

$$-z > \hat{\sigma}(z) > \alpha_-(z, z_0) > 0.$$

The principal message from this theorem is the following:

(i) The interpolation scheme should not produce excessive oscillations.
(ii) The extrapolation scheme should not induce a rapid decay of $\hat{\sigma}(z)$ as $|z| \to \infty$. In fact Theorem 2.2 implies that $1/\hat{\sigma}(z) = O(|z|)$ as $|z| \to \infty$.

In the next section, we propose an algorithm that meets the above requirements. We further address its advantages over the CBOE procedure in Sec. 4.

3. Algorithm

This section describes our algorithm for approximating accurately the annualized expected quadratic variation

$$\frac{1}{T}\mathbb{E}\left[\langle \log(S)\rangle_T\right] = -\frac{2}{T}\mathbb{E}[\log(S_T/F)] = \int \sigma(g(z))^2\phi(z)\mathrm{d}z \tag{3.1}$$

using a set of option prices of the same maturity. It is applied to several options data sets with different maturities, and the results are then combined to attain the volatility index. In a nutshell, the algorithm consists of three steps: (1) selecting a set of *valid* options to be used, (2) approximating the function $z \mapsto \sigma(g(z))^2$ in terms of a set of cubic polynomials and (3) integrating the obtained function with respect to the normal density ϕ. The rest of this section is devoted to describing each step in detail.

3.1. *STEP 1: Selecting the options to be used in the index calculation*

As a preliminary step, a subset of options are selected for the index calculation. We first identify the *at-the-money (ATM)* strike price K_0, and subsequently the *out-of-the-money (OTM)* calls and puts. The forward price is determined through the theoretical put-call parity relationship using ATM put and call options. Both transaction and bid/ask data are used in this step.

The first task of this step is to determine K_0, which corresponds to the strike price at which the difference between put and call option prices is minimized. Here, we use only transactions data. Notice, however, that transactions prices are used only for the purpose of determining K_0, thereby contributing only indirectly to the index calculation. In case of equal put-call differences between two sets of options, the highest strike is selected. The forward price is then calculated using the

formula

$$F = K_0 + e^{rT}(\text{call price} - \text{put price})$$

where the call and put option prices are ATM transaction prices with strike K_0 and we use the 3-months Certificates of Deposit (CD) rates as proxy for the interest rate r.

Using K_0 as the ATM strike price, the sets of OTM puts and calls are then identified. OTM puts are selected among all put options with strike prices *less than or equal to* K_0, whereas OTM calls are selected among all call options with strike prices *greater than* K_0. We consider only options for which bid/ask prices are well-defined. This means that an option is bound to be eliminated unless both bid and ask prices are available. We further sift the data by discarding any option with a ratio of ask to bid prices *greater than or equal to* some constant c so as to focus on the set of reliable mid quotes that will be used later. Here we use $c = 2$. It should be mentioned that the options at K_0 themselves may fail to satisfy these conditions and may not be used for the index calculation.

Table 1 illustrates how the ATM strike price and OTM puts and calls are selected. The strike price 10,000 (in JPY) is associated with the smallest difference 105 (using transaction data) and it hence becomes the ATM strike K_0. The puts and calls are then selected accordingly using the bid-ask data. Here the put options with strike 6,500 and 7,500 and the call options with 12,500 and 12,750 are eliminated by the ratio rule described above. Suppose $r = 0.004825$, which represents the interest rate over the time to maturity $T = 0.11984398782344$, expressed in years, then the forward price F amounts to

$$10000 + \exp(0.004825 \times 0.11984398782344) \times (400 - 295) = 10105.0607335181.$$

3.2. *STEP 2: Calculating volatility*

Using the options data selected in the previous step, we approximate the integrand of the right-hand side of (3.1),

$$z \mapsto \sigma(g(z))^2. \tag{3.2}$$

This is attained in three stages: (i) obtaining from the available options a sequence of data points that will act as "knots" for the regression based on (3.2), (ii) constructing, using these data points, a function that approximates (3.2), and finally (iii) integrating it with respect to the normal density, and obtaining the implied variance using the identity (3.1). Throughout this step, the bid/ask option prices are used.

STEP 2-1: Converting the data

Each options data consists of the option price and strike price $K = Fe^k$, and this is sufficient to compute the corresponding implied volatility $\sigma(k)$ as defined in (2.4).

$$\sigma(k) := P_{\mathrm{BS}}(k, \cdot)^{-1}(P(Fe^k)),$$

Table 1. An example illustrating how the ATM strike and OTM puts and calls are selected.

Strike	Put			Call			Difference	Type	Midprice
	Trans.	Ask	Bid	Trans.	Ask	Bid			
5000	—	1	—	—	5120	5090	—	—	—
5500	—	1	—	—	4620	4590	—	—	—
6000	1	1	—	—	4120	4090	—	—	—
6500	2	2	1	—	3620	3590	—	—	—
7000	3	4	3	—	3120	3100	—	Put	3.5
7500	9	9	4	—	2630	2600	—	—	—
8000	17	17	16	2150	2140	2110	2133	Put	16.5
8250	25	25	20	—	1890	1870	—	Put	22.5
8500	30	35	30	—	1650	1630	—	Put	32.5
8750	45	50	45	—	1420	1390	—	Put	47.5
9000	65	70	65	—	1190	1170	—	Put	67.5
9250	100	105	95	950	970	950	850	Put	100
9500	150	150	145	745	765	745	595	Put	147.5
9750	215	215	205	565	575	560	350	Put	210
10000	**295**	**300**	**295**	**400**	**410**	**400**	**105**	**Put**	**297.5**
10250	405	420	405	280	280	265	125	Call	272.5
10500	550	570	550	165	175	165	385	Call	170
10750	730	750	730	100	105	100	630	Call	102.5
11000	—	955	940	55	60	55	—	Call	57.5
11250	1150	1180	1160	30	35	30	1120	Call	32.5
11500	—	1420	1390	18	19	17	—	Call	18
11750	1650	1660	1640	9	10	9	1641	Call	9.5
12000	—	1910	1880	6	6	5	—	Call	5.5
12250	—	2150	2130	3	4	3	—	Call	3.5
12500	—	2400	2380	2	2	1	—	—	—
12750	—	2650	2620	1	2	1	—	—	—
13000	—	2900	2870	—	1	—	—	—	—
13500	—	3400	3370	—	1	—	—	—	—
14000	—	3900	3880	—	1	—	—	—	—
14500	—	4400	4380	—	1	—	—	—	—

A data point $(d_2(k), \sigma(k)^2)$ is thereby obtained for each option. More precisely, the forward price F is obtained in STEP 1, and then the bisection method with error bound $1.0e-9$ is used to compute implied volatilities.

As is shown in Theorem 2.1, $d_2(k)$ is decreasing in k, thereby decreasing in K as well. In order to satisfy this constraint, we disregard the data points on which this monotonicity fails, and focus on the (longest possible) interval of data points at which the monotonicity holds. Starting from the put option with the highest strike (regularly K_0), each option is visited backward in K to make sure that $d_2(k)$ is decreasing. Once monotonicity breaks down, the corresponding put option and any put option with lower strikes are discarded. The same procedure is applied to call options. Starting from the call option with the lowest strike, each option is visited forward in K; once monotonicity breaks down, the associated call option and any call option with higher strikes are similarly discarded. The first five columns in Table 2 show the option prices selected from Table 1 and their corresponding d_2's

Table 2. The values of d_2 and a, b, c, d representing the coefficients of the polynomials a_j, b_j, c_j, d_j based on the selected option prices listed in Table 1.

Strike	Type	Price	d_2	a (impl. vol.)	b	c	d
7000	P	3.5	2.322589	0.1953966	0	0	0
8000	P	16.5	1.737578	0.1401579	0.1024657	0.1339089	−0.2523994
8250	P	22.5	1.597871	0.1247173	0.0900612	0.3505619	−1.4609950
8500	P	32.5	1.428667	0.1129279	0.0628586	−0.0399028	0.4739328
8750	P	47.5	1.243389	0.1025435	0.0574971	−0.0524102	0.2406433
9000	P	67.5	1.054255	0.0913947	0.0472180	0.1316943	−0.3684178
9250	P	100.0	0.833485	0.0835569	0.0318685	−0.0201518	0.1658297
9500	P	147.5	0.595460	0.0768361	0.0298054	−0.0284511	0.0918246
9750	P	210.0	0.347682	0.0690620	0.0273430	0.0388834	−0.0912490
10000	P	297.5	0.077152	0.0627555	0.0188023	0.0184341	−0.0065281
10250	C	272.5	−0.211813	0.0586251	0.0146191	−0.0178526	0.0578870
10500	C	170.0	−0.516513	0.0540715	0.0102862	0.0316420	−0.0536746
10750	C	102.5	−0.820640	0.0523597	0.0056111	−0.0151997	0.0501673
11000	C	57.5	−1.128248	0.0506391	0.0020201	0.0231773	−0.0375809
11250	C	32.5	−1.410956	0.0510783	−0.0023874	−0.0067401	0.0342762
11500	C	18.0	−1.678436	0.0519399	−0.0026407	−0.0074597	0.0197729
11750	C	9.5	−1.941339	0.0524815	−0.0067655	0.0380046	−0.0764793
12000	C	5.5	−2.158142	0.0549685	−0.0168207	0.0276429	−0.0136939
12250	C	3.5	−2.333800	0.0588631	0	−0.2828918	0.8919309

and implied volatilities. Here, d_2 is found to be monotonically decreasing in the strike price, and therefore no truncation is applied.

STEP 2-2: Constructing a function approximating (3.2) $z \mapsto \sigma(g(z))^2$.

After STEP 2-1, we suppose there are M data points. We denote by $x_1 \leq \cdots \leq x_M$ the ascending sequence of $d_2(\cdot)$ and by $y_1 \leq \cdots \leq y_M$ the corresponding sequence of implied volatilities $\sigma(\cdot)^2$. In the example in Table 2, $M = 19$, $x_1 = -2.3338, \ldots, x_M = 2.322589$ and $y_1 = 0.05886305, \ldots, y_M = 0.1953966$. For each of the following intervals

$$(-\infty, x_1], [x_1, x_2], [x_1, x_2], \ldots, [x_{M-1}, x_M] \text{ and } [x_M, \infty),$$

a cubic polynomial is defined so that their union is continuous and differentiable on $\mathbb{R}$ and goes through each data point.

