The Model-Free Implied Volatility and Its Information Content

George J. Jiang
Eller College of Management, University of Arizona

Yisong S. Tian
Schulich School of Business, York University

Britten-Jones and Neuberger (2000) derived a model-free implied volatility under the diffusion assumption. In this article, we extend their model-free implied volatility to asset price processes with jumps and develop a simple method for implementing it using observed option prices. In addition, we perform a direct test of the informational efficiency of the option market using the model-free implied volatility. Our results from the Standard & Poor's 500 index (SPX) options suggest that the model-free implied volatility subsumes all information contained in the Black–Scholes (B–S) implied volatility and past realized volatility and is a more efficient forecast for future realized volatility.

There has been considerable research on the forecasting ability and information content of the Black–Scholes (B–S; Black and Scholes, 1973) implied volatility. Since option prices reflect market participants' expectations of future movements of the underlying asset, the volatility implied from option prices is widely believed to be informationally superior to the historical volatility of the underlying asset. If the option market is informationally efficient and the B–S model is correct, implied volatility is expected to subsume all information contained in historical volatility and provides a more efficient forecast for future volatility.

Early studies find that implied volatility is a biased forecast of future volatility and contains little incremental information beyond historical volatility. For example, Canina and Figlewski (1993) found that the implied volatility from the Standard & Poor's (S & P) 100 index options is a poor forecast for the subsequent realized volatility of the underlying index. Based on an encompassing regression analysis, they found that implied volatility has virtually no correlation with future realized volatility and thus does not incorporate information contained in historical

We thank Yacine Aït-Sahalia (the editor), an anonymous referee, Stewart Hodges, Chris Lamoureux, Nathaniel O'Connor, Wulin Suo, Shu Yan, Hao Zhou, and seminar participants at the Federal Reserve Board, the University of Toronto, and the University of Warwick for helpful comments and suggestions. The financial support of the Social Sciences and Humanity Research Council of Canada is gratefully acknowledged. Address correspondence to Yisong S. Tian, Finance Area, Schulich School of Business, York University, 4700 Keele Street, Toronto, ON M3J 1P3, or e-mail: ytian@schulich.yorku.ca.

doi:10.1093/rfs/hhi027 Advance Access publication May 25, 2005

volatility. In contrast, Day and Lewis (1992), Lamoureux and Lastrapes (1993), Jorion (1995) and Fleming (1998) report evidence supporting the hypothesis that implied volatility has predictive power for future volatility. They also found that implied volatility is a biased forecast for future realized volatility.

More recent research attempts to correct various data and methodological problems in earlier studies. These studies [e.g., Christensen and Prabhala (1998), Christensen, Hansen, and Prabhala (2001), Blair, Poon, and Taylor (2001), Ederinton and Guan (2002), and Pong et al. (2004)] consider longer time series to take into account possible regime shift around the October 1987 crash, use instrumental variables (IVs) to correct for the errors-in-variable (EIV) problem in implied volatility, adopt high-frequency asset returns to provide a more accurate estimate for realized volatility, or use nonoverlapping samples to avoid the "telescoping overlap" problem. Collectively, these studies present evidence that implied volatility is a more efficient forecast for future volatility than historical volatility.

However, nearly all previous research on the information content of implied volatility focuses on the B–S implied volatility from at-the-money options. Granted, at-the-money options are generally more actively traded than other options and are certainly a good starting point. By concentrating on at-the-money options alone, however, these studies fail to incorporate the information contained in other options. More importantly, tests based on the B–S implied volatility are joint tests of market efficiency and the B–S model. These studies are thus subject to model misspecification errors.

In an important departure from previous research, we perform direct tests of the informational efficiency of the option market using an alternative implied volatility measure that is independent of option pricing models. This model-free implied volatility was derived by Britten-Jones and Neuberger (2000), building on the pioneering work of Breeden and Litzenberger (1978), Derman and Kani (1994, 1998), Rubinstein (1994, 1998), Derman, Kani, and Chriss (1996), and Ledoit and Santa-Clara (1998) on implied distributions. Unlike the traditional concept of implied volatility, their model-free implied volatility is not based on any specific option pricing model. Instead, it is derived entirely from no-arbitrage conditions. In particular, Britten-Jones and Neuberger (2000) showed that the risk-neutral integrated return variance between the current date and a future date is fully specified by the set of prices of options expiring on the future date (their Proposition 2).

However, Britten-Jones and Neuberger (2000) derived the model-free implied volatility under diffusion assumptions. It is unclear whether it is still valid when the underlying asset price process includes jumps. This can be a serious limitation since random jumps are an important aspect of the price dynamics of many financial assets. In this article, we extend

Britten-Jones and Neuberger (2000) and demonstrate that their model-free implied volatility is still valid even if the underlying asset price process has jumps. We first provide a simpler derivation under diffusion assumptions and then generalize it to processes with jumps. Our result ensures the generality of the model-free implied volatility.

In addition, we develop a simple method for implementing the model-free implied volatility using observed option prices. Defined as an integral of option prices, we show that it can be accurately calculated using trapezoidal integration. However, a greater challenge in practice is that options are traded only over a finite range of strike prices while an infinite range is required. We provide theoretical upper bounds for truncation errors when a finite range is used. We also demonstrate how truncation errors vary with the range of strike prices and identify the strike price range required to control such errors. More importantly, we show that the model-free implied volatility can be calculated accurately from observed option prices using a curve-fitting method and extrapolation from endpoint implied volatilities.

Finally, we investigate the forecasting ability and information content of the model-free implied volatility. The model-free implied volatility facilitates a direct test of the informational efficiency of the option market, rather than a joint test of market efficiency and the assumed option pricing model. It also aggregates information across options with different strike prices and should be informationally more efficient. We conduct our empirical tests using the S & P's 500 index (SPX) options traded on the Chicago Board Options Exchange (CBOE). Following prior research, we minimize measurement errors by using tick-by-tick data, commonly used data filters, nonoverlapping samples, and realized volatility estimated from high-frequency index returns.

Consistent with previous research, we find that the B–S implied volatility contains more information than the historical volatility of the underlying asset but is an inefficient forecast of future realized volatility. In contrast, we find that the model-free implied volatility subsumes all information contained in both the B–S implied volatility and the past realized volatility and is a more efficient forecast for future realized volatility. Our results provide support for the informational efficiency of the option market. Information in historical volatility is correctly incorporated into option prices, although the use of a single option is not sufficient to extract all relevant information. These findings are robust to alternative estimation methods, volatility series over different horizons, the choice of actual or implied index values, and alternative methods for calculating realized volatility.

The rest of the article proceeds as follows. The next section describes and generalizes the model-free implied volatility. A simple method for implementing the model-free implied volatility is also developed.

Numerical examples are provided to demonstrate the validity and accuracy of this method. In Section 2, we discuss the data used in this study and the calculation of volatility measures. The forecasting ability and information content of the model-free implied volatility are investigated in Section 3 using monthly nonoverlapping samples. Section 4 conducts robustness tests to ensure the generality of our results. Conclusions are in the final section.

1. Model-Free Implied Volatility

Britten-Jones and Neuberger (2000) derived the model-free implied volatility under diffusion assumptions. In this section, we first provide a simpler derivation under diffusion assumptions and then generalize it to processes with jumps. We also examine various implementation issues.

1.1 Diffusions and jumps

Suppose call options with a continuum of strike prices (K) for a given maturity (T) are traded on an underlying asset. Following Dumas, Fleming, and Whaley (1998) and Britten-Jones and Neuberger (2000), we consider the forward asset price and forward option price, denoted by F_t and $C^F(T,K)$, respectively, under the forward probability measure. The forward price F_t is a martingale under the forward measure even when both interest rates and dividends are stochastic. Britten-Jones and Neuberger's (2000) model-free implied volatility is defined in the following proposition:

Proposition 1. *The integrated return variance between the current date 0 and a future date T is fully specified by the set of prices of call options expiring on date T:*

$$E_0^F\left[\int_0^T \left(\frac{dF_t}{F_t}\right)^2\right] = 2\int_0^\infty \frac{C^F(T,K) - \max(0,F_0 - K)}{K^2}\,dK \qquad (1)$$

where the superscript F denotes the forward probability measure.

Although Britten-Jones and Neuberger (2000) relied on the diffusion assumption to derive Proposition 1, we show in the Appendix that it also holds when asset prices contain jumps.[1] Since a martingale can be decomposed canonically into the orthogonal sum of a purely continuous martingale and a purely discontinuous martingale [Jacod and Shiryaev (1987) and Protter (1990)], we generalize Britten-Jones and Neuberger's result to all martingale processes. Equation (1) is thus valid for a very general class of asset price processes and provides a model-free relationship between

[1] We are indebted to an anonymous referee for providing the idea and outline for the proof to Proposition 1.

asset return variance and option prices. Following Britten-Jones and Neuberger (2000), the right-hand side (RHS) of Equation (1) will be referred to as the model-free implied variance and its square root the model-free implied volatility.

1.2 Implementation issues

As shown on the RHS of Equation (1), the model-free implied volatility is defined as an integral of option prices over an infinite range of strike prices. If option prices are available for all strike prices, the required integral is straightforward to calculate using numerical integration methods. Of course, only a finite number of strike prices are actually traded in the marketplace which may lead to inaccuracies in the calculation of the model-free implied volatility. Limited availability of strike prices and other implementation issues are discussed below.

1.2.1 Truncation errors. Suppose the interval $[K_{\min}, K_{\max}]$ defines the range of available strike prices. To avoid trivial cases, we further require that $0 < K_{\min} < F_0 < K_{\max} < +\infty$. In order to focus on truncation errors, we also assume that all strike prices in the interval are available. Truncation errors are present when we ignore the tails of the distribution and approximate the RHS of Equation (1) by the following integral:

$$2\int_{K_{\min}}^{K_{\max}} \frac{C^F(T, K) - \max(0, F_0 - K)}{K^2} dK. \tag{2}$$

In the Appendix, we show that the truncation errors are related to the local variations in the tails of the asset return distribution. The results are summarized in the following proposition:

Proposition 2. *The right and left truncation errors beyond the strike price range $[K_{min}, K_{max}]$ have the following upper bounds, respectively:*

$$2\int_{K_{\max}}^{+\infty} \frac{C^F(T, K) - \max(0, F_0 - K)}{K^2} dK \le E_0^F\left[\left(\frac{F_T - K_{\max}}{K_{\max}}\right)^2 \middle| F_T > K_{\max}\right], \tag{3}$$

and

$$2\int_{0}^{K_{\min}} \frac{C^F(T, K) - \max(0, F_0 - K)}{K^2} dK \le E_0^F\left[\left(\frac{F_T - K_{\min}}{F_T}\right)^2 \middle| F_T < K_{\min}\right]. \tag{4}$$

Both upper bounds are quite intuitive as they reflect the local variations in the tails of the return distribution. Although these upper bounds are shown subsequently to be quite tight, they are not model free. We also derive model-free upper bounds for truncation errors in the Appendix. However, the model-free upper bounds are not as tight as the ones in equations (3) and (4).