In defining the cubic polynomial for a given interval, say $[x_j, x_{j+1}]$ for some $j = 1, \ldots, M-1$, there are two more "degrees of freedom" remaining after satisfying the requirements that it goes through (x_j, y_j) and (x_{j+1}, y_{j+1}). It should be noted that there are various ways of using these degrees of freedom. A typical example uses *cubic splines* as in Jiang and Tian [14], which realizes the twice-differentiability at these joint points. We choose, however, to obtain polynomials in the way we shall discuss hereafter. For $(-\infty, x_1]$ and $[x_M, \infty)$, we use constant extrapolation. In the next section, we will discuss this method in comparison to the one by Jiang and Tian [14].

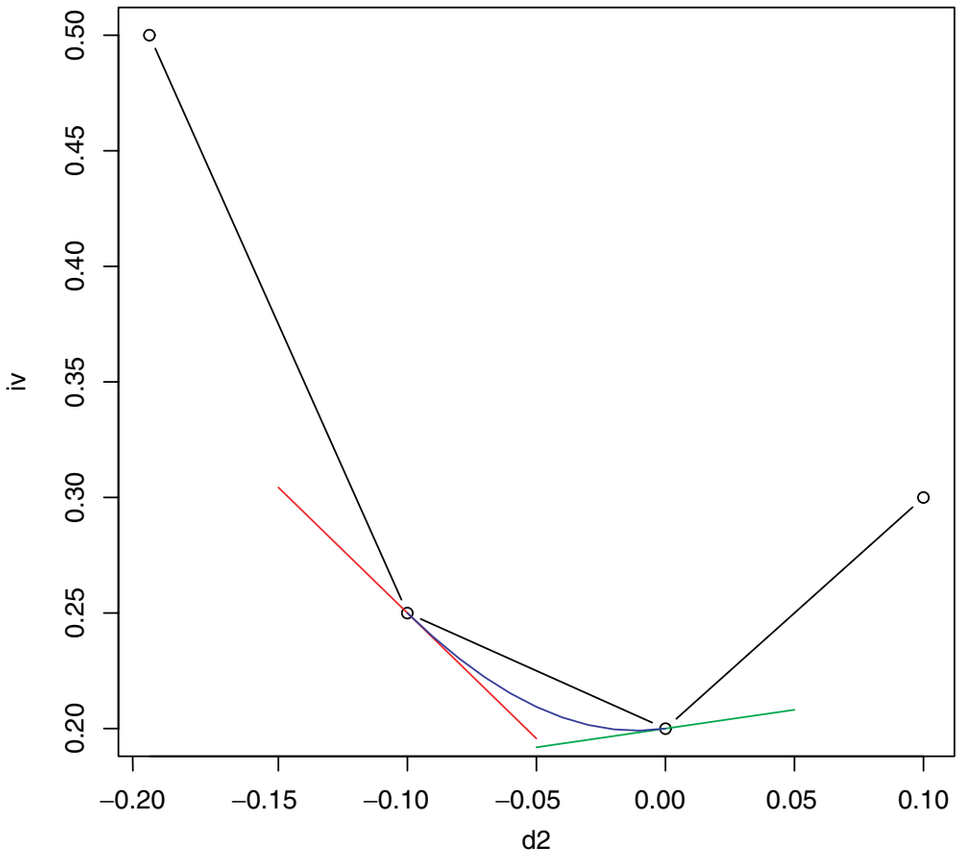

Fig. 1. The incidence angle becomes equal to the reflected angle. The y-axis refers to IV, which represents Implied Variance in this case.

入射角 = 反射角

The slope of the cubic polynomial at each joint point $\{(x_j, y_j); j = 1, \ldots, M\}$ is first determined. We initially set $y'(x_1) = y'(x_M) = 0$. For $2 \leq j \leq M-1$, the slope is chosen so that "the incidence angle becomes equal to the reflected angle". In other words, we construct a polynomial so that the angle made by its tangent line at (x_j, y_j) and the line going through (x_{j-1}, y_{j-1}) and (x_j, y_j) is equal to the angle made by the same tangent line and the line going through (x_j, y_j) and (x_{j+1}, y_{j+1}), as shown in Fig. 1.

This can be calculated as in the following:

slope

$$y'(x_j) = -\left(\frac{x_{j+1} - x_j}{l_{j+1}} - \frac{x_j - x_{j-1}}{l_j}\right) \Bigg/ \left(\frac{y_{j+1} - y_j}{l_{j+1}} - \frac{y_j - y_{j-1}}{l_j}\right)$$

where l_j and l_{j+1} are Euclidean distances from (x_j, y_j) to (x_{j-1}, y_{j-1}) and (x_{j+1}, y_{j+1}), respectively.

For every $1 \leq j \leq M$, the polynomial on $[x_j, x_{j+1}]$ is now uniquely determined by the two end-points (x_j, y_j) and (x_{j+1}, y_{j+1}) and the slopes at these points. We represent the polynomial by

$$y(x) = a_j + b_j(x - x_j) + c_j(x - x_j)^2 + d_j(x - x_j)^3, \quad x_j \leq x \leq x_{j+1}$$

where a_j, b_j, c_j and d_j are the coefficients to be determined. It can be easily shown that

$$\begin{aligned} a_j &= y_j, \\ b_j &= y'(x_j), \\ c_j &= (3\Delta y_j - \Delta x_j y'(x_{j+1}) - 2\Delta x_j y'(x_j))/(\Delta x_j^2), \\ d_j &= (\Delta y_j - y'(x_j)\Delta x_j - c_j \Delta x_j^2)/(\Delta x_j^3), \end{aligned}$$

where $\Delta x_j = x_{j+1} - x_j$ and $\Delta y_j = y_{j+1} - y_j$. For the polynomials on $(-\infty, x_0]$ and $[x_M, \infty)$, we set $y = x_0$ and $y = x_M$, respectively. Figure 2 shows the polynomial based on the example in Table 2.

3.3. *STEP 3: Integrating*

Now we obtain the integral of the function. Notice that

$$\begin{aligned} \int_{-\infty}^{\infty} \sigma(g(z))^2 \phi(z) \mathrm{d}z = \sum_{j=1}^{M-1} \int_{x_j}^{x_{j+1}} \sigma(g(z))^2 \phi(z) \mathrm{d}z + \int_{-\infty}^{x_1} \sigma(g(z))^2 \phi(z) \mathrm{d}z \\ + \int_{x_M}^{\infty} \sigma(g(z))^2 \phi(z) \mathrm{d}z. \end{aligned}$$

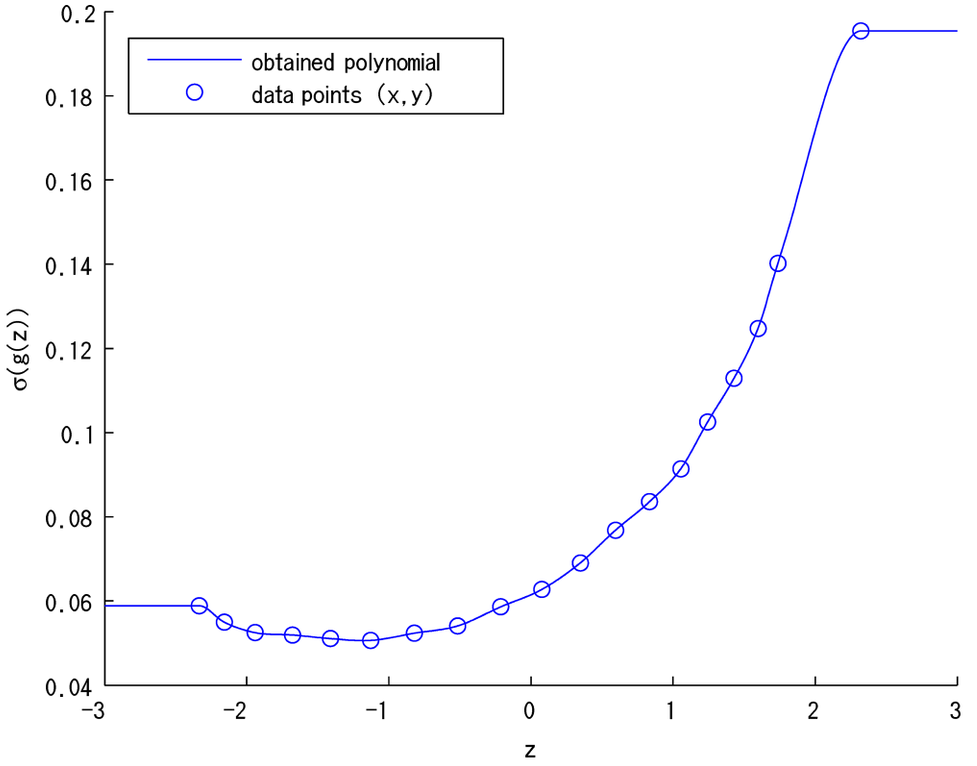

Fig. 2. The cubic polynomial obtained based on the option data in Table 2.

Here, in particular, for every $1 \leq j \leq M-1$, we have

$$\begin{aligned}\int_{x_j}^{x_{j+1}} \sigma(g(z))^2\phi(z)\mathrm{d}z &\approx \int_{x_j}^{x_{j+1}} y(z)\phi(z)\mathrm{d}z \\ &= \int_{x_j}^{x_{j+1}} (a_j + b_j(z-x_j) + c_j(z-x_j)^2 + d_j(z-x_j)^3)\phi(z)\mathrm{d}z \\ &= a_jA_j + b_jB_j + c_jC_j + d_jD_j\end{aligned}$$

where

$$\begin{aligned}A_j &:= \Phi(x_{j+1}) - \Phi(x_j), \\ B_j &:= -(\phi(x_{j+1}) - \phi(x_j)) - x_j(\Phi(x_{j+1}) - \Phi(x_j)), \\ C_j &:= -(x_{j+1}\phi(x_{j+1}) - x_j\phi(x_j)) + 2x_j(\phi(x_{j+1}) - x_j\phi(x_j)) \\ &\quad + (1+x_j^2)(\Phi(x_{j+1}) - \Phi(x_j)), \\ D_j &:= (1-x_{j+1}^2)\phi(x_{j+1}) - (1-x_j^2)\phi(x_j) + 3x_j(x_{j+1}\phi(x_{j+1}) - x_j\phi(x_j)) \\ &\quad - 3(1+x_j^2)(\phi(x_{j+1}) - \phi(x_j)) - x_j(3+x_j^2)(\Phi(x_{j+1}) - \Phi(x_j)).\end{aligned}$$

The implied variance (3.1) can thus be approximated using the above formula.

4. Comparison to the CBOE Procedure

This section discusses the new method in comparison to the CBOE procedure. An apparent disadvantage of the CBOE approximation (2.3) is that it leads the underestimation of the expected quadratic variation due to the reduced contributions of $P(K)$ on $K < K_{\min}$ and $C(K)$ on $K > K_{\max}$. This is exacerbated even further when the range of available strike prices is not sufficiently wide.

Let us illustrate this issue by using the Nikkei 225 options in comparison to the S&P 500 options. On October 9, 2008, the S&P 500 index dropped to 909.92 USD, and it was followed by the Nikkei stock average, which decreased to 8276.43 JPY on the next day. With both indexes dropping around 10% on a single day, the ranges of strikes for the corresponding options are however significantly different. As shown in Table 3, the lowest strike of the Nikkei 225 index option $K_{\min}$ approaches the spot price of the underlying index; it is as large as 88% of the spot price, which is tremendously larger than the ratio for the S&P case. As a result, as can be

Table 3. The lowest strikes of the S&P 500 and the Nikkei 225 index options and the spot prices of their underlying market indexes.

	Spot Price		$K_{\min}$ and Ratio to Spot Price on Oct. 9 (Oct. 10)	
	Oct. 8 (Oct. 9)	Oct. 9 (Oct. 10)	Nov. Maturity	Dec. Maturity
S&P	984.94	909.92	400 (40%)	400 (40%)
Nikkei	9157.49	8276.43	7250 (88%)	7250 (88%)

Table 4. Put option prices of the Nikkei 225 index options with the lowest strikes on Oct. 10, 2008 maturing in Nov. 2008 and Dec. 2008.

Strike	Nov. Maturity			Dec. Maturity		
	Trans.	Ask	Bid	Trans.	Ask	Bid
7250	500	520	500	695	695	655
7500	580	600	580	605	605	570
7750	680	700	680	790	790	760
8000	805	805	790	785	900	855
8250	925	925	905	930	1020	970
⋮	⋮	⋮	⋮	⋮	⋮	⋮

observed in Table 4, the price of the put option at $K_{\min}$ becomes significantly high in comparison to the single-digit prices reported in Table 1 under normal market conditions. In this case, the cutting-off of the integral over $[0, K_{\min}]$ severely affects the approximation because the integrand $P(K)/K^2$ is not sufficiently small around $K = K_{\min}$. Moreover, a sharp decline in asset value tends to be accompanied with an increase in expected volatility, or equivalently a fatter tail of $P(K)$. This further amplifies the effect of the cutting-off. The CBOE procedure (2.3) is therefore likely to lead to a non-negligible negative bias in options markets with unaccommodative listing rules, particularly during financial crises. A more accurate estimation of implied volatility is indeed required for this reason.

Our approach improves this approximation through the process of extrapolation. Jiang and Tian [15] employed the cubic spline for the Black-Scholes implied volatility with linear extrapolation. We remark that their method may cause a positive bias due to its inconsistency with the theoretical asymptotic behavior of the volatility surface described by Lee [16], where he obtained a model-free bound $\sigma(k) < \sqrt{2|k|}$ for large $|k|$. See also Lemma A.4 for a refinement of this bound. In order to be consistent with Theorem 2.2 that gives a counterpart of those model-free bounds for our integration scale, we choose a constant extrapolation as described in STEP 2-2.

Another source of approximation errors in (2.3) is the discretization of the integral with respect to K. A continuous interpolation would be a more natural treatment since, in theory, $P(K) = \mathbb{E}[(K - S_T)_+]$ and $C(K) = \mathbb{E}[(S_T - K)_+]$ are continuous functions. They are C^1 if S_T admits a density. Our algorithm employs a C^1-interpolation by cubic polynomials in the Black-Scholes implied volatility scale. The algorithm by Jiang and Tian [15] consists of the C^2-interpolation and numerical integration with the original scale (2.2). However, the C^2-property of these functions is not guaranteed in general, and adhering to this may cause undesirable oscillations. On the other hand, our C^1-interpolation approach is expected to prevent these oscillations, which is in fact required in the light of Theorem 2.2.

We proceed to integration in the same Black-Scholes implied volatility scale using the formula (1.1). Since the integral is with the standard normal density, we can utilize the well-known identities for the Hermite polynomial system to avoid

numerical integration. The model-free formula (1.1) describes a direct link between the expected quadratic variation and the curvature of the Black-Scholes implied volatility surface that is familiar to practitioners.

To examine the extent to which our method indeed improves the approximation, let us consider the *Heston model*:

$$\begin{aligned} \mathrm{d}S_t &= S_t\sqrt{V_t}[\rho \mathrm{d}W_t^1 + \sqrt{1-\rho^2}\mathrm{d}W_t^2], \\ \mathrm{d}V_t &= \lambda(v - V_t)\mathrm{d}t + \eta\sqrt{V_t}\mathrm{d}W_t^1, \end{aligned} \tag{4.1}$$

where (W^1, W^2) is a two-dimensional standard Brownian motion. This model enables us to calculate easily the theoretical call and put option prices and obtain the following explicit formula for the expected annualized quadratic variation:

$$\frac{1}{T}\mathbb{E}[\langle \log(S)\rangle_T] = \frac{1}{T}\mathbb{E}\left[\int_0^T V_t \mathrm{d}t\right] = v + \frac{1-e^{-\lambda T}}{\lambda T}(V_0 - v). \tag{4.2}$$

See, e.g., Gatheral [12] for more details. Using artificial options data generated by the Heston pricing formula, we evaluate the approximations in comparison to the true value of the expected quadratic variation. We generate call and put option prices with the same set of strikes and nearest two maturities as the above-mentioned options data on October 10, 2008. The first and second maturities falling in November and December correspond to $T = 0.0951864535768645$ and $T = 0.171898782343988$, respectively. The initial price S_0 of the underlying asset on October 10, 2008 is set to 8276.43 JPY and the interest rate is set to 0. See Table 5 for the generated data for the first maturity. To further approach normal market conditions, we randomize the bid-ask spread using geometric random variables with success probability $p = 0.8$. The ask price is sampled so that it becomes the lowest price above the true value (consistent with the tick-size used on the OSE) with probability p, and becomes the second lowest price with probability $(1-p)p$. It becomes the subsequent prices in the same geometric manner, and the bid price is chosen similarly. We use the corresponding mid-quote price for each transaction price.

Table 6 shows the expected annualized quadratic variations and the approximated values by the CBOE procedure and our method based on the following parameter sets:

Parameter Set A: $\lambda = 1$, $v = 0.2$, $\eta = 0.5$, $\rho = -0.8$ and $V_0 = 0.6$;
Parameter Set B: $\lambda = 1$, $v = 0.2$, $\eta = 1.0$, $\rho = -0.4$ and $V_0 = 0.6$;
Parameter Set C: $\lambda = 5$, $v = 0.04$, $\eta = 1.0$, $\rho = -0.4$ and $V_0 = 0.6$;
Parameter Set D: $\lambda = 1.5$, $v = 0.04$, $\eta = 0.3$, $\rho = -0.7$ and $V_0 = 0.04$.

The column "New" stands for the values resulting from the application of our new method. The parameters in D are typical values obtained by calibrating on a regular day with instantaneous volatility $\sqrt{V_0} = 20\%$. Under this average volatility level, the tails of the distribution of S_T are light, and hence the approximation errors due to the cutting-off of tails are negligible. The main source of error

Table 5. Artificial data with maturity November 2008 and parameter set A.

Strike	Call			Put		
	Ask	Theoretical	Bid	Ask	Theoretical	Bid
7250	1370	1361.58	1360	345	335.15	335
7500	1200	1197.41	1190	430	420.98	420
7750	1050	1046.16	1040	525	519.73	515
8000	915	907.99	905	640	631.56	630
8250	790	782.88	780	765	756.45	755
8500	675	670.56	665	910	894.13	890
8750	580	570.59	570	1050	1044.16	1040
9000	485	482.35	475	1210	1205.92	1200
9250	415	405.12	405	1380	1378.69	1370
9500	345	338.08	335	1570	1561.65	1560
9750	290	280.34	280	1760	1753.91	1750
10000	240	231.01	230	1960	1954.58	1950
10250	195	189.19	185	2170	2162.76	2060
10500	160	153.99	140	2380	2377.56	2370
10750	130	124.60	120	2610	2598.17	2590
11000	110	100.22	100	2830	2823.79	2820
11250	90	80.15	80	3060	3053.72	3050
11500	70	63.73	60	3290	3287.30	3280
11750	55	50.40	45	3540	3523.97	3520
12000	40	39.63	30	3770	3763.20	3760
12250	40	31.00	30	4010	4004.57	3900
12500	30	24.12	20	4250	4247.69	4140
12750	20	18.67	18	4500	4492.24	4290
13000	16	14.38	14	4740	4737.95	4730
13250	13	11.02	11	4990	4984.59	4980
13500	10	8.40	8	5240	5231.97	5230
13750	8	6.38	6	5480	5479.95	5370
14000	8	4.82	4	5730	5728.39	5720
14250	5	3.62	3	6000	5977.19	5970
14500	4	2.71	2	6230	6226.28	6120
15000	2	1.50	—	6730	6725.07	6720
15500	1	0.81	—	7240	7224.38	7220
16000	1	0.44	—	7740	7724.01	7720
16500	1	0.23	—	8230	8223.80	8220
17000	1	0.12	—	8730	8723.69	8720
17500	1	0.06	—	9230	9223.63	9120

Table 6. Expected annualized quadratic variations.