To illustrate typical truncation errors, we consider the following stochastic volatility and random jump (SVJ) model:

$$\begin{aligned} dF_t/F_t &= V_t^{1/2} dW_t + J_t dN_t - \mu_J \lambda dt, \\ dV_t &= (\theta_v - \kappa_v V_t)dt + \sigma_v V_t^{1/2} dW_t^v, \\ dW_t dW_t^v &= \rho dt, \end{aligned} \tag{5}$$

where N_t~i.i.d. Poisson (λ) and $\ln(1 + J_t)$~i.i.d. $\mathrm{N}\left[\ln(1+\mu_J) - \frac{1}{2}\sigma^{2J}, \sigma^J\right]$. If $\lambda = 0$, the SVJ model reduces to the Heston (1993) model. We use the following two sets of parameters to illustrate the results: (I) $k_v = 1$, $\sigma_v = 0.25$, $\rho = 0$, $\lambda = 4$, $\mu_J = 0.0375$, $\theta_v = 0.1854^2 K_v$, $V_0 = \theta_v/k_v$ and (II) $k_v = 1$, $\sigma_v = 0.25$, $\rho = 0$, $\lambda = 0.5$, $\mu_J = -0.075$, $\theta_v = 0.1854^2 k_v$, $V_0 = \theta_v/k_v$. These parameters are consistent with empirical estimates from SPX options by Bakshi, Cao, and Chen (1997) and are chosen to have sufficient variations in implied volatility patterns (Figure 1).[2] Both sets of parameters imply an annualized volatility of 20%. We fix the initial forward price (F_0) at \$100. The implied volatility function and truncation errors are plotted in Figure 1 using these parameters and various option maturities.

For each set of parameters, we first calculate call prices using the SVJ model for various maturities and strike prices. The B–S implied volatility is then calculated and plotted against option moneyness (panels A and C). Option moneyness is defined as the ratio of strike price over asset price. Both smile and smirk patterns are illustrated in these plots, reflecting the behavior of the asset return distribution (such as fat tails and skewness). Next, we calculate the true truncation errors [the left-hand side (LHS) of equations (3) and (4)] for a given truncation interval. The left and right truncation points, $K_{\min}$ and $K_{\max}$, are both expressed as multiples of standard deviations (SDs) from the initial forward price (F_0), and the truncation errors are plotted against the corresponding multiples. As shown in panels B and D of Figure 1, the truncation error declines monotonically and diminishes as the truncation point moves away from the initial forward price. In particular, the left and right truncation errors

[2] The skewness and kurtosis of the daily return distribution implied by the first set of parameters are -0.03 and 7.43, respectively. The corresponding numbers for the second set are -2.22 and 41.87.

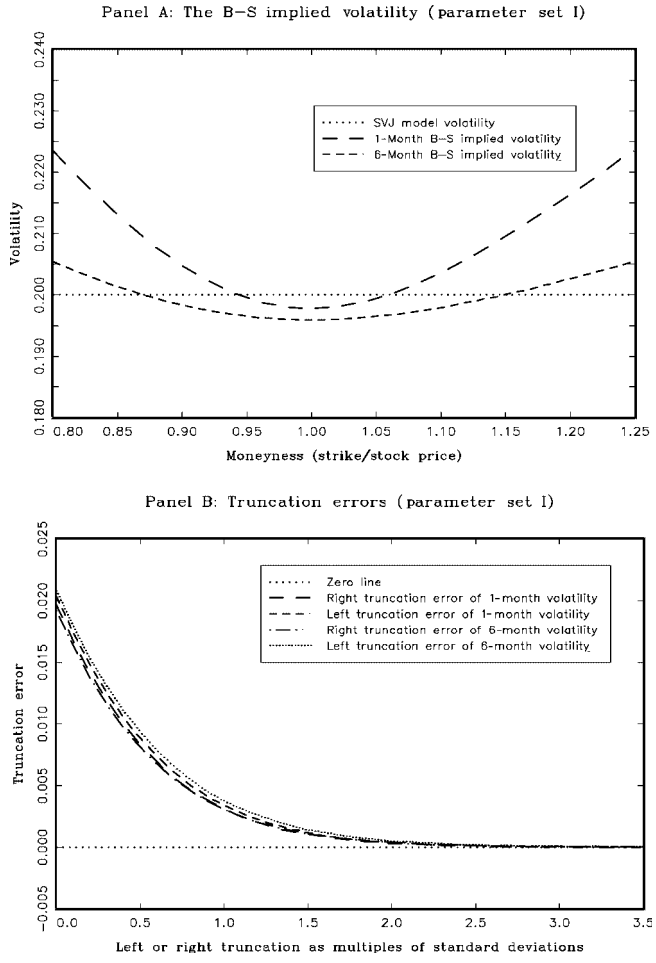

Figure 1
Truncation errors in calculating the model-free implied volatility
The left and right truncation errors of the model-free implied volatility are plotted against the respective truncation point ($K_{\min}$ or $K_{\max}$). The stochastic volatility with random jump (SVJ) model is assumed for the underlying asset price: $dF_t/F_t = V_t^{1/2}dW_t + J_t dN_t - \mu_J \lambda dt$, $dV_t = (\theta_v - k_v V_t)dt + \sigma_v V_t^{1/2} dW_t^v$, and $dW_t dW_t^v = \rho dt$ where $N_t \sim$ i.i.d. Poisson (λ) and $\ln(1+J_t) \sim$ iid $\mathrm{N}\left[\ln(1+\mu_J) - \frac{1}{2}\sigma_J^2, \sigma_J\right]$. Two sets of parameters are used: (I) $k_v = 1$, $\sigma_v = 0.25$, $\rho = 0$, $\lambda = 4$, $\mu_J = 0$, $\sigma_J = 0.0375$, $\theta_v = 0.1854^2 k_v$, $V_0 = \theta/k_v$ and (II) $k_v = 1$, $\sigma_v = 0.25$, $\rho = 0$, $\lambda = 0.5$, $\mu_J = -0.075$, $\sigma_J = 0.075$, $\theta_v = 0.1854^2 k_v$, $V_0 = \theta/k_v$. The two panels on the left (panels A and C) illustrate the shape of the Black–Scholes (B–S) implied volatility function for one- and six-month maturities while the two panels on the right (panels B and D) show the truncation errors at various truncation levels. The truncation points are stated in multiples of standard deviations (SDs).

are of similar magnitude (panel B) and exhibit nearly identical convergence properties when the return distribution is nearly symmetric (panel A). In comparison, the left truncation error is larger than the corresponding right truncation error (panel D) when the return distribution is skewed to the left (panel C). In this case, the left truncation errors converges at a

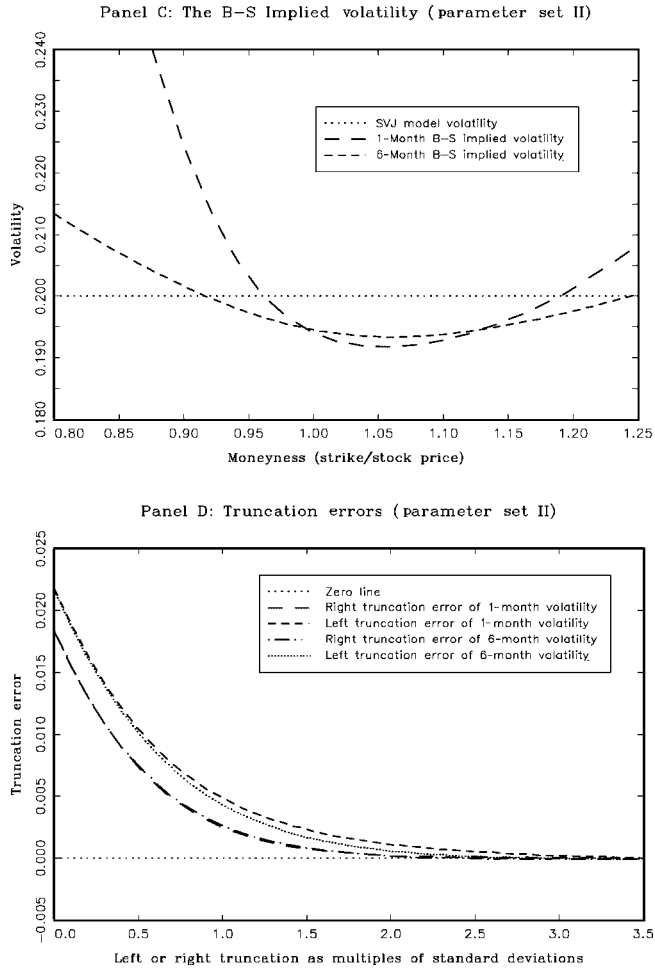

Figure 1
Continued

slower rate than the right truncation error does. With a fatter left tail, a larger range of strike prices is needed on the left of F_0 if we wish to have identical truncation errors from both sides. Nevertheless, the differences between left and right truncation errors are relatively small and convergence is rapid in all cases. In general, the truncation error is negligible if the truncation points are more than two SDs from F_0.

Figure 1 also shows that the model-free implied volatilities calculated from one-month and six-month maturities have similar truncation errors. With a longer maturity, however, the same multiple of SDs translates into a larger range of strike prices. This is because integrated

variance is an increasing function of maturity. As a result, a larger range of strike prices are needed to control the truncation errors for longer maturities.

We also calculate the upper bounds for truncations errors [the RHS of Equations (3) and (4)] for both the one-month and six-month maturities. In general, the difference between the upper bound and the true truncation error is small. In particular, the upper bound is almost indistinguishable from the true truncation error when the truncation point is more than two SDs from F_0. For this reason, the upper bounds are not plotted in Figure 1.

1.2.2 Discretization errors. In addition to truncation errors, the implementation of the model-free implied volatility also involves discretization errors due to numerical integration. Consider the numerical integration of Equation (2) using the trapezoidal rule:

$$2\int_{K_{\min}}^{K_{\max}} \frac{C^F(T,\,K)-\max(0,\,F_0-K)}{K^2}dK \approx \sum_{i=1}^{m}[g(T,\,K_i)+g(T,\,K_{i-1})]\Delta K, \tag{6}$$

where $\Delta K = (K_{\max}-K_{\min})/m$, $K_i = K_{\min} + i\Delta K$ for $0 \le i \le m$, and $g(T,K_i) = [C^F(T,K_i) - \max(0,F_0 - K_i)]/K_i^2$. For a finite m (or $\Delta K > 0$), the numerical integration scheme in Equation (6) leads to discretization errors.

Figure 2 illustrates typical discretization errors at various levels of ΔK We set the truncation points at 3.5 SDs from F_0. As shown in Figure 1, truncation errors are virtually zero beyond 3.5 SDs. We use the second set of parameter values to plot Figure 2 since it exhibits a more severe volatility smile or smirk pattern. Discretization errors are smaller if we use the first set of parameter values. In order to provide a useful benchmark for discretization errors, we measure ΔK in SD units. As shown in Figure 2, discretization errors diminish quickly as ΔK decreases. For both one-month and six-month maturities, discretization errors are negligible when $\Delta K \le 0.35$ SDs (or $m \ge 20$). With the initial asset price at \$100 and the annualized volatility of 0.2, 0.35 SD translates into a strike price increment of roughly \$2 and \$5 for one-month and six-month options, respectively. This is generally consistent with actual strike price increments in the marketplace. Since we use a much smaller ΔK (or larger m) in our empirical implementation, discretization errors are unlikely to have any impact on the calculation of the model-free implied volatility.