Parameter	2008/11			2008/12		
	True	CBOE	New	True	CBOE	New
A	0.5816	0.4639	0.5767	0.5676	0.4041	0.5504
B	0.5816	0.4604	0.5692	0.5676	0.4023	0.5460
C	0.4856	0.3930	0.4633	0.4157	0.3112	0.4023
D	0.0400	0.0420	0.0408	0.0400	0.0378	0.0394

lies in the numerical integration and our method provides better numerical accuracy. The other specifications A, B and C represent the market conditions at the onset of financial crises, with instantaneous volatility $\sqrt{V_0} \approx 77\%$. The results are more striking. The CBOE method severely underestimates the expected quadratic variation where our method is found to significantly improve the estimation accuracy.

5. New Volatility Index and Empirical Evidence

In this section, we evaluate the in-sample forecasting performance of the volatility index from the Nikkei 225 options based on the new method in comparison to that using the CBOE procedure. The options data were obtained from the Nikkei Economic Electronic Database System (NEEDS) FinancialQuest database, and contain call and put option prices quoted at the close of trading days from December 22, 1997 to December 1, 2009 (2,935 trading days). Throughout this section, we denote by CSFI-VXJ^2 our volatility index and by VXJ^2 the index computed based on the CBOE approach.[1] Here, the values of CSFI-VXJ^2 and VXJ^2 are measured as annualized 30-days quadratic variation, while the volatility indexes usually refer to their square roots in percentage points.

Table 7 summarizes the descriptive statistics of the CSFI-VXJ^2 and VXJ^2 volatility indexes. The mean value of CSFI-VXJ^2 is close to that of VXJ^2, but CSFI-VXJ^2 has a slightly larger standard deviation than VXJ^2. The skewness and kurtosis moments indicate that both CSFI-VXJ^2 and VXJ^2 are not close to the normal distribution. The large Ljung-Box statistics up to tenth order LB(10) indicate that the null hypothesis of no autocorrelations is rejected.

For the sake of estimating market variance over the fixed expiration period of 30 days, we employ ex-post realized variances using high-frequency data. These intraday measures of realized volatility are usually found to provide more accurate variance estimates than those derived from lower-frequency data such as daily, weekly or monthly observations. The NEEDS historical tick data include

Table 7. Descriptive Statistics of the CSFI-VXJ^2 and VXJ^2.

	CSFI-VXJ^2	VXJ^2
Mean	0.0816	0.0822
Standard deviation	0.0824	0.0771
Minimum	0.0120	0.0133
Maximum	0.9454	0.8363
Skewness	4.90	4.55
Kurtosis	31.98	27.31
LB(10)	24432	25271

[1]We use these names because the same indexes are computed, updated and made available online to the public under these titles by the Center for the Study of Finance and Insurance, Osaka University. http://www-csfi.sigmath.es.osaka-u.ac.jp/structure/activity/vxj.php.

the intraday Nikkei 225 stock average recorded every minute. The Tokyo Stock Exchange (TSE) where the Nikkei 225 index components are traded is open for 9:00–11:00 (morning session) and 12:30–15:00 (afternoon session). In order to take into consideration the non-trading hours, we employ the adjustment introduced by Hansen and Lunde [13], using the past 441 daily returns and realized variances to estimate the time-varying adjustment parameter.

In order to perform tests that are robust to the microstructure noise, we use as many as 15 estimators of realized variance listed in Table 8. A more detailed description of these estimators is provided in Appendix B. We summarize in Table 9 the descriptive statistics of annualized realized variances computed over fixed periods of 30 days, consistent with the expiry of the hypothetical options underlying the implied volatility indexes. The realized variances have lower means than the CSFI-VXJ^2 and VXJ^2 indexes, although the standard deviations are rather close to these volatility forecasts. The distribution of realized variance is also found to be positively skewed and long-tailed. Judging from LB(10) values, there are signs of autocorrelations. Figure 3 plots the time-series of market variance along with the CSFI-VXJ^2 and VXJ^2 indexes. There are sudden surges associated with significant economic and financial shocks such as the Russian default and Long-Term Capital Management crises in 1998 and the U.S. housing and credit crisis in 2008.

Table 8. List of estimators of realized variance using high-frequency data.

1. Realized variance with returns sampled at the highest frequencies, RV
2. Realized variance with 5-minute returns, $RV_{5\,\mathrm{min}}$
3. Realized variance with 15-minute returns, $RV_{15\,\mathrm{min}}$
4. Optimally-sampled realized variance as proposed in Bandi and Russell [2], $RV^{}_{BR}$
5. The Bartlett kernel estimator in Barndorff-Nielsen *et al.* [4] with a finite sample optimal number of autocovariances proposed by Bandi and Russell [3], RV^{BK}_{BR}
6. The two-scale estimator with an asymptotically optimal number of subsamples proposed by Zhang *et al.* [20], RV^{ZMA}
7. The two-scale estimator with a finite sample optimal number of subsamples proposed by Bandi and Russell [3], RV^{ZMA}_{BR}
8. The bias-corrected two-scale estimator with an asymptotically optimal number of subsamples proposed by Zhang *et al.* [20], RV^{BC-ZMA}
9. The bias-corrected two-scale estimator with a finite sample optimal number of subsamples proposed by Bandi and Russell [3], RV^{BC-ZMA}_{BR}
10. The flat-top Bartlett kernel estimator with an asymptotically optimal number of autocovariances proposed by Barndorff-Nielsen *et al.* [5], RV^{FBK}
11. The flat-top Bartlett kernel estimator with a finite sample optimal number of autocovariances proposed by Bandi and Russell [3], RV^{FBK}_{BR}
12. The flat-top cubic kernel estimator with an asymptotically optimal number of autocovariances proposed by Barndorff-Nielsen *et al.* [5], RV^{FCK}
13. The flat-top cubic kernel estimator with a finite sample optimal number of autocovariances proposed by Bandi and Russell [3], RV^{FCK}_{BR}
14. The flat-top modified Tukey-Hanning kernel estimator with an asymptotically optimal number of autocovariances proposed by Barndorff-Nielsen *et al.* [5], RV^{FMTH}
15. The flat-top modified Tukey-Hanning kernel estimator with a finite sample optimal number of autocovariances proposed by Bandi and Russell [3], RV^{FMTH}_{BR}

Table 9. Descriptive statistics of realized variances.

	Mean	Standard Error	Minimum	Maximum	Skewness	Kurtosis	LB(10)
RV	0.0640	0.0806	0.0078	0.9334	7.10	61.12	25977
$RV_{5\,\mathrm{min}}$	0.0628	0.0819	0.0063	0.9346	7.03	60.09	25914
$RV_{15\,\mathrm{min}}$	0.0624	0.0928	0.0067	1.0650	7.21	61.30	26031
$RV^{}_{BR}$	0.0632	0.0870	0.0061	1.0066	7.37	65.08	25750
RV^{BK}_{BR}	0.0628	0.0896	0.0064	1.0577	7.55	67.60	25685
RV^{ZMA}	0.0633	0.0871	0.0067	1.0265	7.30	63.67	25889
RV^{ZMA}_{BR}	0.0630	0.0907	0.0061	1.0684	7.50	66.92	25689
RV^{BC-ZMA}	0.0643	0.0807	0.0064	0.9305	7.06	61.48	25695
RV^{BC-ZMA}_{BR}	0.0634	0.0851	0.0062	0.9865	7.28	64.26	25663
RV^{FBK}	0.0636	0.0834	0.0062	0.9601	7.15	62.43	25690
RV^{FBK}_{BR}	0.0636	0.0845	0.0062	0.9762	7.23	63.51	25672
RV^{FCK}	0.0636	0.0860	0.0061	0.9946	7.28	64.15	25664
RV^{FCK}_{BR}	0.0635	0.0841	0.0062	0.9716	7.21	63.33	25670
RV^{FMTH}	0.0636	0.0856	0.0062	0.9888	7.23	63.43	25697
RV^{FMTH}_{BR}	0.0636	0.0841	0.0062	0.9694	7.15	62.41	25700

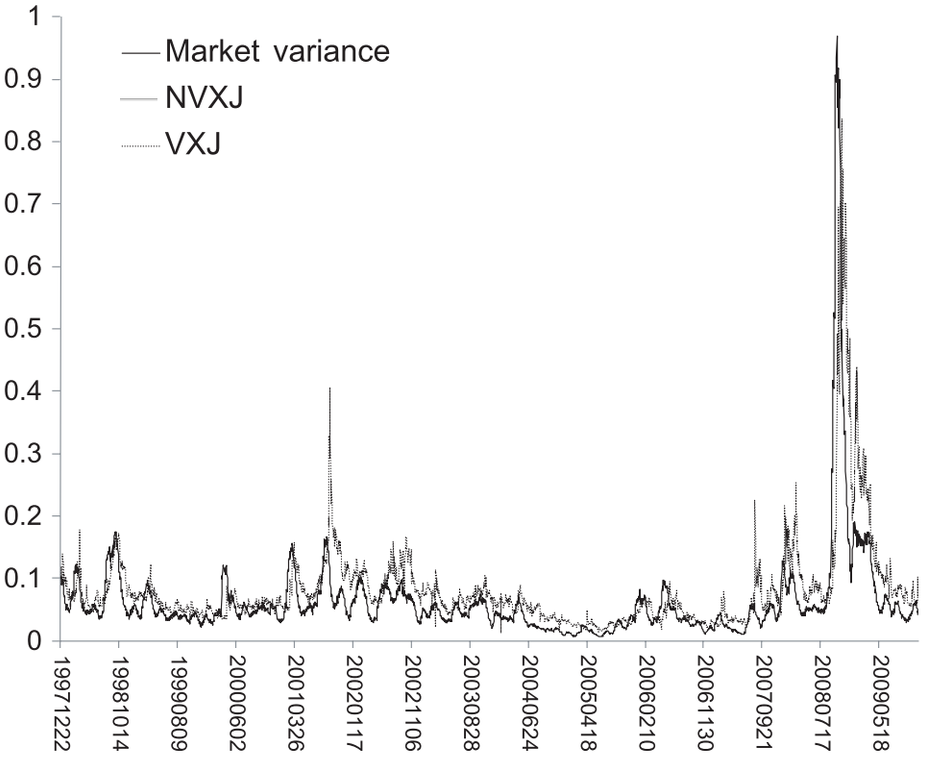

Fig. 3. Time-series of realized variance and implied variance indexes.

A formal evaluation of forecasting performance of the various implied volatility indexes can be made using the following regression model

$$\hat{V}_t = \alpha + \beta\, IV_t + u_t, \tag{5.1}$$

where $\hat{V}$ and IV represent the annualized realized variance and implied variance, respectively. Table 10 reports the coefficient of determination R^2 from the regression

Table 10. Coefficients of determination R^2 from regressions of realized variance on implied variance indexes.

Market Variance	CSFI-VXJ^2	VXJ^2
RV	0.488	0.457
$RV_{5\,\text{min}}$	0.501	0.471
$RV_{15\,\text{min}}$	0.427	0.398
RV_{BR}	0.482	0.451
RV_{BR}^{BK}	0.460	0.430
RV^{ZMA}	0.448	0.417
RV_{BR}^{ZMA}	0.458	0.427
RV^{BC-ZMA}	0.504	0.473
RV_{BR}^{BC-ZMA}	0.491	0.459
RV^{FBK}	0.501	0.469
RV_{BR}^{FBK}	0.496	0.464
RV^{FCK}	0.490	0.458
RV_{BR}^{FCK}	0.497	0.465
RV^{FMTH}	0.490	0.459
RV_{BR}^{FMTH}	0.498	0.466

equation (5.1). Irrespective of the estimator of realized variance used, the R^2 values obtained using CSFI-VXJ^2 are found to be higher than those obtained using VXJ^2. These results show that the CSFI-VXJ^2 forecasts of return variability are more accurate than those based on VXJ^2.

Recall that the indexes are constructed as approximations of the expected quadratic variation. Although the expectation is with respect to a risk-neutral measure, by assuming that it coincides with the physical measure, we further assess the forecast error of CSFI-VXJ^2 and VXJ^2 using four loss functions: the root mean squared error (RMSE), the root mean squared percentage error (RMSPE), the mean absolute error (MAE) and the mean absolute percentage error (MAPE) defined, respectively, as

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(IV - \hat{V})^2}, \quad \text{RMSPE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{IV - \hat{V}}{\hat{V}}\right)^2},$$

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|IV - \hat{V}|, \quad \text{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{IV - \hat{V}}{\hat{V}}\right|.$$

Here we use $N = 2{,}935$ for the number of observations of realized variances and implied variance indexes over the sample period. Table 11 displays the estimates of biases and the loss functions for CSFI-VXJ^2 and VXJ^2. The bias is always positive, but it is smaller in magnitude for CSFI-VXJ^2 than for VXJ^2. The values of loss functions for CSFI-VXJ^2 are uniformly smaller than those for VXJ^2. Judging from these results, the volatility index proposed in this paper provides a better approximation

Table 11. Estimates of bias and four loss functions for the CSFI-VXJ^2 and VXJ^2 indexes.

Market Variance	Bias		RMSE		RMSER		MAE		MAPE	
	CSFI-VXJ^2	VXJ^2	CSFI-VXJ^2	VXJ^2	CSFI-VXJ^2	VXJ^2	CSFI-VXJ^2	VXJ^2	CSFI-VXJ^2	VXJ^2
RV	0.0175	0.0182	0.0657	0.0661	0.6832	0.7395	0.0319	0.0331	0.5171	0.5673
$RV_{5\,\mathrm{min}}$	0.0188	0.0194	0.0655	0.0660	0.7383	0.8015	0.0322	0.0335	0.5548	0.6064
$RV_{15\,\mathrm{min}}$	0.0192	0.0198	0.0760	0.0769	0.8474	0.9086	0.0363	0.0377	0.6510	0.7051
RV_{BR}	0.0184	0.0190	0.0688	0.0697	0.7541	0.8192	0.0329	0.0343	0.5656	0.6188
RV_{BR}^{BK}	0.0187	0.0193	0.0718	0.0727	0.7426	0.8004	0.0336	0.0349	0.5669	0.6172
RV^{ZMA}	0.0183	0.0189	0.0715	0.0722	0.7597	0.8270	0.0340	0.0352	0.5771	0.6309
RV_{BR}^{ZMA}	0.0185	0.0192	0.0724	0.0734	0.7615	0.8226	0.0340	0.0353	0.5812	0.6328
RV^{BC-ZMA}	0.0173	0.0179	0.0645	0.0649	0.7092	0.7785	0.0312	0.0324	0.5241	0.5773
RV_{BR}^{BC-ZMA}	0.0181	0.0188	0.0674	0.0682	0.7273	0.7929	0.0324	0.0337	0.5490	0.6012
RV^{FBK}	0.0180	0.0186	0.0659	0.0666	0.7249	0.7925	0.0320	0.0333	0.5459	0.5985
RV_{BR}^{FBK}	0.0180	0.0186	0.0667	0.0674	0.7273	0.7944	0.0322	0.0335	0.5480	0.6003
RV^{FCK}	0.0180	0.0186	0.0677	0.0686	0.7335	0.7996	0.0325	0.0338	0.5536	0.6059
RV_{BR}^{FCK}	0.0180	0.0186	0.0665	0.0672	0.7260	0.7933	0.0321	0.0334	0.5473	0.5996
RV^{FMTH}	0.0180	0.0186	0.0675	0.0684	0.7372	0.8032	0.0326	0.0339	0.5566	0.6090
RV_{BR}^{FMTH}	0.0180	0.0186	0.0664	0.0671	0.7309	0.7987	0.0323	0.0336	0.5521	0.6047

of short-term market volatility and has a better forecasting performance than the index computed using the CBOE procedure.

6. Conclusion

This study proposes a new model-free approach to approximating the expected quadratic variations of asset prices based on related options premia. The relation between quadratic variations and option prices is model-free and it is illustrated within the framework of the familiar Black-Scholes implied volatility scale. The new approximation method avoids numerical integration by taking advantage of the form of the integral with respect to the standard normal density and using the polynomial interpolation of the integrand. Based on this approach, a new volatility index is developed using the Nikkei 225 index options. The new volatility benchmark provides good approximations of the true values generated under the Heston stochastic volatility model. This is indicative of better numerical efficiency than alternative approaches such as the CBOE procedure, which rely on discretization to evaluate integrals. Based on various estimators of realized variance, the empirical evidence from market prices also suggests that the new volatility index is associated with better forecast accuracy. The model-free formula for the expected quadratic variations is thus conducive to the reduction of approximation errors and to improved numerical efficiency and forecasting accuracy.

Appendix A. Pricing Formula for Log Contract

In this appendix, we give the proofs of Theorems 2.1 and 2.2. At first, let us observe that the Black-Scholes implied volatility $\sigma : \mathbb{R} \to [0, \infty)$ is well-defined. The value $P_{\mathrm{BS}}(k, \sigma)$ coincides with the undiscounted version of the Black-Scholes put option price with strike $K = Fe^k$, maturity T and volatility $\sigma > 0$. It is well-known that the Black-Scholes price is an increasing function of the volatility parameter. Hence the inverse function $P_{\mathrm{BS}}(k, \cdot)^{-1}$ is well-defined. Besides, it holds that

$$\lim_{\sigma \to 0} P_{\mathrm{BS}}(k, \sigma) = F(e^k - 1)_+ \quad \text{and} \quad \lim_{\sigma \to \infty} P_{\mathrm{BS}}(k, \sigma) = Fe^k.$$

Because

$$(K - F)_+ \leq P(K) < K$$

for all $K > 0$ due to Jensen's inequality, $\sigma(k)$ is well-defined for all $k \in \mathbb{R}$.

Since S_T admits a density by assumption, we have

$$D(K) := \mathbb{E}[K > S_T] = \frac{\mathrm{d}P}{\mathrm{d}K}(K).$$

Notice that D is an increasing function of K. Let us suppose, for simplicity, that $T = 1$ hereafter. Let $D_{\mathrm{BS}}(K)$ be a function of $K = Fe^k$ defined as

$$D_{\mathrm{BS}}(K) = \frac{\mathrm{d}P_{\mathrm{BS}}}{\mathrm{d}K}(\log(K/F), \Sigma)|_{\Sigma = \sigma(\log(K/F))} = \Phi(-d_2(k, \sigma(k))).$$

By definition,

$$\begin{aligned} D(K) &= \frac{\mathrm{d}P}{\mathrm{d}K}(K) \\ &= \frac{\mathrm{d}}{\mathrm{d}K} P_{\mathrm{BS}}(\log(K/F), \sigma(\log(K/F))) \\ &= D_{\mathrm{BS}}(K) + \frac{1}{K}\frac{\partial P_{\mathrm{BS}}}{\partial \Sigma}(\log(K/F), \sigma(\log(K/F)))\frac{\mathrm{d}\sigma}{\mathrm{d}k}(\log(K/F)) \\ &= D_{\mathrm{BS}}(K) + \phi(-d_2(\log(K/F), \sigma(\log(K/F))))\frac{\mathrm{d}\sigma}{\mathrm{d}k}(\log(K/F)). \end{aligned} \tag{A.1}$$

Lemma A.1. *For every $k \in \mathbb{R}$, we have*

$$-d_2(k, \sigma(k))\frac{\mathrm{d}\sigma}{\mathrm{d}k}(k) < 1.$$

Proof. Put $f(k) := -d_2(k, \sigma(k))$. The inequality is trivial when $f(k) = 0$. If $f(k) > 0$, it follows from (A.1) that

$$f(k)\frac{\mathrm{d}\sigma}{\mathrm{d}k}(k) = f(k)\frac{D(Fe^k) - D_{\mathrm{BS}}(Fe^k)}{\phi(f(k))} \leq f(k)\frac{1 - \Phi(f(k))}{\phi(f(k))} < 1.$$