1.2.3 Spot prices versus forward prices. Equation (1) is stated in forward prices. When applications require the use of spot prices, appropriate

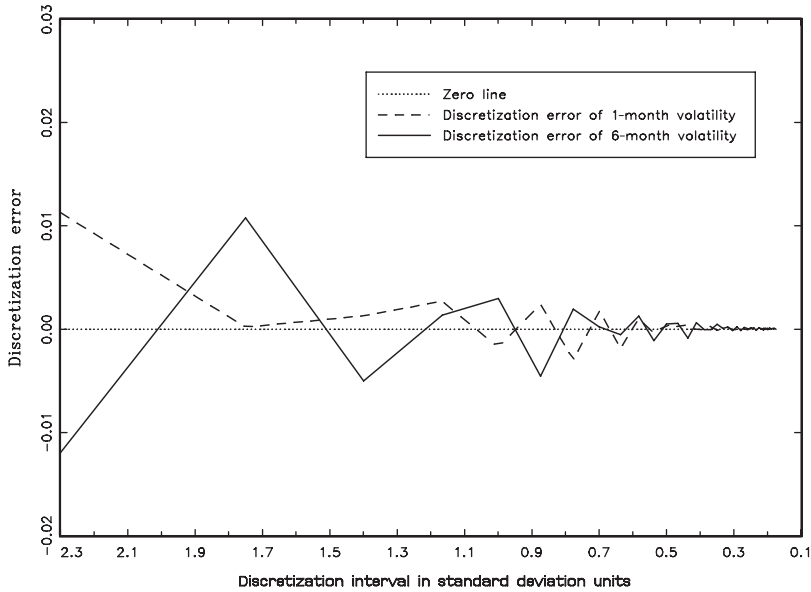

Figure 2
Discretization errors in calculating the model-free implied volatility
The discretization errors of the model-free implied volatility are plotted against strike price increment (ΔK, measured in standard deviation [SD] units) for options with one- and six-month maturities. The SVJ model is assumed for the underlying asset price: $dF_t/F_t = V_t^{1/2}dW_t + J_t dN_t - \mu_J \lambda dt$, $dV_t = (\theta_v - k_v V_t)dt + \sigma_v V_t^{1/2} dW_t^v$, and $dW_t dW_t^v = \rho dt$ where $N_t \sim$ i.i.d. Poisson (λ) and $\ln(1 + J_t) \sim$ iid $\mathrm{N}\left[\ln(1 + \mu_J) - \frac{1}{2}\sigma_J^2, \sigma_J\right]$. The second set of parameters with a more pronounced volatility skew is used: $k_v = 1$, $\sigma_v = 0.25$, $\rho = 0$, $\lambda = 0.5$, $\mu_J = -0.075$, $\mu_J = 0.075$, $\theta_v = 0.1854^2 k_v$, $V_0 = \theta/k_v$. The model-free implied volatility is calculated using the trapezoidal integration method. To minimize truncation errors, we choose truncation points sufficiently far from the forward price (3.5 SDs on both sides of the forward price F_0). The discretization error is the difference between the calculated volatility and the true volatility.

adjustments are needed. Under the assumption of deterministic interest rates and dividends, we write the forward asset price and forward option price at time t as $F_t = S_t/B(t, T)$ and $C^F(T, K) = C(T, K)/B(t, T)$, respectively, where S_t is the spot asset price minus the present value of all expected future dividends to be paid prior to the option maturity, $B(t, T)$ is the time t price of a zero-coupon bond that pays \$1 at time T, and $C(T,K)$ is the spot option price. By substitution, Equation (1) changes to:

$$E_0^F\left[\int_0^T \left(\frac{dS_t}{S_t}\right)^2\right] = 2\int_0^\infty \frac{C(T,K)/B(0,T) - \max[0, S_0/B(0,T) - K]}{K^2} dK.$$

Through a change of variables, the above equation is restated as:

$$E_0^F\left[\int_0^T \left(\frac{dS_t}{S_t}\right)^2\right] = 2\int_0^\infty \frac{C[T,K/B(0,T)] - \max(0, S_0 - K)}{K^2} dK. \quad (7)$$

Equation (2) provides an alternative definition of the model-free implied variance in spot prices.

1.2.4 Limited availability of strike prices. Up to this point, we have assumed that prices are available for options with any strike price in a given range $[K_{\min}, K_{\max}]$. This is, however, not realistic because only strike prices at a fixed increment are listed for trading in the marketplace. We must deal with not only a limited range but also a sparse set of discrete strike prices. Following prior research, we apply a curve-fitting method to interpolate between available strike prices. To illustrate this method, we consider the strike price structure of SPX options on September 23, 1988 and use it as a prototype.[3] On this date, the closing index value is approximately 270 (S_0) and the listed strike prices (K) are 200, 220 to 310 with a 5-point increment, 325 and 350.

We continue to use the SVJ model with the second set of parameter values as described previously. Without loss of generality, interest rates and dividends are assumed to be zero. We calculate the model prices of call options with the listed strike prices and a specified maturity (T). We use model prices (instead of market prices) as inputs in the implementation of the model-free implied volatility so that the true volatility is known and the approximation error due to our numerical implementation can be accurately measured.

To calculate the model-free implied volatility, we need to evaluate the RHS of Equation (6). This requires prices of call options with strike prices K_i for $0 \leq i \leq m$. Because these options are not listed on our sample date, their prices are not directly available (even though we could calculate their prices using the SVJ model). Instead, their prices must be inferred from the prices of listed options. Among the approaches used in previous research, the curve-fitting method is the most practical and effective. Although some studies apply the curve-fitting method directly to option prices [e.g., Bates (1991)], the severely nonlinear relationship between option price and strike price often leads to numerical difficulties. Following Shimko (1993) and Aït-Sahalia and Lo (1998), we apply the curve-fitting method to implied volatilities instead of option prices. Prices of listed calls are first translated into implied volatilities using the B–S model. A smooth function is then fitted to the implied volatilities. We then extract implied volatilities at strike prices K_i from the fitted function. The B–S model is used once more to translate the extracted implied volatilities into call prices. Note that this curve-fitting procedure does not make the assumption that the B–S model is the true model underlying option prices. It

[3] The choice of the sampling date is not important because we do not use observed option prices in our implementation here. We only need the set of strike prices listed on that date to provide a sense of availability of strike prices in the marketplace.

is merely used as a tool to provide a one-to-one mapping between option prices and implied volatilities. With the extracted call prices, the model-free implied volatility is calculated using the RHS of Equation (6). Following Bates (1991) and Campa, Chang, and Reider (1998), we use cubic splines in the curve-fitting of implied volatilities. Using cubic splines has the advantage that the obtained volatility function is smooth everywhere and provides an exact fit to the known implied volatilities.[4]

It is also important to note that the curve-fitting method is only effective for extracting option prices between the maximum and minimum available strike prices. For options with strike prices beyond the available range, we use the endpoint implied volatility to extrapolate their option values. In other words, the volatility function is assumed to be constant beyond the maximum and minimum strike prices. Extrapolation may be necessary in empirical applications as the range of available strike prices may not be sufficiently large on all trading days. Of course, this extrapolation procedure introduces an approximation error that is different from the truncation error due to Equation (2). Whereas the truncation method ignores options with strike prices beyond the available range, the extrapolation method incorporates these options by approximating their prices using the implied volatility at the endpoint strike price (either $K_{\min}$ or $K_{\max}$).

Table 1 presents the approximation errors from both the truncation method (panel A) and the extrapolation method (panel B) for various maturities (T) and truncation intervals ($[K_{\min}, K_{\max}]$). The available strike price range of [200, 350] on September 23, 1988 is also included in the table, which is equivalent to a truncation interval of $[0.74S_0, 1.30S_0]$. The discretization parameter (m) is fixed at 100. The approximation error is calculated as the estimated annualized volatility minus the true volatility. As expected, the magnitude of the approximation error from both methods is positively related to maturity and negatively related to truncation interval. In addition, the extrapolation method is in general more accurate than the truncation method. Consider for example the longest maturity of 180 days and the smallest truncation interval of $[0.9S_0,1.1S_0]$. The approximation error from the truncation method is –0.0325 or 16.3% of the true volatility of 0.2. The large error is not surprising because the truncation points are roughly 0.7 SD from the initial asset price. From the results in Figure 1, we know that the truncation error may not diminish until the truncation interval is at least plus/minus 2 SDs. In comparison, the approximation error from the extrapolation method is much smaller. As shown in panel B, the corresponding error is only –0.0033 or 1.7% of the true volatility. Additional calculations show that similar results are

[4] For a more detailed review of curve-fitting and other methods, see the survey article by Jackwerth (1999).

Table 1
Approximation errors in the calculation of the mode-free implied volatility

		Maturity (days)						
$K_{\min}$	$K_{\max}$	30	45	60	75	90	120	180
Panel A: the truncation method								
$0.9S_0$	$1.1S_0$	−0.0048	−0.0079	−0.0113	−0.0146	−0.0177	−0.0234	−0.0325
$0.8S_0$	$0.2S_0$	0.0001	−0.0001	−0.0005	−0.0011	−0.0018	−0.0034	−0.0072
200	350	0.0000	−0.0002	−0.0005	−0.0010	−0.0014	−0.0027	−0.0056
$0.7S_0$	$1.3S_0$	0.0001	0.0001	−0.0000	−0.0001	−0.0002	−0.0006	−0.0018
$0.6S_0$	$1.4S_0$	0.0001	0.0000	−0.0000	−0.0001	−0.0002	−0.0004	−0.0010
Panel B: the extrapolation method								
$0.9S_0$	$1.1S_0$	−0.0013	−0.0015	−0.0017	−0.0019	−0.0021	−0.0026	−0.0033
$0.8S_0$	$1.2S_0$	0.0005	0.0003	0.0001	0.0000	−0.0001	−0.0004	−0.0011
200	350	0.0005	0.0003	0.0002	0.0001	−0.0000	−0.0003	−0.0008
$0.7S_0$	$1.3S_0$	0.0005	0.0003	0.0002	0.0001	−0.0000	−0.0003	−0.0008
$0.6S_0$	$1.4S_0$	0.0005	0.0003	0.0002	0.0001	−0.0000	−0.0003	−0.0008

The approximation error is defined as the estimated annualized volatility minus the true volatility (0.2). The SVJ model is implemented with the following parameters: $k_v = 1$, $\sigma_v = 0.25$, $\rho = 0$, $\lambda = 0.5$, $\mu_J = -0.075$, $\sigma_J = 0.075$, $\theta_v = 0.1854^2$ k_v, $V_0 = \theta_v/k_v$. The current asset price (S_0) is 270 and the listed strike prices range from 200 to 350, based on options listed on the Standard & Poor's (S & P 's) 500 index on September 23, 1988. The discretization parameter (m) used is 100.

obtained over a wide range of parameters for the SVJ model. The accuracy and robustness of our extrapolation method is thus assured.

As shown in Table 1, the extrapolation method (panel B) leads to smaller approximation errors than the truncation method (panel A) does. The truncation method ignores strike prices beyond the truncation interval, which leads to the underestimation of the true volatility. The extrapolation method corrects the underestimation error by assuming a flat implied volatility function beyond the truncation point. In other words, implied volatilities outside the truncation interval are extrapolated from the implied volatility at the respective truncation point. We further argue that this correction mechanism in the extrapolation method is likely to lead to smaller approximation errors in most empirical settings. Consider, for example, the three most commonly observed shapes of the implied volatility function: smile, smirk, and skew. With a volatility smile, the implied volatility rises as the strike price increases or decreases from the asset price. By extrapolating on the implied volatility at the truncation points, the extrapolation method underestimates the true volatility beyond the truncation interval. The underestimation is, however, less severe than that in the truncation method, leading to a more accurate approximation. With a volatility smirk, the implied volatility rises faster at low strike prices than at high strike prices. The extrapolation method is again more accurate than the truncation method. In this case, however, extrapolation creates a more severe kink at the left truncation

point, which may lead to larger approximation errors if $K_{\min}$ is too close to S_0. Numerical examples in Table 1 show that the extrapolation method is, nevertheless, reasonably accurate even in the presence of a severe volatility smirk. Finally, we consider a volatility skew. A typical volatility skew exhibits high implied volatility for low strike prices and low implied volatility for high strike prices. This means that the extrapolation method underestimates the volatility beyond the minimum strike price but overestimates it beyond the maximum strike price. As the two effects offset or partially offset each other, the extrapolation method should be more accurate than the truncation method. Consequently, we adopt the extrapolation method in our subsequent empirical tests.