Here we used the fact that $0 \leq D(K) \leq 1$ by definition and a well-known estimate

$$1 - \Phi(x) < x^{-1}\phi(x), \quad x > 0. \tag{A.2}$$

For the case $f(k) < 0$, we have

$$\begin{aligned} f(k)\frac{\mathrm{d}\sigma}{\mathrm{d}k}(k) &= f(k)\frac{D(Fe^k) - D_{\mathrm{BS}}(Fe^k)}{\phi(f(k))} \leq -f(k)\frac{\Phi(f(k))}{\phi(f(k))} \\ &= -f(k)\frac{1 - \Phi(-f(k))}{\phi(-f(k))} < 1. \end{aligned}$$

□

Lemma A.2. *For all k with $d_1(k, \sigma(k)) \geq 0$, we have*

$$-d_1(k, \sigma(k))\frac{\mathrm{d}\sigma}{\mathrm{d}k}(k) < 1.$$

Proof. By definition, it holds that for all $K > 0$,

$$KD(K) \geq P(K).$$

Combining this and (A.1), we have

$$F\Phi(-d_1(k, \sigma(k))) + K\phi(-d_2(k, \sigma(k)))\frac{\mathrm{d}\sigma}{\mathrm{d}k}(k) \geq 0$$

with $k = \log(K/F)$. Since $K\phi(-d_2) = F\phi(-d_1)$, we obtain by (A.2)

$$\frac{\mathrm{d}\sigma}{\mathrm{d}k}(k) \geq -\frac{1 - \Phi(d_1(k, \sigma(k)))}{\phi(d_1(k, \sigma(k)))} > -\frac{1}{d_1(k, \sigma(k))}.$$

□

Lemma A.3. *The mapping $k \mapsto -d_2(k, \sigma(k))$ is increasing.*

Proof. Put $f(k) := -d_2(k, \sigma(k))$. By definition,

$$f(k) = \frac{k}{\sigma(k)} + \frac{\sigma(k)}{2},$$

and therefore

$$\begin{aligned}\frac{\mathrm{d}f}{\mathrm{d}k}(k) &= \frac{1}{\sigma(k)}\left\{1 - \frac{\mathrm{d}\sigma}{\mathrm{d}k}(k)\frac{k}{\sigma(k)}\right\} + \frac{1}{2}\frac{\mathrm{d}\sigma}{\mathrm{d}k}(k) \\ &= \frac{1}{\sigma(k)}\left\{1 - \frac{\mathrm{d}\sigma}{\mathrm{d}k}(k)f(k)\right\} + \frac{\mathrm{d}\sigma}{\mathrm{d}k}(k). \qquad \text{(A.3)}\end{aligned}$$

Hence, by Lemma A.1, we have

$$\frac{\mathrm{d}f}{\mathrm{d}k}(k) > \frac{\mathrm{d}\sigma}{\mathrm{d}k}(k).$$

It suffices then to treat the case $\mathrm{d}\sigma/\mathrm{d}k < 0$. By rewriting (A.3), we have

$$\frac{\mathrm{d}f}{\mathrm{d}k}(k) = \frac{1}{\sigma(k)}\left\{1 + \frac{\mathrm{d}\sigma}{\mathrm{d}k}(k)d_1(k, \sigma(k))\right\}.$$

If $d_1(k, \sigma(k)) < 0$, we have $\mathrm{d}f/\mathrm{d}k > 0$ under $\mathrm{d}\sigma/\mathrm{d}k < 0$. If $d_1(k, \sigma(k)) \geq 0$, we can use Lemma A.2 to obtain the same inequality. □

Lemma A.4. *It holds for $k \geq 0$ that*

$$-d_2(k, \sigma(k)) \geq \sqrt{2k} \quad \textit{and} \quad \frac{\mathrm{d}\sigma}{\mathrm{d}k}(k) < \frac{1}{\sqrt{2k}}.$$

Moreover, there exists $k^* > 0$ such that $\sigma(k^*) = \sqrt{2k^*}$ and it holds for all $k > k_0 \geq k^*$ that

$$\sigma(k) < \sigma(k_0) - \sqrt{2k_0} + \sqrt{2k} \leq \sqrt{2k}.$$

Proof. The first inequality comes from the fact that the arithmetic mean exceeds the geometric mean. The second follows from this and Lemma A.1. To consider the last claim, observe that

$$\mathbb{E}[(S_T - K)_+] = \mathbb{E}[(K - S_T)_+] + F - K = F\Phi(d_1(k, \sigma(k))) - K\Phi(d_2(k, \sigma(k)))$$

by definition, or "Call-Put Parity". The left-hand side goes to 0 as $K \to \infty$ and by (A.2)

$$\begin{aligned}K\Phi(d_2(k, \sigma(k))) &= Fe^k\phi(-d_2(k, \sigma(k)))\frac{1 - \Phi(-d_2(k, \sigma(k)))}{\phi(-d_2(k, \sigma(k)))} \\ &< \frac{F}{2\sqrt{\pi k}} \to 0 \quad \text{as } k \to \infty.\end{aligned}$$

Hence $d_1(k) \to -\infty$ as $k \to \infty$, which implies that $\sigma(k) < \sqrt{2k}$ for sufficiently large k. The last inequality holds since $\sqrt{2k} - \sigma(k)$ is increasing by the second inequality.

Here, notice that results similar to the previous lemma were obtained by Lee [16] and Rogers and Tehranchi [19]. □

Lemma A.5. *If there exists $p > 0$ such that S_T^{-p} is integrable, then we have*

$$\lim_{k\to\pm\infty} \sigma(k)\phi(d_2(k,\sigma(k))) = 0 \quad \textit{and} \quad \lim_{k\to\pm\infty} k\frac{\mathrm{d}\sigma}{\mathrm{d}k}(k)\phi(d_2(k,\sigma(k))) = 0. \tag{A.4}$$

Proof. By the results of Lee [16], there exist $k^* < 0$ and $\beta \in (0,2)$ such that

$$-\frac{\sigma(k)^2}{k} < \beta$$

for all $k < k^*$. It follows that

$$d_2(k,\sigma(k)) = -\frac{k}{\sigma(k)} - \frac{\sigma(k)}{2} > \sqrt{\frac{|k|}{\beta}} - \frac{\sqrt{\beta|k|}}{2} = \frac{2-\beta}{2\sqrt{\beta}}\sqrt{|k|}.$$

With the aid of Lemma A.4, these estimates imply that $\phi(d_2(k,\sigma(k)))$ decays exponentially fast as $|k| \to \infty$ and that the first claim of (A.4) holds. To see the second claim, notice that by (A.1)

$$D(Fe^k) - D_{\mathrm{BS}}(Fe^k) = \phi(-d_2(k,\sigma(k)))\frac{\mathrm{d}\sigma}{\mathrm{d}k}(k),$$

so that it now suffices to show

$$\lim_{k\to\infty} kD(Fe^k) = 0, \quad \lim_{k\to\infty} k(1 - D(Fe^k)) = 0,$$

$$\lim_{k\to\infty} kD_{\mathrm{BS}}(Fe^k) = 0, \quad \lim_{k\to\infty} k(1 - D_{\mathrm{BS}}(Fe^k)) = 0.$$

The first-line equations follow from the integrability condition of S_T, whereas the remaining two equations are obvious from the fact that $D_{\mathrm{BS}}(Fe^k) = \Phi(-d_2(k,\sigma(k)))$. □

Proof of Theorem 2.1. Since the second derivative of P coincides with the density of S_T,

$$\begin{aligned} -2\mathbb{E}[\log(S_T/F)] &= -2\int_0^\infty \log(K/F)\frac{\mathrm{d}^2P}{\mathrm{d}K^2}(K)\mathrm{d}K \\ &= -2\int_{-\infty}^\infty k\frac{\mathrm{d}^2P}{\mathrm{d}K^2}(Fe^k)Fe^k\mathrm{d}k. \end{aligned}$$

By Lemma A.3, the inverse function g of d_2 is well-defined. Using (A.1), we have

$$\begin{aligned} \frac{\mathrm{d}^2P}{\mathrm{d}K^2}(Fe^k) &= \frac{\mathrm{d}D}{\mathrm{d}K}(Fe^k) \\ &= \frac{1}{Fe^k}\phi(g^{-1}(k))\left\{-\frac{\mathrm{d}g^{-1}}{\mathrm{d}k}(k)\left(1 + g^{-1}(k)\frac{\mathrm{d}\sigma}{\mathrm{d}k}(k)\right) + \frac{\mathrm{d}^2\sigma}{\mathrm{d}k^2}(k)\right\}. \end{aligned}$$

Since

$$\frac{\mathrm{d}}{\mathrm{d}k}\phi(g^{-1}(k)) = -\phi(g^{-1}(k))g^{-1}(k)\frac{\mathrm{d}g^{-1}}{\mathrm{d}k}(k),$$

we have

$$-\int_{-\infty}^{\infty} k\phi(g^{-1}(k))g^{-1}(k)\frac{\mathrm{d}g^{-1}}{\mathrm{d}k}(k)\frac{\mathrm{d}\sigma}{\mathrm{d}k}(k)\mathrm{d}k$$
$$= \left[k\frac{\mathrm{d}\sigma}{\mathrm{d}k}(k)\phi(g^{-1}(k))\right]_{-\infty}^{\infty} - \int_{-\infty}^{\infty}\left\{\frac{\mathrm{d}\sigma}{\mathrm{d}k}(k) + k\frac{\mathrm{d}^2\sigma}{\mathrm{d}k^2}(k)\right\}\phi(g^{-1}(k))\mathrm{d}k.$$

By (A.4), we obtain

$$-2\mathbb{E}[\log(S_T/F)] = 2\int_{-\infty}^{\infty}\phi(g^{-1}(k))\left\{k\frac{\mathrm{d}g^{-1}}{\mathrm{d}k}(k) + \frac{\mathrm{d}\sigma}{\mathrm{d}k}(k)\right\}\mathrm{d}k.$$