2. Data and Volatility Calculation

Data used in this study are from several sources. Intraday data on SPX options are obtained from the CBOE. Daily cash dividends are obtained from the S & P's DRI database. High-frequency data at 5-minute and 30-minute intervals for the SPX are extracted from the contemporaneous index levels recorded with the quotes of SPX options. In addition, daily Treasury bill yields (our risk-free rates) are obtained from the Federal Reserve Bulletin. Our sample period is from June 1988 to December 1994. Following Harvey and Whaley (1991, 1992a, 1992b), we use daily cash dividends instead of constant dividend yield. We use mid bid-ask quotes instead of actual transaction prices in order to avoid the bid-ask bounce problem [Bakshi, Cao, and Chen (1997, 2000)].

To select our final sample of SPX options, several data filters are applied. First, option quotes less than 3/8 are excluded from the sample. These prices may not reflect true option value due to proximity to tick size. Second, options with less than a week remaining to maturity are excluded from the sample. These options may have liquidity and market microstructure concerns. Third, option quotes that are time-stamped later than 3:00 P.M. (Central Standard Time) are excluded. As the stock market closes at 3:00 P.M., these option prices cannot be matched with simultaneously observed index values. Next, following Aït-Sahalia and Lo (1998), we exclude in-the-money options from the sample. In-the-money options are more expensive and often less liquid than at-the-money or out-of-the-money options. We define in-the-money options as call options with strike prices less than 97% of the asset price and put options with strike prices more than 103% of the asset price. In addition, options violating the boundary conditions are eliminated from the sample. These options are significantly undervalued and the B–S implied volatility is in fact negative for these options. Finally, only option quotes from 2:00 to 3:00 P.M. are included in the sample. To reduce computational burden, we construct one implied volatility surface for

each day in our sample. As option quotes over a range of strike prices and maturities are needed, we use all option quotes in the final trading hour before the stock market closes. The last trading hour is used because trading volume is typically higher than at other times of the day.

Our empirical tests require the construction of implied volatility surface from available option prices. In theory, this is straightforward to do if the set of traded options covers a sufficient range of strike prices and maturities. For each available option, we use the B–S model to back out the implied volatility from option quote. If option quotes are available for all strike prices and maturities, the calculated implied volatilities naturally form a surface. However, only a limited number of strike prices and maturities are listed for trading. A curve-fitting method is needed to extract the implied volatility surface from available option prices. To fit a smooth surface to the available B–S implied volatilities, the curve-fitting method is implemented in two steps. Cubic splines are first applied in the moneyness dimension to fit a smooth curve to the B–S implied volatilities as in Section 1.2.4. That is, implied volatility as a smooth function of moneyness (x), $\hat{\sigma}(x|T)$, is obtained for each option maturity (T). In the second step, we fix moneyness and apply cubic splines in the maturity dimension. The result is a smooth implied volatility surface, $\bar{\sigma}(x,T)$.

To avoid the telescoping overlap problem described by Christensen, Hansen, and Prabhala (2001), we extract implied volatilities from the implied volatility surface at predetermined fixed maturity intervals. The model-free implied volatility is then calculated using the RHS of Equation (6) for fixed maturities of 30, 60, 120, and 180 (calendar) days. The B–S implied volatility is extracted directly from the implied volatility surface for the same fixed maturities and moneyness levels from 0.94 to 1.06.

In addition to implied volatilities, our empirical analysis also requires monthly series of realized volatility and historical volatility. Realized volatility is needed to assess the forecasting ability and information content of implied volatilities. We calculate the realized volatility over periods of 30 to 180 days, matching the maturities of the corresponding implied volatilities. Historical volatility serves as a competing forecast, to implied volatilities, for future realized volatility. We use the realized volatility on the latest trading day as a proxy for historical volatility since the latest day's volatility is likely to contain the most relevant information for future volatility. In other words, the lagged daily realized volatility is our proxy for historical volatility.

Previous research adopts different sampling frequencies of asset returns for the calculation of realized volatility. Some earlier studies such as Canina and Figlewski (1993) and Christensen and Prabhala (1998) calculated these volatilities using daily returns. More recent studies argue that realized volatility should be calculated using intraday returns instead of

daily returns. For example, Andersen and Bollerslev (1998), Andersen, Bollerslev, Diebold, and Labys (2001, 2003) and Barndorff-Nielsen and Shephard (2003) find that there is considerable advantage in using high-frequency data over daily data in estimating realized volatility. In particular, Andersen and Bollerslev (1998) show that the typical squared returns method for calculating realized volatility produces inaccurate forecasts if daily returns are used. The inaccuracy is a result of noise in daily returns. They further show that the impact of the noise component is diminished if high-frequency returns (e.g., 5-minute returns) are used. These studies, however, draw their conclusions using evidence from foreign exchange rates. Intraday data on foreign exchange rates are relatively well behaved and exhibit little or no autocorrelation. This is not true for stock or stock index returns. Stoll and Whaley (1990) show that 5-minute returns on the SPX during the sample period from July 23, 1984 to December 31, 1986 have a first-order autocorrelation of 0.45 and a second-order autocorrelation of 0.14. The positive correlation is mainly due to the infrequent trading of the stocks in the index. As market liquidity tends to improve over time, we should expect this positive correlation to become smaller in more recent period. Indeed, in our sample period from June 1, 1988 to December 31, 1994, the first-order autocorrelation drops to 0.31 while the second-order autocorrelation drops to 0.04. Nevertheless, the correlation is still significant and a naive estimate of realized volatility based on 5-minute returns is likely to be downward biased. In the extreme case when returns are sampled continuously, Zhang, Mykland, and, Aït-Sahalia (2003) show that the standard realized volatility estimator captures the volatility of the market microstructure noise rather than the true integrated volatility if the noise is not accounted for.

The optimal sampling frequency of intraday returns and the appropriate procedure for dealing with the related market microstructure problem are the subject of several recent studies. Aït-Sahalia, Mykland, and Zhang (2005) demonstrate that more data does not necessarily lead to a better estimate of realized volatility in the presence of market microstructure noise. They show that the optimal sampling frequency is jointly determined by the magnitude of the market microstructure noise and the horizon of the realized volatility. For a given level of noise, the realized volatility for a longer horizon (e.g., one month) should be estimated with less frequent sampling than the realized volatility for a shorter horizon (e.g., one day). In addition, they also show that correcting for market microstructure noise may restore the first-order statistical effect that it is optimal to sample as often as possible.

Consistent with the findings from Aït-Sahalia, Mykland, and Zhang (2003), we use 30-minute index returns to calculate the realize volatility over one month or longer horizons and 5-minute index returns to calculate

the lagged daily realized volatility (our proxy for historical volatility). To correct for the bias in estimated realized volatility due to autocorrelation in intraday returns, we adopt a correction method suggested, in various forms, by French, Schwert, and Stambaugh (1987), Zhou (1996), and Hansen and Lunde (2004). In this correction method, the annualized realized variance over the period $[t, t + \tau]$ is calculated as:

$$V_{t,\tau} = \frac{1}{\tau}\sum_{i=1}^{n} R_i^2 + \frac{2}{\tau}\sum_{h=1}^{l}\left(\frac{n}{n-h}\right)\sum_{i=1}^{n-h} R_i R_{i+h}, \tag{8}$$

where R_i is the index return during the i-th interval, n is the total number of intervals in the period, and l is the number of correction terms included.[5] When the volatility (e.g, the lagged daily realized volatility) is calculated with 5-minute returns, we use Equation (8) with one correction term (i.e., $l = 1$). This is because 5-minute returns in our sample have a first-order autocorrelation of 0.31 while higher-order autocorrelations are much smaller. In contrast, we do not correct for autocorrelation if volatility (e.g., the one-month realized volatility) is calculated with 30-minute returns (i.e., $l = 0$). These less frequent returns have a first-order autocorrelation of only 0.06 in our sample while higher-order autocorrelations are all negligible. Although all reported results in this study are based on realized volatility calculated this way, we do consider alternative correction terms subsequently in robustness tests and the results are not materially affected.

Table 2 provides summary statistics for the model-free implied volatility, the B–S implied volatility and the realized volatility. All volatility measures are estimated monthly on the Wednesday immediately following the expiry date of the month. For the B–S model, implied volatility is reported for at-the-money options only. For other moneyness levels, the results are similar but exhibit the well-known smile pattern. The realized volatility is calculated using 30-minute index returns for the period matching the maturity of the corresponding option(s) in the implied volatility calculation. As shown in Table 2, both the B–S and model-free implied volatilities are on average higher than the realized volatility over all horizons. These implied volatilities are thus likely biased forecast for realized volatility, with a slightly larger bias for the latter. The larger bias from the model-free implied volatility is expected as the B–S implied volatility is known to be a downward biased measure of risk-neutral expected variance under stochastic volatility by Jensen's inequality. By correcting this downward bias, the model-free implied volatility tends to

[5] Dividends are ignored in the above calculation since daily dividends on the SPX are typically small and much less volatile than the index itself. Unreported results confirm that adjustment for dividends does not lead to material change to realized volatility estimates.

Table 2
Summary Statistics of monthly volatility series

τ	N	Mean	Standard deviation	Skewness	Excess kurtosis	Minimum	Maximum
Panel A: σ^{BS}							
30	78	14.78	4.396	1.1258	2.0636	6.912	31.21
60	79	15.71	4.399	1.0210	0.9198	8.325	29.90
120	78	16.82	4.254	0.8038	0.6587	8.761	30.81
180	78	17.31	3.844	0.7213	0.3137	10.62	29.21
Panel B: σ^{RE}							
30	79	13.96	4.078	1.4165	2.9300	6.953	28.78
60	79	14.08	3.751	1.2724	2.1930	7.350	26.92
120	79	14.10	3.682	1.3448	2.2043	8.248	26.54
180	79	14.13	3.241	1.1327	0.9593	9.208	25.32
Panel C: σ^{MF}							
30	78	15.68	4.450	1.5966	3.8305	8.589	34.61
60	79	16.11	3.964	1.1685	1.3999	10.16	28.72
120	78	17.51	3.976	0.7028	0.1716	10.06	28.24
180	78	17.33	3.566	0.4522	−0.1985	9.666	25.80
Panel D: $\ln(\sigma^{BS})$							
30	78	2.653	0.284	0.1563	0.2628	1.933	3.441
60	79	2.719	0.264	0.3500	−0.1444	2.119	3.397
120	78	2.792	0.246	0.1050	0.0401	2.170	3.427
180	78	2.828	0.216	0.1901	−0.3050	2.363	3.374
Panel E: $\ln(\sigma^{RE})$							
30	79	2.596	0.272	0.3959	0.6496	1.938	3.360
60	79	2.612	0.255	0.2439	−0.0728	1.995	3.292
120	79	2.621	0.244	0.5847	0.3935	2.110	3.278
180	79	2.608	0.217	0.6220	−0.0051	2.220	3.154
Panel F: $\ln(\sigma^{MF})$							
30	78	2.718	0.255	0.6607	0.6865	2.150	3.544
60	79	2.752	0.229	0.5321	0.1377	2.318	3.357
120	78	2.838	0.221	0.1528	−0.2150	2.308	3.341
180	78	2.832	0.205	−0.0519	−0.2103	2.368	3.250

σ^{BS}, σ^{RE}, and σ^{MF} are the at-the-money Black–Scholes (B–S) implied volatility, the realized volatility, and the model-free implied volatility, respectively. τ is the time horizon and N is the sample size. All volatilities are expressed in annualized percentage terms.

be higher than the B–S implied volatility. In addition, the reported skewness and excess kurtosis in Table 2 reveal that the log volatility is the most conformable with the normal distribution. Regressions based on the log volatility are thus statistically better specified than those based on volatility or variance.