Since

$$\int_{-\infty}^{\infty}\phi(g^{-1}(k))\frac{\mathrm{d}\sigma}{\mathrm{d}k}(k)\mathrm{d}k = \left[\phi(g^{-1}(k))\sigma(k)\right]_{-\infty}^{\infty}$$
$$+\int_{-\infty}^{\infty}\phi(g^{-1}(k))g^{-1}(k)\sigma(k)\frac{\mathrm{d}g^{-1}}{\mathrm{d}k}(k)\mathrm{d}k$$

and

$$k + g^{-1}(k)\sigma(k) = k + d_2(k,\sigma(k))\sigma(k) = -\frac{\sigma(k)^2}{2}$$

by definition, we obtain

$$-2\mathbb{E}[\log(S_T/F)] = -\int_{-\infty}^{\infty}\phi(g^{-1}(k))\sigma(k)^2\frac{\mathrm{d}g^{-1}}{\mathrm{d}k}(k)\mathrm{d}k$$
$$= \int_{-\infty}^{\infty}\sigma(g(z))^2\phi(z)\mathrm{d}z.$$

Here we have used (A.4). □

Proof of Theorem 2.2. The first inequality for the derivative of $\hat{\sigma}$ follows from (A.3) and Lemma A.1. The existence of z^* is equivalent to the existence of k^* in Lemma A.4. The other inequalities follow from the fact that

$$k \mapsto d_2(k,\sigma(k))^2 - d_1(k,\sigma(k))^2 = 2k$$

is an increasing function. □

Appendix B. Realized Variance Estimators

Here, we start with a brief review of various estimators of realized variance employed in Sec. 5. It is assumed that the logarithmic equilibrium price follows a continuous semimartingale process $dp^*(s) = \mu(s)ds + \sigma(s)dW(s)$, where $\mu(s)$ is a drift coefficient, $W(s)$ is a standard Brownian motion, and $\sigma^2(s)$ is the instantaneous variance

of the equilibrium price p^*. We consider that the parameter of interest is an integrated variance over the tth trading day, $[t, t+1)$, defined as $V_t := \int_t^{t+1} \sigma_t^2 \mathrm{d}t$. Throughout this section, let E be the expectation operator under the physical measure.

B.1. *RV: Realized variance with returns sampled at the highest frequencies*

Let the ith intraday price on a day t be denoted by $p(t_i)$. One way to estimate the integrated variance V is to use the sum of squared returns:

$$RV = \sum_{i=1}^{n} \left(p(t_i) - p(t_{i-1})\right)^2, \tag{B.1}$$

where n represents the number of observed intraday returns in $[t, t+1)$. The estimator is generally called realized variance or realized volatility. If $p(t_i)$ is equal to the equilibrium price $p^*(t_i)$, then RV provides a consistent estimate of V.

However, RV fails to satisfy the consistency condition when there is market microstructure noise as usually documented in real high-frequency data. The microstructure noise can be induced by various market frictions such as the discreteness of price changes, bid-ask bounces, and asymmetric information across traders, inter alia. For the rest of this appendix, assume the ith intraday return r_i is contaminated by microstructure noise as follows:

$$r_i = \underbrace{p^*(t_i) - p^*(t_{i-1})}_{r_i^*} + \underbrace{\eta(t_i) - \eta(t_{i-1})}_{\epsilon_i} \tag{B.2}$$

where η represents microstructure noise. A growing literature attempts to examine different estimators of integrated variance from microstructure noise-contaminated high-frequency data, including Zhang *et al.* [20], inter alia.

B.2. *$RV_{5\,\mathrm{min}}$ and $RV_{15\,\mathrm{min}}$: Realized variance with 5- and 15-minute returns*

Another classical method uses the realized variance constructed from intraday returns sampled at moderate frequencies rather than at the highest frequency because the realized variance with finer frequency is more sensitive to microstructure noise. This approach can partially offset the bias induced by the microstructure effects. In practice, researchers are necessarily forced to select a moderate sampling frequency. For example, it may be regarded as around those frequencies for which realized variance signature plots under alternative sampling frequencies are leveled off. Evidence from previous studies suggests that it is optimal to use 5 to 30-minute return data. Hence, we employ $RV_{5\,\mathrm{min}}$ and $RV_{15\,\mathrm{min}}$ which are equal to the sum of squared 5- and 15-minute returns.

B.3. RV_{BR}: *Optimally-sampled realized variance*

There is a trade-off between the microstructure noise-induced bias and variance reduction at high sampling frequencies. To take this trade-off into account, Bandi and Russell [2] provide a theoretical justification for the choice of optimal sampling frequency based on the mean squared error (MSE) criterion. They derive the following approximation of the optimal number of observations n^* based on MSE minimization in a finite sample $n^* \approx (\frac{IQ}{(\mathrm{E}(\epsilon^2))^2})^{\frac{1}{3}}$, where IQ represents an integrated quarticity of the equilibrium price process ($IQ = \int_0^T \sigma^4(s)\mathrm{d}s$). It is estimated by realized quarticity $\hat{IQ} = \frac{n}{3}\sum_{i=1}^{n} r_i^4$ with low frequency returns such as 15-minute returns. Following the consistent estimator of noise moment as shown by Bandi and Russell [2], $\mathrm{E}(\epsilon^2)$ can be estimated by $\hat{\mathrm{E}}(\epsilon^2) = \frac{1}{n}\sum_{i=1}^{n} r_i^2$ at the highest frequency. Thus, RV_{BR} is equal to RV with $\hat{n}^* = (\hat{IQ}/(\hat{\mathrm{E}}(\epsilon^2))^2)^{1/3}$.

B.4. RV_{BR}^{BK}: *The Bartlett-type kernel estimator in Barndorff-Nielsen et al.* [4] *with a finite sample optimal number of autocovariances proposed by Bandi and Russell* [3]

$RV_{5\,\mathrm{min}}$, $RV_{15\,\mathrm{min}}$ and RV_{BR} have the obvious drawback that they do not incorporate all available observations, and useful information may thereby be lost. The problem of estimating the integrated variance under microstructure noise is similar to the autocorrelation corrections that are used in the estimation of long-run variance in stationary time-series (see, for instance, Newey and West [18] and Andrews [1]). So it is natural to consider kernel-based estimators of integrated variance under microstructure noise. Barndorff-Nielsen *et al.* [4] examine the Bartlett-type kernel estimator defined as

$$RV^{BK} = \left(\frac{n-1}{n}\frac{H-1}{H}\right)\gamma_0 + 2\sum_{h=1}^{H}\left(\frac{H-h}{H}\right)\gamma_h, \tag{B.3}$$

where $\gamma_h = \sum_{i=1}^{n-h} r_i r_{i+h}$ is the hth autocovariance of intraday returns and γ_0 is equal to realized variance using returns sampled at the highest frequencies. This estimator weights the realized variance and the Hth return autocovariances by Bartlett weights. The optimal number of autocovariances is obtained through MSE minimization in finite sample (see Eqs. (7) to (10) in Bandi and Russell [3] for the exact MSE minimization expressions). There is a convenient rule-of-thumb for choosing H in practice as proposed in Bandi and Russell [3]. The expression is obtained as $H^{BR} \approx (\frac{3V^2}{2n^2 IQ})^{\frac{1}{3}} n$. V and IQ are estimated using realized variance and realized quarticity with lower frequency returns such as 15-minute returns. Hence, RV^{BK} with finite-sample optimal number of autocovariances H^{BR} leads to RV_{BR}^{BK}.

B.5. RV^{ZMA}: The two-scale estimator with an asymptotically optimal number of subsamples proposed by Zhang et al. [20]

A two-scale or subsampling estimator is proposed by Zhang *et al.* [20] in the spirit of the estimation of the long-run variance studied by Carlstein [6]. Denote the original grid of observation times as $\Psi = \{t_0, t_1, t_2, \ldots, t_n\}$. Consider Ψ is partitioned into $\tilde{K}$ nonoverlapping subgrids, $\Psi^{(j)}_{\tilde{K}} = \{t_{j-1}, t_{j-1+\tilde{K}}, t_{j-1+2\tilde{K}}, \ldots\}$, $j = 1, \ldots, \tilde{K}$, Then, the realized variance for the subgrid $\Psi^{(j)}_{\tilde{K}}$ is defined as

$$RV^{(j)}_{\tilde{K}} = \sum_{i=1}^{n_j} (p(t_{(j-1)+i\tilde{K}}) - p(t_{(j-1)+(i-1)\tilde{K}}))^2. \tag{B.4}$$

The two-scale estimator in Zhang *et al.* [20] is given by

$$ZMA = (1/\tilde{K}) \sum_{j=1}^{\tilde{K}} RV^{(j)}_{\tilde{K}} - (\bar{n}/n) RV, \tag{B.5}$$

where $\bar{n} = (n - \tilde{K} + 1)/\tilde{K}$ and RV is the realized variance for the full grid Ψ. The second term corrects the bias in the first term. The asymptotic optimal number of subsamples $\tilde{K}^{ZMA}$ derived by minimizing the estimator's asymptotic variance is given by $\tilde{K}^{ZMA} = (\frac{3(\mathrm{E}(\epsilon^2))^2}{IQ})^{1/3} n^{2/3}$. IQ and $\mathrm{E}(\epsilon^2)$ are estimated by realized quarticity with 15-minute returns and $\hat{\mathrm{E}}(\epsilon^2) = \frac{1}{n}\sum_{i=1}^{n} r_i^2$ at the highest frequency, respectively. Thus, ZMA with $\tilde{K}^{ZMA}$ leads to RV^{ZMA}.

B.6. RV^{ZMA}_{BR}: The two-scale estimator with a finite-sample optimal number of subsamples proposed by Bandi and Russell [3]

Barndorff-Nielsen *et al.* [4] show that the two-scale estimator is almost identical to the modified Bartlett kernel estimator. Bandi and Russell [3] additionally show that the finite sample MSEs of RV^{BK} and ZMA are very similar in practice. Hence, the ZMA with $\tilde{K} = H^{BR}$ is represented by RV^{ZMA}_{BR}.