Table 3 summarizes the correlation matrix of the monthly 30-day volatility series for the B–S implied volatility (from options with different moneyness levels), the model-free implied volatility, and the realized volatility. Correlations for monthly volatility series with longer maturities have similar properties and are not reported. Overall, all B–S implied volatility series are highly correlated with the model-free implied volatility. As expected, the B–S implied volatility from at-the-money options has the highest correlation with the model-free implied volatility (94.1%). While all implied volatility series are highly correlated with the corresponding realized volatility, the model-free implied volatility has the

Table 3
Correlation matrix of monthly 30-day volatility series

	$\sigma^{BS}(0.94)$	$\sigma^{BS}(0.97)$	$\sigma^{BS}(1.00)$	$\sigma^{BS}(1.03)$	σ^{MF}	σ^{RE}
Panel A: correlation matrix of volatility ($N = 60$)						
$\sigma^{BS}(0.94)$	1.000	—	—	—	—	—
$\sigma^{BS}(0.97)$	0.879	1.000	—	—	—	—
$\sigma^{BS}(1.00)$	0.861	0.887	1.000	—	—	—
$\sigma^{BS}(1.03)$	0.864	0.872	0.898	1.000	—	—
σ^{MF}	0.922	0.927	0.941	0.936	1.000	—
σ^{RE}	0.767	0.778	0.814	0.803	0.857	1.000
Panel B: correlation matrix of log volatility ($N = 60$)	$\ln(\sigma^{BS})$ (0.94)	$\ln(\sigma^{BS})$ (0.97)	$\ln(\sigma^{BS})(1.00)$	$\ln(\sigma^{BS})(1.03)$	$\ln(\sigma^{MF})$	$\ln(\sigma^{RE})$
$\ln(\sigma^{BS})$ (0.94)	1.000	—	—	—	—	—
$\ln(\sigma^{BS})$ (0.97)	0.847	1.000	—	—	—	—
$\ln(\sigma^{BS})(1.00)$	0.813	0.852	1.000	—	—	—
$\ln(\sigma^{BS})(1.03)$	0.828	0.829	0.872	1.000	—	—
$\ln(\sigma^{MF})$	0.889	0.903	0.918	0.916	1.000	—
$\ln(\sigma^{RE})$	0.721	0.755	0.764	0.771	0.843	1.000

σ^{BS}, σ^{RE}, and σ^{MF} are the Black–Scholes (B–S) implied volatility, the realized volatility and the model-free implied volatility, respectively. Numbers in parentheses after σ^{BS} are option moneyness. The $\sigma^{BS}(1.06)$ is excluded in the calculation of correlation matrix as it has fewer observations.

highest correlation (85.7%), followed by the at-the-money B–S implied volatility (81.4%). We also test and reject the null hypothesis that the B–S implied volatility is an unbiased estimator of the model-free implied volatility at the 1% significance level. Figure 3 plots the two implied volatility time series over our sample period. For the B–S implied volatility, we only illustrate the at-the-money series. It is clear that the two volatility series track one another fairly closely. The B–S implied volatility tends to be lower than the model-free implied volatility and exhibits greater local fluctuations than the latter does. Overall, these results imply that while both the B–S and model-free implied volatilities have similar features, the two implied volatility series do contain sufficiently different information content.

3. The Information Content of Implied Volatility

Prior research has extensively examined the information content of the B–S implied volatility. While early studies produce mixed results, recent empirical studies seem to agree that the B–S implied volatility is a more efficient forecast for future realized volatility than historical volatility. They also find that the B–S implied volatility does not subsume all information contained in historical volatility and is an (upward) biased forecast for future realized volatility. However, these studies typically adopt a point estimate of the B–S implied volatility obtained from the price of a single option or prices of several similar options. Since information contained in other options is discarded, tests based on such volatility

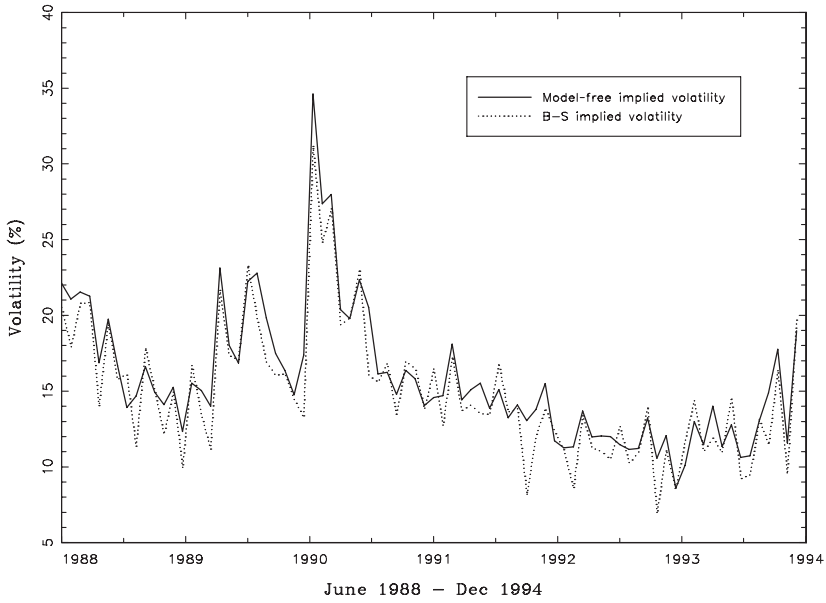

Figure 3
Time-series plot of the model-free and at-the-money Black–Scholes (B–S) implied volatilities
The figure plots the time series of the 30-day model-free and at-the-money B–S implied volatilities over the sample period from June 1988 to December 1994. The two implied volatilities are calculated using the implied volatility surface. The implied volatility surface is constructed by interpolating implied volatilities of traded options using cubic splines, first in the moneyness dimension and then in the maturity dimension.

measures are likely biased towards rejecting informational efficiency. In this study, we examine the information content of the model-free implied volatility. Because it aggregates information from options across all strike prices, the model-free implied volatility should be informationally more efficient. To nest previous research within our framework, we compare and contrast three competing volatility forecasts—the model-free implied volatility, the B–S implied volatility, and the historical volatility.

Following Christensen and Prabhala (1998) and Christensen, Hansen, and Prabhala (2001), we test our null hypotheses using monthly nonoverlapping samples. This avoids the so-called telescoping overlap problem which may render the *t*-statistics and other diagnostic statistics in the regression analysis invalid. To extract our monthly nonoverlapping sample, we choose the Wednesday immediately following the expiration date of the month. If it happens to be a nontrading day, we go to the following Thursday then the proceeding Tuesday. We select the monthly sample this way because option trading seems to be more active during the week following the expiration date and Wednesday has the fewest holidays among all weekdays. When calculating the monthly volatility measures,

we fix the option maturity at 30 calendar days. Time series obtained this way are neither overlapping nor telescoping.

Consistent with prior research [e.g., Canina and Figlewski (1993) and Christensen and Prabhala (1998)], we employ both univariate and encompassing regressions to analyze the information content of volatility forecasts. In a univariate regression, the realized volatility is regressed against a single volatility forecast. In comparison, two or more volatility forecasts are used as explanatory variables in encompassing regressions. While the univariate regression focuses on the forecasting ability and information content of one volatility forecast, the encompassing regression addresses the relative importance of competing volatility forecasts and whether one volatility forecast subsumes all information contained in other volatility forecast(s). Since a univariate regression is a restricted version of the corresponding encompassing regression, we only state the encompassing regressions. These regressions are specified as follows, in three alternative specifications:

$$\sigma_{t,\tau}^{RE} = \alpha_{x,\tau} + \beta_{x,\tau}^{MF}\sigma_{t,\tau}^{MF} + \beta_{x,\tau}^{BS}\sigma_{t,x,\tau}^{BS} + \beta_{x,\tau}^{LRE}\sigma_{t,\tau}^{LRE} + \varepsilon_{t,x,\tau}, \tag{9}$$

$$V_{t,\tau}^{RE} = \alpha_{x,\tau} + \beta_{x,\tau}^{MF}V_{t,\tau}^{MF} + \beta_{x,\tau}^{BS}V_{t,x,\tau}^{BS} + \beta_{x,\tau}^{LRE}V_{t,\tau}^{LRE} + \varepsilon_{t,x,\tau}, \tag{10}$$

$$\ln\,\sigma_{t,\tau}^{RE} = \alpha_{x,\tau} + \beta_{x,\tau}^{MF}\,\ln\,\sigma_{t,\tau}^{MF} + \beta_{x,\tau}^{BS}\,\ln\,\sigma_{t,x,\tau}^{BS} + \beta_{x,\tau}^{LRE}\ln\,\sigma_{t,\tau}^{LRE} + \varepsilon_{t,x,\tau}, \tag{11}$$

where σ and V are asset return volatility and variance, respectively. The superscripts *RE*, *MF*, *BS*, and *LRE* stand for REalized, Model-Free, Black–Scholes, and Lagged REalized, respectively. The subscripts t, x, and τ are observation date, moneyness, and maturity, respectively. Univariate regressions are obtained if two of the three regressors are dropped.

Lagged realized volatility σ^{LRE} is our proxy for historical volatility. Previous studies [e.g., Canina and Figlewski (1993) and Christensen and Prabhala (1998)] adopted the realized volatility over a matching period (30 calendar days here) immediately proceeding the current observation date as the lagged realized volatility. As the volatility process is likely a Markov process, the latest day's realized volatility is the most relevant for forecasting future volatility. Following this logic, we adopt the realized volatility on the latest trading day as the lagged realized volatility. This lagged realized volatility is calculated from 5-minute index returns using Equation (8) with one correction term. The robustness of our results with respect to alternative correction terms and other proxies for historical volatility is examined subsequently.

Table 4 summarizes the OLS regression results from both univariate and encompassing regressions using the monthly nonoverlapping sample described previously. Results from the three specifications are presented in separate panels in the table. While the estimate for the other two volatility measures is only a function of maturity, the estimate for the

Table 4
Univariate and encompassing regressions of 30-day volatility (OLS)

N	a	β^{MF}	β^{BS}	β^{LRE}	Adjusted R^2	Durbin-Watson	χ^2 test(a)	χ^2 test(b)
Panel A: σ_t								
78	2.49 (1.15)	—	0.77*** (0.080)	—	0.64	1.98	14.85 (0.000)	—
78	2.66 (1.16)	—	0.70*** (0.093)	0.06***** (0.027)	0.65	2.04	—	11.91 (0.003)
78	10.8 (0.80)	—	—	0.23*** (0.036)	0.15	1.65	204.6 (0.000)	—
78	1.30 (0.79)	0.80*** (0.053)	—	—	0.74	1.98	70.37 (0.000)	—
78	1.36 (0.78)	0.83 (0.116)	0.04 (0.123)	—	0.74	2.01	—	15.17 (0.001)
78	1.32 (0.80)	0.80*** (0.060)	—	0.01 (0.033)	0.74	2.00	—	15.12 (0.001)
78	1.39 (0.84)	0.84 (0.122)	0.05 (0.126)	0.01 (0.032)	0.74	2.02	—	16.21 (0.001)
Panel B: V_t								
78	32.1 (18.9)	—	0.75*** (0.063)	—	0.62	2.07	15.10 (0.000)	—
78	32.8 (19.1)	—	0.73*** (0.104)	0.02 (0.029)	0.62	2.08	—	10.68 (0.005)
78	176.2 (16.6)	—	—	0.15*** (0.037)	0.11	1.75	652.2 (0.000)	—
78	19.9 (13.4)	0.71*** (0.058)	—	—	0.73	1.95	67.86 (0.000)	—
78	18.8 (15.9)	0.69*** (0.119)	0.03 (0.152)	—	0.73	1.90	—	48.11 (0.000)
78	20.4 (14.4)	0.73*** (0.072)	—	0.02 (0.020)	0.73	1.91	—	49.32 (0.000)
78	19.3 (14.7)	0.71*** (0.121)	0.03 (0.152)	0.02 (0.025)	0.73	1.91	—	49.14 (0.000)
Panel C: In V_t								
78	0.55 (0.22)	—	0.77*** (0.080)	—	0.60	1.90	23.20 (0.000)	—
78	0.40 (0.21)	—	0.70*** (0.082)	0.14***** (0.051)	0.62	1.91	—	13.37 (0.001)
78	1.88 (0.18)	—	—	0.29*** (0.071)	0.20	1.51	111.4 (0.000)	—
78	0.10 (0.16)	0.92 (0.057)	—	—	0.75	1.94	71.21 (0.000)	—
78	0.11 (0.16)	0.99 (0.135)	0.09 (0.120)	—	0.74	1.96	—	2.262 (0.323)
78	0.09 (0.16)	0.89* (0.065)	—	0.04 (0.042)	0.75	1.97	—	2.541 (0.281)
78	0.09 (0.16)	0.95 (0.147)	0.07 (0.123)	0.03 (0.043)	0.74	1.98	—	3.625 (0.305)

The at-the-money Black–Scholes (B–S) implied volatility and one-day lagged historical volatility are used in the regressions. The numbers in parentheses beside the parameter estimates are the standard errors, which are computed following a robust procedure taking into account of the heteroscedast1icity [see White (1980)]. The χ^2 test(a) is for the joint hypothesis H_0: $\alpha = 0$ and $\beta^j = 1$ ($j = MF, BS, LRE$) in univariate regressions, and the χ^2 test(b) is for the joint hypothesis H_0: $\beta^{BS} = 1$ and $\beta^{LRE} = 0$ or H_0: $\beta^{MF} = 1$, and $\beta^{BS} = \beta^{LRE} = 0$ in encompassing regressions. The test statistics are reported with the p-values in the brackets beside the statistic.

*, **, and *** indicate that the leading term β coefficient of the regressionis significantly different from one at the 10%, 5%, and 1% level, and ****, *****, and ****** indicate that the β coefficient of the remaining terms is significantly different from zero at the 10%, 5%, and 1% level, respectively.

B–S implied volatility is a function of both maturity and moneyness. To keep it manageable, we only present the results for at-the-money options ($x = 1$) since the B–S implied volatility from these options has the highest correlation with the realized volatility (Table 3). Results for other moneyness levels are not materially different. Numbers in brackets beside the parameter estimates are the standard errors, which are estimated following a robust procedure taking into account of heteroscedasticity [White (1980)]. The Durbin-Watson (DW) statistic is not significantly different from two in most regressions, indicating that the regression residuals are not autocorrelated.

Similar to Christensen and Prabhala (1998), we formulate and test several testable hypotheses associated with the information content of volatility measures. We begin our discussion with the results from univariate regressions. First, if a volatility forecast contains no information about future volatility, the slope coefficient (β) for the volatility forecast should be zero. This leads to our first testable hypothesis H_0: $\beta = 0$. Results in Table 4 strongly reject this hypothesis. In all univariate regressions, the slope coefficient (β) is positive and significantly different from zero at all conventional significance levels. This implies that all three volatility measures contain important information for future volatility. Secondly, if a given volatility forecast is an unbiased estimator of future realized volatility, the slope coefficient (β) should be one and the intercept (α) should be zero. This testable hypothesis is formulated as a joint hypothesis H_0: $\alpha = 0$ and $\beta = 1$. A χ^2 test is used to examine this hypothesis, with test statistics reported in the column with the heading "χ^2 test(a)." Numbers in brackets beside the test statistics are p-values. The null hypothesis is strongly rejected by the χ^2 test for every volatility forecast in every regression specification. This result is not surprising because summary statistics in Table 2 indicate that both the model-free and B–S implied volatilities are on average greater than the realized volatility. The evidence is also consistent with the existing option pricing literature which documents that stochastic volatility is priced with a negative market price of risk (or equivalently a positive risk premium). The volatility implied from option prices is thus higher than their counterpart under the objective measure due to investor risk aversion.

Some notable differences also exist in univariate regressions across model specifications and volatility measures. For instance, the R^2 is the highest for the model-free implied volatility regressions ranging from 73 to 74% while it is the lowest for the lagged realized volatility regressions ranging from 11 to 20%. This evidence suggests that, among the three volatility measures, the model-free implied volatility explains the most of the variations in future realized volatility. In other words, the model-free implied volatility contains the most information among the three volatility measures while the lagged realized volatility the least.

The higher R^2 in the model-free implied volatility regression also implies that previous studies based on the B–S implied volatility have underestimated the information content of implied volatility. By aggregating information across options with different strike prices, the model-free implied volatility retains more information for future realized volatility than the B–S implied volatility does.

We next consider the results from encompassing regressions involving only the B–S implied volatility and the lagged realized volatility. In these regressions, we investigate the informational efficiency of the B–S implied volatility relative to the lagged realized volatility. If the B–S implied volatility is informationally efficient, we should expect the lagged realized volatility to be statistically insignificant. This leads to the following null hypothesis H_0:$\beta^{LRE} = 0$. If this hypothesis is supported, the historical volatility of the underlying asset is redundant and its information content has been impounded in the B–S implied volatility. As shown in Table 4 (the second regression in each panel), the null hypothesis is rejected at the 5% level in two out of three specifications. Only the variance regression (panel B) fails to reject the null hypothesis. In addition, the informational efficiency of the B–S implied volatility can be formulated as a joint hypothesis H_0: $\beta^{BS} = 1$ and $\beta^{LRE} = 0$. It states that the B–S implied volatility is not only informationally efficient but also fully subsumes the information contained in the lagged realized volatility. Test statistics from a χ^2 test [column "χ^2 test(b)"] strongly reject this null hypothesis in all three specifications. These results, together with those from the univariate regressions, provide evidence that the B–S implied volatility is a biased forecast for future realized volatility and does not subsume all information contained in past realized volatility. This finding is consistent with previous research [e.g., Lamoureux and Lastrapes (1993), Jorion (1995), and Christensen and Prabhala (1998)].

Finally, we examine the results from encompassing regressions involving the model-free implied volatility. A total of nine encompassing regressions involving the model-free implied volatility are run and analyzed for alternative regression specifications and choices of volatility measures. Combining the results from all nine regressions (the last three regressions in each panel), we find strong evidence in support of the hypothesis that the model-free implied volatility subsumes all information contained in both the B–S implied volatility and the lagged realized volatility and is a more efficient forecast for future realized volatility. To begin with, once the model-free implied volatility is included in the regression, the addition of either the B–S implied volatility or the lagged realized volatility or both does not improve the regression goodness-of-fit (adjusted R^2) at all. This is clear from a direct comparison between univariate regressions and the related encompassing regressions involving

the model-free implied volatility. Secondly, in all nine encompassing regressions involving the model-free implied volatility, the t-statistics cannot reject the hypothesis that the slope coefficients for the B–S implied volatility and the lagged realized volatility are zero at any conventional significance levels. In fact, the estimated coefficients of the B–S implied volatility and the lagged realized volatility are all close to zero in the nine encompassing regressions. In addition, the t-test statistics in the log encompassing regressions cannot reject the hypothesis that the slope coefficient for the model-free implied volatility is one at the 5% significance level. Furthermore, the χ^2 test [column "χ^2 test(b)"] in the log encompassing regressions does not reject the null hypothesis that the slope coefficient is one for the model-free implied volatility and zero for all other slope coefficients. These results provide evidence that the model-free implied volatility fully subsumes information contained in both the B–S implied volatility and the lagged realized volatility and is a more efficient forecast for future realized volatility.

4. Robustness Tests

Results from the previous section provide strong support for the informational efficiency of the model-free implied volatility. Unlike the B–S implied volatility, the model-free implied volatility extracts information from options across all available strike prices. By aggregating information contained in individual options, the model-free implied volatility exhibits superior forecasting ability and is informationally more efficient. These findings are striking and provide new insight on volatility forecasting and market efficiency. We now conduct robustness tests to ensure the generality of our findings.

4.1 IV regressions

Christensen and Prabhala (1998) employ IV regressions to correct for potential EIV problems. They recognize that the B–S implied volatility (their volatility forecast) may contain significant measurement errors due to either the early exercise premium in the American style S & P 100 index options, the possible nonsynchronous observations of option quotes and index levels in their data set, or the misspecification error of the B–S model. It is well known that the EIV problem tends to drive the slope coefficient downward (biased toward zero). This may explain why the coefficient of the B–S implied volatility in their univariate and encompassing regressions is below one. Christensen and Prabhala (1998) found substantial differences between the estimation results from the OLS and IV regressions, supporting the existence of measurement errors in the B–S implied volatility.

Compared to Christensen and Prabhala (1998) and other previous studies, our data sample is arguably less prone to measurement errors for reasons discussed previously. In fact, our regressions involving the B–S implied volatility, both univariate and encompassing, have much higher R^2 than the corresponding figures in previous studies. Nevertheless, it is still important to investigate whether there is any EIV problem in the B–S and model-free implied volatilities in our sample and whether the IV regressions support the findings from the OLS regressions.

Following Christensen and Prabhala (1998), we apply a two-stage least squares regression to implement the IV estimation procedure. For the B–S implied volatility, we use the lagged realized volatility and lagged B–S implied volatility as IVs. Similarly, we use the lagged realized volatility, the lagged model-free implied volatility and the lagged B–S implied volatility as IVs for the model-free implied volatility. Table 5 reports the results from the IV regressions under the log specification as it is econometrically better specified than the other two specifications. Results from the OLS regressions in the first stage are reported in panel A while those from the second stage are reported in panel B.

Consider first the forecasting ability and information content of the B–S implied volatility. The IV regression results are found in the top half of panel B in Table 5. From the univariate regression, the joint hypothesis H_0: $\alpha = 0$ and $\beta^{BS} = 1$ is strongly rejected by the χ^2 test [column "χ^2 test(a)"], consistent with the result from the OLS regression reported in Table 4. This result confirms that the B–S implied volatility is a biased forecast for future realized volatility. In comparison, the results from the encompassing regression contradict the corresponding results from the OLS regression. The χ^2-statistic [column "χ^2 test(b)"] cannot reject the joint hypothesis H_0: $\beta^{BS} = 1$ and $\beta^{LRE} = 0$ at any conventional significance levels. This is evidence that the B–S implied volatility is informationally efficient and subsumes all information in lagged realized volatility. The *t*-statistics also support this conclusion. However, estimates of both coefficients have noticeably larger standard errors. The much increased standard errors make it more difficult to reject the null hypothesis. The Hausman (1978) χ^2-statistics (one degree of freedom) for testing the EIV problem are 6.58 and 6.65 with *p*-values of 0.0103 and 0.0099 for the two regressions, respectively. These test statistics strongly suggest the presence of measurement errors and the IV regressions are less efficient than the OLS regressions. Taken together, these results suggest that the B–S implied volatility is more efficient than historical volatility in forecasting future volatility but does not subsume all information contained in historical volatility. This is consistent with the findings in Christensen and Prabhala (1998) from the S & P 100 index options.