B.7. RV^{BC-ZMA}: The bias-corrected two-scale estimator with an asymptotically optimal number of subsamples proposed by Zhang et al. [20]

The two-scale estimator ZMA has a finite-sample bias as shown in Zhang *et al.* [20] who provide an approximate correction for this bias. On the other hand, Bandi and Russell [3] report the exact bias-correction form. The bias-corrected estimator can be defined, following a suggestion by Bandi and Russell [3], as

$$BC(ZMA) = c(\tilde{K}, n) ZMA, \quad c(\tilde{K}, n) = \left(\frac{\tilde{K}n - 1 + 2\tilde{K} - \tilde{K}^2 - n}{\tilde{K}n} \right)^{-1}. \tag{B.6}$$

Since $BC(ZMA)$ is asymptotically equivalent to ZMA, the asymptotically optimal number of subsamples is given by $\tilde{K}^{ZMA}$. Thus, $BC(ZMA)$ with $\tilde{K}^{ZMA}$ can be described by RV^{BC-ZMA}.

B.8. RV_{BR}^{BC-ZMA}: The bias-corrected two-scale estimator with a finite-sample optimal number of subsamples proposed by Bandi and Russell [3]

Since $BC(ZMA)$ is unbiased in a finite-sample, the optimal number of subsamples is provided by minimizing the finite-sample variance of $BC(ZMA)$. Bandi and Russell [2, 3] show that the optimal number of subsamples is defined as

$$\tilde{K}^{BR} = \underset{0<\tilde{K}/n\leq 1/2}{\arg\min}\ [\mathrm{Var}(BC(ZMA))] = \underset{0<\tilde{K}/n\leq 1/2}{\arg\min}\ [(c(K,n))^2\ \mathrm{Var}(ZMA)], \quad \text{(B.7)}$$

where $\mathrm{Var}(ZMA)$ can be expressed in Eq. 6 of Bandi and Russell [2]. Hence, $BC(ZMA)$ with $\tilde{K}^{BR}$ can be represented by RV_{BR}^{BC-ZMA}.

B.9. RV^{FBK}: The flat-top Bartlett kernel estimator with an asymptotically optimal number of autocovariances proposed by Barndorff-Nielsen et al. [5]

Barndorff-Nielsen *et al.* [5] examine the following unbiased flat-top kernel type estimator (called realized kernel)

$$RK = \gamma_0 + \sum_{h=1}^{H} k(x)(\gamma_h + \gamma_{-h}), \quad \text{(B.8)}$$

where the non-stochastic $k(x) \in [0,1]$ for $x = \frac{h-1}{H}$ is a weight function and $\gamma_h = \sum_{i=1}^{n} r_i r_{i-h}$ with $h = -H, \ldots, H$. The flat-top Bartlett kernel estimator is equivalent to RK in case where $k(x) = 1 - x$. For this class of kernels, Barndorff-Nielsen *et al.* [5] show that the asymptotic distribution of $RK - V$ is mixed normal with zero mean and rate of convergence $n^{1/6}$ when $H = cn^{2/3}$ where c is a constant. Then, the asymptotically optimal value of c which minimizes the asymptotic variance is given by $c^* \approx 2.28\xi^{\frac{4}{3}}$, where $\xi^2 = \sigma_\eta^2/\sqrt{IQ}$, $\sigma_\eta^2 = \frac{1}{2n}\sum_{i=1}^{n} r_i^2$. Hence, RK with $k(x) = 1 - x$ and $H = c^* n^{2/3}$ corresponds to RV^{FBK}.

B.10. RV^{FCK} and RV^{FMTH}: The flat-top cubic kernel estimator and the flat-top modified Tukey-Hanning kernel estimator with an asymptotically optimal number of autocovariances proposed by Barndorff-Nielsen et al. [5]

The estimators based on the cubic kernel and the modified Tukey-Hanning kernel are equivalent to RK with $k(x) = 1 - 3x^2 + 2x^3$ and $k(x) = \{1 - \cos\pi(1-x)^2\}/2$. When $H = c\xi n^{1/2}$, RK for this class of kernels is consistent at the rate of convergence

$n^{1/4}$ as shown in Barndorff-Nielsen *et al.* [5]. The asymptotically optimal value of c is expressed as

$$c^* = \sqrt{\rho \frac{k_\bullet^{1,1}}{k_\bullet^{0,0}} \left\{ 1 + \sqrt{1 + \frac{3k_\bullet^{0,0} k_\bullet^{2,2}}{\rho (k_\bullet^{1,1})^2}} \right\}}, \tag{B.9}$$

where $\rho = V/\sqrt{IQ}$, $k_\bullet^{0,0} = \int_0^1 k(x)^2 dx$, $k_\bullet^{1,1} = \int_0^1 k'(x)^2 dx$ and $k_\bullet^{2,2} = \int_0^1 k''(x)^2 dx$, where the primes represent derivatives. The values of $(k_\bullet^{0,0}, k_\bullet^{1,1}, k_\bullet^{2,2})$ amount to $(k_\bullet^{0,0}, k_\bullet^{1,1}, k_\bullet^{2,2}) = (0.371, 1.20, 12.0)$ for cubic kernel and $(k_\bullet^{0,0}, k_\bullet^{1,1}, k_\bullet^{2,2}) = (0.219, 1.71, 41.7)$ for modified Tukey-Hanning kernel. We define RV^{FCK} and RV^{FMTH} as RK with $H = c^* \xi n^{1/2}$ at $k(x) = 1 - 3x^2 + 2x^3$ and $k(x) = \{1 - \cos\pi(1-x)^2\}/2$.

B.11. *RV_{BR}^{FBK}, RV_{BR}^{FCK} and RV_{BR}^{FMTH}: The flat-top Bartlett kernel estimator, the flat-top cubic kernel estimator and the flat-top modified Tukey-Hanning kernel estimator with a finite-sample optimal number of autocovariances proposed by Bandi and Russell* [3]

Bandi and Russell [3] provide an alternative way to choose the number of autocovariances in finite samples. Denote H as ϕn with $0 < \phi \leq 1$. The optimal value of ϕ is defined as

$$\phi^* = \underset{0<\phi\leq 1}{\arg\min} \left[(\text{bias}(RK))^2 + \text{Var}(RK) \right], \tag{B.10}$$

where $\text{bias}(RK) = 0$ and $\text{Var}(RK)$ can be expressed in Theorem 3 of Bandi and Russell [3]. Thus, RK with $H = \phi^* n$ for Bartlett kernel, cubic kernel and modified Tukey-Hanning kernel leads to RV_{BR}^{FBK}, RV_{BR}^{FCK} and RV_{BR}^{FMTH}, respectively.

Acknowledgments

The authors are thankful for the financial support from the Osaka Securities Exchange and the Japan Science and Technology Agency CREST. Masaaki Fukasawa and Kazutoshi Yamazaki acknowledge the grants-in-aid for scientific research No. 21740074 and No. 22710143, respectively, from the Japanese Ministry for Education, Culture, Sports, Science and Technology. Kosuke Oya and Masato Ubukata are thankful for the grants-in-aid for scientific research No. 22243021 from the Japanese Society for the Promotion of Science. Isao Ishida also acknowledges the grants-in-aid for scientific research No. 20530265. Nabil Maghrebi is thankful for the grants-in-aid No. 21330077, and the Center for the Study of Finance and Insurance where this study was completed while he was Visiting Professor. The authors also wish to thank the Editor and anonymous referee for useful comments and suggestions from which this paper has greatly benefited.

References

[1] D. W. K. Andrews, Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica* **59** (1991) 817–858.

[2] F. M. Bandi and J. R. Russell, Microstructure noise, realized variance, and optimal sampling, *Review of Economic Studies* **75** (2008) 339–369.

[3] F. M. Bandi and J. R. Russell, Market microstructure noise, integrated variance estimators, and the accuracy of asymptotic approximations, *Journal of Econometrics* **160** (2011) 145–149.

[4] O. E. Bandorff-Nielsen, P. R. Hansen, A. Lunde and N. Shephard, Regular and modified kernel-based estimators of integrated variance: The case with independent noise? Economic Papers 2004-W28, Economics Group, Nuffield College, Oxford University (2004).

[5] O. E. Barndorff-Nielsen, P. R. Hansen, A. Lunde and N. Shephard, Designing realized kernels to measure the ex-post variation of equity prices in the presence of noise, *Econometrica* **76** (2008) 1481–1536.

[6] E. Carlstein, The use of subseries values for estimating the variance of a general statistic from a stationary sequence, *Annals of Statistics* **14** (1986) 1171–1179.

[7] P. Carr and R. Lee, Volatility derivatives, *Annual Review of Financial Economics* **1** (2009) 1–21.

[8] P. Carr, R. Lee and L. Wu, Variance swaps on time-changed Lévy processes, *Finance and Stochastics*, forthcoming (2009).

[9] P. Carr and L. Wu, A tale of two indices, *Journal of Derivatives* **13**(3) (2006) 13–29.

[10] N. Chriss and W. Morokoff, Market risk for volatility and variance swaps, *Risk* October (1999).

[11] M. Fukasawa, The normalizing transformation of the implied volatility smile, *Mathematical Finance*, forthcoming (2010).

[12] J. Gatheral, *The Volatility Surface; A Practitioner's Guide* (Wiley, 2006).

[13] P. R. Hansen and A. Lunde, A forecast comparison of volatility models: Does anything beat a GARCH(1,1)? *Journal of Applied Econometrics* **20** (2005) 873–889.

[14] S. L. Heston, A Closed-form solution for options with stochastic volatility with applications to bond and currency options, *The Review of Financial Studies* **6**(2) (1993) 327–343.

[15] G. T. Jiang and Y. S. Tian, Extracting model-free volatility from option prices: An examination of the VIX index, *Journal of Derivatives* (2007) 35–60.

[16] R. W. Lee, The moment formula for implied volatility at extreme strikes, *Mathematical Finance* **14**(3) (2004) 469–480.

[17] W. Morokoff, Y. Akesson and Y. Zhou, Risk management of volatility and variance swaps, *Firmwide Risk Quantitative Modeling Notes*, Goldman Sachs & Co (1999).

[18] W. K. Newey and K. D. West, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* **55** (1987) 703–708.

[19] C. Rogers and M. Tehranchi, Can the implied volatility surface move by parallel shifts, *Finance and Stochastics* **14**(2) (2010) 235–248.

[20] L. Zhang, P. A. Mykland and Y. Aït-Sahalia, A tale of two time scales: Determining integrated volatility with noisy high-frequency data, *Journal of the American Statistical Association* **100** (2005) 1394–1411.