Next, we examine the forecasting ability and information content of the model-free implied volatility. The IV regression results are found in the

Table 5
Univariate and encompassing regressions of 30-day log volatility instrumental variable (IV)

N	α	β^{LBS} or β^{LMF}	β^{LRE}	β^{LBS}	Adjusted R^2	DW		
Panel A: first stage regression								
Dependent variable: in V_t^{BS}								
77	0.93 (0.22)	0.65 (0.082)	—	—	0.39	2.38	—	—
77	0.93 (0.23)	0.67 (0.107)	−0.03 (0.072)	—	0.39	2.33	—	—
Dependent variable: in V_t^{MF}								
77	0.67 (0.15)	0.75 (0.057)	—	—	0.56	2.25	—	—
77	0.68 (0.15)	0.79 (0.074)	−0.05 (0.065)	—	0.56	2.19	—	—
77	0.66 (0.16)	0.64 (0.150)	—	0.11 (0.156)	0.56	2.28	—	—
77	0.67 (0.16)	0.68 (0.165)	−0.05 (0.064)	0.12 (0.154)	0.56	2.22	—	—

N	α	β^{MF}	β^{BS}	β^{LRE}	adjusted R^2	DW	χ^2 test(a)	χ^2 test(b)
Panel B: second stage IV estimates								
Instrument variable for the B–S implied volatility								
77	0.08 (0.40)	—	0.94 (0.146)	—	0.35	1.90	9.855 (0.007)	—
77	0.21 (0.41)	—	0.79 (0.194)	0.10 (0.080)	0.36	2.05	—	2.474 (0.290)
Instrument variables for both model-free and B–S implied volatility								
77	0.06 (0.32)	0.93 (0.116)	—	—	0.58	1.94	31.13 (0.000)	—
77	0.19 (0.34)	1.10 (0.267)	–0.22 (0.288)	—	0.64	1.94	—	0.774 (0.679)
77	0.12 (0.35)	0.89 (0.173)	—	0.02 (0.082)	0.63	1.95	—	0.505 (0.776)
77	0.23 (0.35)	0.99 (0.294)	–0.16 (0.305)	0.04 (0.076)	0.64	1.94	—	1.180 (0.758)

The at-the-money Black–Scholes (B–S) implied volatility and one-day lagged historical volatility are used in the regressions. The numbers in parentheses beside the parameter estimates are the standard errors, which are computed following a robust procedure taking into account of the heteroscedastic and autocorrelated error structure [see Newey and West (1987)]. The χ^2 test(a) is for the joint hypothesis H_0:$\alpha = 0$ & $\beta^j = 1$ ($j = MF, BS, LRE$) in univariate regressions, and the χ^2 test(b) is for the joint hypothesis H_0: $\beta^{BS} = 1$ and $\beta^{LRE} = 0$ or H_0: $\beta^{MF} = 1$ and $\beta^{BS} = 1$ and $\beta^{LRE} = 0$ in encompassing regressions. The test statistics are reported with the p-values in the brackets beside the statistic. Durbin-Watson (DW) denotes the Durbin-Watson statistic.

bottom half of panel B in Table 5. In both univariate and encompassing regressions, we find no material change in statistical inferences between IV and OLS regressions. What is changed is the increased standard errors and reduced regression R^2. However, the magnitude of the change is relatively small compared to that in corresponding IV regressions associated with the B–S implied volatility. The Hausman (1978) χ^2 test also confirms that there is no evidence of any EIV problem in the model-free implied volatility. As a result, the IV regressions continue to support the hypothesis that the model-free implied volatility subsumes all information contained in both the B–S implied volatility and the lagged realized volatility. Our findings for the model-free implied volatility are thus robust to the estimation method used. In comparison, there is evidence that the B–S implied volatility contains measurement errors and these errors are likely due to model misspecification and the use of the B–S implied volatility from a single option.

4.2 Samples with longer maturities

In addition to the monthly nonoverlapping samples of 30-day options, we also extract monthly samples of volatility measures for options with a fixed maturity of 60, 120 or 180 days. We use the same monthly series of implied volatility surface constructed in Section 2 to extract the required implied volatility series. Again, the realized volatility is calculated using 30-minute index returns over the matching time horizon and the lagged realized volatility is calculated using 5-minute index returns on the latest trading day. The latter is corrected for autocorrelation using Equation (8) with one correction term.

The monthly samples obtained this way exhibit varying degrees of overlap. The longer the option maturity is, the more overlapped the sample becomes. We are interested in finding out whether or not regression results change if samples with longer horizons are used in the analysis instead of the nonoverlapping samples used previously. Overlapping samples may lead to severe serial correlation and render the OLS test statistics invalid. As pointed out by Richardson and Smith (1991), however, correct test statistics can be computed using the generalized method of moments (GMM). It is also important to note that our monthly overlapping samples do not have the telescoping overlap problem described by Christensen, Hansen and Prabhala (2001). Although these monthly samples are overlapping, they are not telescoping because the time to maturity rather than the time of maturity is fixed.

Table 6 summarizes the OLS results from the univariate and encompassing regressions for the log volatility measures over the 60-day, 120-day and 180-day maturity horizons. Because the log regression is econometrically better specified, we again concentrate on the log regressions in the analysis. As indicated by the DW statistics, there is an increased

Table 6
Univariate and encompassing regressions of 60-, 120- and 180-day log volatility (OLS)

N	α	β^{MF}	β^{BS}	β^{LRE}	Adjusted R^2	Durbin-Watson	χ^2 test(a)	χ^2 test(b)
Panel A: In V_t (τ = 60)								
79	0.51 (0.21)	—	0.76*** (0.076)	—	0.57	1.66	55.56 (0.000)	—
79	0.41 (0.22)	—	0.72*** (0.073)	0.09**** (0.051)	0.58	1.64	—	10.77 (0.004)
79	2.14 (0.20)	—	—	0.18*** (0.085)	0.10	1.09	129.7 (0.000)	—
79	0.03 (0.19)	0.93 (0.068)	—	—	0.70	1.61	114.6 (0.000)	—
79	0.03 (0.19)	0.87 (0.127)	0.07 (0.108)	—	0.70	1.62	—	1.313 (0.518)
79	0.01 (0.20)	0.92 (0.065)	—	0.03 (0.038)	0.70	1.69	—	1.433 (0.488)
79	–0.00 (0.20)	0.85 (0.128)	0.07 (0.110)	0.03 (0.039)	0.70	1.72	—	2.401 (0.493)
Panel B: In V_t (τ = 120)								
78	0.77 (0.21)	—	0.67*** (0.076)	—	0.50	1.51	67.57 (0.000)	—
78	0.67 (0.22)	—	0.63*** (0.074)	0.08**** (0.049)	0.50	1.49	—	19.55 (0.000)
78	2.21 (0.19)	—	—	0.16** (0.77)	0.08	0.83	160.4 (0.000)	—
78	0.20 (0.19)	0.87 (0.097)	—	—	0.65	1.25	118.7 (0.000)	—
78	0.18 (0.19)	0.80* (0.121)	0.08 (0.086)	—	0.65	1.27	—	3.954 (0.138)
78	0.19 (0.21)	0.86 (0.091)	—	0.01 (0.034)	0.65	1.25	—	3.499 (0.173)
78	0.18 (0.20)	0.79* (0.127)	0.08 (0.086)	0.01 (0.033)	0.64	1.27	—	4.251 (0.235)
Panel B: In V_t (τ = 180)								
78	0.62 (0.20)	—	0.70*** (0.073)	—	0.45	0.91	142.8 (0.000)	—
78	0.55 (0.21)	—	0.68*** (0.076)	0.05 (0.037)	0.44	0.89	—	12.82 (0.002)
78	2.29 (0.13)	—	—	0.13** (0.055)	0.06	0.53	354.0 (0.000)	—
78	0.19 (0.27)	0.85* (0.092)	—	—	0.61	0.94	214.3 (0.000)	—
78	0.20 (0.25)	0.79 (0.194)	0.06 (0.122)	—	0.61	0.94	—	4.778 (0.091)
78	0.25 (0.27)	0.86 (0.096)	—	–0.04 (0.035)	0.60	0.98	—	4.341 (0.114)
78	0.21 (0.25)	0.82 (0.195)	0.05 (0.125)	–0.03 (0.034)	0.60	0.96	—	4.423 (0.219)

The numbers in parentheses beside the parameter estimates are the standard errors, which are computed following a robust procedure taking into account of the heteroscedastic and autocorrelated error structure [see Newey and West (1987)]. The χ^2 test(a) is for the joint hypothesis H_0:$\alpha = 0$ and $\beta^j = 1$ ($j = MF, BS, LRE$) in univariate regressions, and the χ^2 test(b) is for the joint hypothesis H_0:$\beta^{BS} = 1$ and $\beta^{LRE} = 0$ or H_0:$\beta^{MF} = 1$ and $\beta^{BS} = \beta^{LRE} = 0$ in encompassing regressions. The test statistics are reported with the p-values in the brackets beside the statistic.

*, **, and *** indicate that the leading term β coefficient of the regression is significantly different from one at the 10%, 5%, and 1% level and the ****,*****,and ****** indicate that the β coefficient of the remaining terms is significantly different from zero at the 10%, 5%, and 1% level, respectively.

autocorrelation in regression errors from the overlapping samples. Not surprisingly, the DW statistic declines as the forecast horizon increases from 60 to 180 days, indicating that the degree of autocorrelation increases as the sample becomes more overlapped. The standard errors in all regressions are thus estimated following a robust procedure taking into account of both heteroscedasticity and autocorrelation [Newey and West (1987)]. The number of lags used in the estimation is set equal to the number of overlapping periods in the regression.

The most striking finding from Table 6 is that parameter estimates and statistical inferences from the overlapping samples are quite similar to those from the nonoverlapping samples reported in Table 4. For example, the slope coefficient from univariate regressions is consistently positive and statistically significant, supporting the hypothesis that all three volatility measures contain information for future realized volatility. In addition, the encompassing regressions using overlapping samples continue to show that the B–S implied volatility does not subsume all information contained in the lagged realized volatility. The lagged realized volatility is, however, only marginally significant, which is not unexpected since the daily realized volatility is much more volatile than integrated volatility over 60 to 180 days. On the other hand, the model-free implied volatility is found to subsume all information contained in both the B–S implied volatility and lagged realized volatility. As long as the model-free implied volatility is included in the regression, the B–S implied volatility and the lagged realized volatility are no longer statistically significant. These findings are consistent with the results from the nonoverlapping samples.

4.3 Implied index values

Previous research [e.g., Longstaff (1995)] found that the price of the underlying asset implied by option prices may be substantially different from the corresponding actual market price. These differences may be caused by market frictions (e.g., commission fees, bid-ask spread, illiquidity, and taxes), measurement errors, or market inefficiency. In our empirical tests, the B–S and model-free implied volatilities are both calculated using actual index values rather than implied index values. It is prudent to investigate whether the regression results change materially if implied index values are used to calculate implied volatility.

Longstaff (1995) calculated the implied value of the S & P 100 index from option prices using either the B–S model or more generalized models with a four-parameter distribution function. Implied index values obtained this way are nevertheless subject to model misspecification errors. Following Aït-Sahalia and Lo (1998), we use the put-call parity to calculate the implied index value from option prices. Model misspecification errors are avoided as the put-call parity is model free. In addition,

Aït-Sahalia and Lo (1998) documented that the put-call parity is rarely violated for the SPX options.

To use the put-call parity, we need to find matched pairs of call and put options. For each observed call price, we find the closest temporally matched price from a corresponding put option with identical maturity and strike price. Matched pairs with option prices observed more than 5 minutes apart are eliminated. Nonsynchronous trading is unlikely to be a problem with this matching criterion. Finally, we minimize the effect of illiquidity on option prices by only including options that are near the money. Specifically, we only use options whose strike price is within 3% of the index value. These options are traded more actively and their prices are likely to be more efficient and reliable. For strike prices further away from the index value, either the call or the put is deep in the money and may not have sufficient liquidity.

Comparing the actual and implied index values over the sample period, we find that the implied index value is within 0.64% of the actual index value in 98% of the sample. The mean and median of the absolute percentage difference are 0.12 and 0.08%, respectively.[6] In addition, the distribution of the percentage difference is also nearly symmetric with an approximately 50-50 split between positive and negative differences. These results indicate that the implied index value is generally consistent with the actual index value. The choice of actual or implied index values should have little impact on the calculation of the model-free implied volatility. Unreported results indeed confirm this assertion. The choice of actual or implied index value has little impact on the calculated model-free implied volatility or the regression results.

4.4 Alternative measures for realized volatility

All reported results so far are based on realized volatility calculated from 30-minute index returns and lagged daily realized volatility (our proxy for historical volatility) calculated from 5-minute index returns. The latter is corrected for autocorrelation with one correction term while the former is not. Although recent research on high frequency data and our sample estimates for autocorrelations of 5-minute and 30-minute returns suggest that this is a reasonable approach, it is prudent to consider alternative measures for realized volatility and find out whether our regression results are robust to them. As high frequency index returns typically have positive autocorrelation due to the infrequent trading problem, the estimate for realized volatility would be downward biased if no correction is made. We now investigate the robustness of our regression results over alternative estimates for realized volatility.

[6] Compared to the corresponding numbers of 40.0 and 31.1% reported in Longstaff (1995), these differences are negligible.

We first examine the impact of different number of correction terms on our regression analysis. Since autocorrelation more or less disappears after the first lag in our intraday return series, we examine the effect of including zero to three correction terms in Equation (8) for both 5-minute and 30-minute index returns. Realized volatility and lagged realized volatility series are recalculated and the results in Table 4 are reproduced using the new volatility estimates. Unreported results show that test statistics continue to support our key finding that the model-free implied volatility subsumes all information contained in the B–S implied volatility and lagged realized volatility and is a more efficient forecast for future realized volatility. There are also some noticeable differences as more correction terms are added. The regression R^2 is reduced and the lagged realized volatility becomes less significant. This is an indication of an increasing level of noise in the volatility series as more correction terms are included. We interpret these results as evidence supporting our previous findings.

In addition, previous studies such as Andersen, Bollerslev, Diebold, and Ebens (2001) suggest that intraday returns should be cleaned up using an MA(1) filter before they are used to calculate realized volatility. The filtered returns have much reduced autocorrelation and are thus better suited for realized volatility calculation. Applying the MA(1) filter to both 5-minute and 30-minute returns, we recalculate realized volatility and lagged realized volatility series using the filtered series and rerun our univariate and encompassing regressions. Unreported results show that all parameter estimates and test statistics are similar to those reported in Table 4 and there is no material change in statistical inferences.

Finally, we also consider another proxy for historical volatility commonly used in previous research [e.g., Christensen and Prabhala (1998)]—lagged realized volatility over a matching horizon immediately proceeding the observation date. For the 30-day realized volatility calculated on date t, the matching lagged realized volatility is calculated over the 30-day period proceeding date t. In comparison, the lagged realized volatility on the latest trading day is used as our proxy for historical volatility. To see whether the regression results are influenced by our choice of the proxy, we reproduce results in Table 4 using the lagged 30-day realized volatility calculated using 30-minute index returns. Unreported results show that parameter estimates and test statistics are similar when the lagged monthly volatility is used as the proxy for historical volatility. In particular, the lagged monthly volatility is still significant in encompassing regressions involving only the B–S implied volatility but remains insignificant in encompassing regressions involving the model-free implied volatility. Our main findings are thus robust to the choice of lagged realized volatility as well.

5. Conclusions

In this article, we empirically test the forecasting ability and information content of implied volatility. Instead of relying on the B–S implied volatility as in previous research, we implement the model-free implied volatility derived by Britten-Jones and Neuberger (2000). This new implied volatility has several advantages over its predecessor. First, it is independent of any option pricing model, whereas the commonly used B–S implied volatility is model-specific. Second, the model-free implied volatility extracts information from options across all strike prices instead of a single option as in the case of the B–S implied volatility. By aggregating information across options, the model-free implied volatility should be informationally more efficient than the B–S implied volatility. Third, tests based on the model-free implied volatility are direct tests of market efficiency instead of joint tests of market efficiency and the assumed option pricing model. Evidence on market efficiency from the model-free implied volatility is thus not subject to model misspecification errors.

Our research makes several contributions to the related literature. We provide a simpler derivation of the model-free implied volatility under Britten-Jones and Neuberger's (2000) diffusion assumption and then generalize it to processes with jumps. We thus establish the validity of the model-free implied volatility and ensure that it applies to general asset price processes. In addition, we develop a simple method for implementing the model-free implied volatility using observed option prices. We provide theoretical bounds on truncation errors due to the finite range of available strike prices and illustrate the minimum range of strike prices required to control such errors. We show that the model-free implied volatility can be calculated accurately using our implementation method. Finally, we use univariate and encompassing regressions to investigate the forecasting ability and information content of the model-free implied volatility. Our findings from the SPX options support the hypothesis that the model-free implied volatility subsumes all information contained in the B–S implied volatility and past realized volatility and is a more efficient forecast for future realized volatility. These results are robust to alternative estimation methods, volatility series over different horizons, the choice of actual or implied index values, and alternative methods for calculating realized volatility. Our findings also provide theoretical and empirical support for the recent decision by the CBOE to modify its widely watched VIX index. The new VIX index is based on the model-free implied volatility instead of the B–S implied volatility of at-the-money options.

Appendix

Proof of Proposition 1. We first prove that Equation (1) holds under the diffusion assumption. The no-arbitrage argument implies that there exists a forward measure F such that:

$$dF_t/F_t = \sqrt{V_t}dW_t, \tag{A1}$$

where V_t is the instantaneous variance and W_t is a standard F-Brownian motion. Applying Ito's lemma, we have:

$$d \ln F_t = -\frac{1}{2}V_t dt + \sqrt{V_t}dW_t.$$

Integrating over time and taking expectations, we further have

$$E_0^F\left(\int_0^T V_t dt\right) = 2\left[\ln F_0 - E_0^F(\ln F_T)\right],$$

where the LHS is equivalent to $E_0^F\left[\int_0^T \left(\frac{dF_t}{F_t}\right)^2\right]$. Thus we only need to prove that:

$$\int_0^\infty \frac{C^F(T, K) - \max(0, F_0 - K)}{K^2}dK = \ln F_0 - E_0^F(\ln F_T). \tag{A2}$$

Applying integration by parts to the LHS of (A2), we have

$$\int_0^\infty \frac{C^F(T, K) - \max(0, F_0 - K)}{K^2}dK = -\frac{C^F(T, K) - \max(F_0 - K, 0)}{K}\Bigg|_0^\infty + \int_0^\infty \frac{C_K^F(T, K) + 1_{F_0 > K}}{K}dK, \tag{A3}$$

where $C_K^F(T,K)$ is the partial derivative of option price with respect to strike price. The first term on the RHS of Equation (A3) vanishes under the mild condition that the density of the forward price distribution is bounded. Since $C_K^F(T, K) = E_0^F\left[\frac{\partial \max(0, F_T - K)}{\partial K}\right] = -E_0^F(1_{F_T > K})$, the second term on the RHS of Equation (A3) can be simplified to:

$$\int_0^\infty \frac{C_K^F(T, K) + 1_{F_0 > K}}{K}dK = \ln F_0 - E_0^F(\ln F_T).$$

Equation (A2) thus holds. This completes the proof for diffusion processes.

We now consider processes with jumps. Under certain regularity conditions, the forward price as a martingale can be decomposed canonically into two orthogonal components: a purely continuous martingale and a purely discontinuous martingale [see Jacod and Shiryaev (1987) and Protter (1990)]. Incorporating the discontinuous component (i.e., jumps), we restate the forward price process as:

$$dF_t/F_t = \sqrt{V_t}dW_t + J_t dN_t - \mu_t\lambda_t dt, \tag{A4}$$

where N_t is a pure jump process with time-varying intensity λ_t, J_t, is the jump size with instantaneous mean μ_t, and the jump component is uncorrelated with the diffusion component.

Applying *Ito*'s lemma, we have:

$$d \ln F_t = -\frac{1}{2}V_t dt + \sqrt{V_t}dW_t + \ln(1 + J_t)dN_t - \mu_t\lambda_t dt.$$

Since $\ln(1 + J_t) \approx J_t - \frac{1}{2}J_t^2$, we further have:

$$E_0^F\left[\int_0^T d\ln F_t\right] \approx -\frac{1}{2}E_0^F\left[\int_0^T (V_t + \lambda_t J_t^2)dt\right],$$

or equivalently:

$$E_0^F\left[\int_0^T (V_t + \lambda_t J_t^2)dt\right] \approx 2\left[\ln F_0 - E_0^F(\ln F_T)\right].$$

Since the LHS of the above equation is the integrated return variance, we now have:

$$E_0^F\left[\int_0^T \left(\frac{dF_t}{F_t}\right)^2\right] \approx 2\left[\ln F_0 - E_0^F(\ln F_T)\right]. \tag{A5}$$

Note that Equation (A2) holds even when asset returns include jumps because its derivation does not require any knowledge of the asset return process. Combining Equations (A2) and (A5), we complete the proof for processes with jumps.[7]

Proof of Proposition 2. We first consider the right truncation error, $2\int_{K_{\max}}^{+\infty}\frac{C^F(T,K)-\max(0,F_0-K)}{K^2}dK$. From Equation (A3), this error can be rewritten as:

$$-2\frac{C^F(T,K)-\max(0,F_0-K)}{K}\Bigg|_{K_{\max}}^{+\infty} + 2\int_{K_{\max}}^{+\infty}\left[C_K^F(T,K) + 1_{F_0>K}\right]d\ln K, \tag{A6}$$

which further simplifies to:

$$2\frac{C^F(T,K_{\max})-\max(0,F_0-K_{\max})}{K_{\max}} - 2E_0^F\left[\max\left(0,\ln\left(\frac{F_T}{K_{\max}}\right)\right)\right] + \max\left(0,\ln\left(\frac{F_0}{K_{\max}}\right)\right).$$

With $K_{\max} > F_0$, we have:

$$2\int_{K_{\max}}^{+\infty}\frac{C^F(T,K)-\max(0,F_0-K)}{K^2}dK = 2\int_{K_{\max}}^{+\infty}\left[\frac{F_T-K_{\max}}{K_{\max}} - \ln\left(1+\frac{F_T-K_{\max}}{K_{\max}}\right)\right]\phi(F_T)dF_T,$$

where $\Phi(F_T)$ is the density of the forward price. Using the properties of the Taylor series expansion for the log function, we have the following upper bound for the right truncation error:

$$2\int_{K_{\max}}^{+\infty}\frac{C^F(T,K)-\max(0,F_0-K)}{K^2}dK \le \int_{K_{\max}}^{+\infty}\left(\frac{F_T-K_{\max}}{K_{\max}}\right)^2\phi(F_T)dF_T. \tag{A7}$$

where the RHS can be rewritten as $E_0^F\left[\left(\frac{F_T-K_{\max}}{K_{\max}}\right)^2\middle|F_T>K_{\max}\right]$, which proves Equation (3). The upper bound for the left truncation error in Equation (4) is derived in a similar fashion.

[7] Note the Equation (A5) does not hold exactly. The approximation error is due to ignoring higher order ($\geq$3) terms in the ln (1+J_t) expansion, which are mainly determined by the third moment of the jump size J_t. When the jump size is negatively skewed, the model-free implied volatility [the RHS of Equation (A5)] tends to overestimate the asset price variance. However, our numerical evaluation suggests that the approximation error is negligible in commonly encountered cases (see the results in Figure 1 and Table 1).

It is also straightforward to derive model-free upper bounds based on available option prices. Invoking the monotonicity and convexity of option prices, we have

$$2\int_{K_{\max}}^{+\infty}\frac{C^F(T,\,K)-\max(0,\,F_0-K)}{K^2}\,dK \le \frac{2C^F(T,\,K_{\max})}{K_{\max}},$$

$$2\int_{K_{\min}}^{+\infty}\frac{C^F(T,K)-\max(0,\,F_0-K)}{K^2}\,dK \le \frac{2P^F(T,\,K_{\min})}{K_{\min}}\ln\,(K_{\min}/K^*)+\varepsilon(K^*),$$

where $P^F(T,K)$ is the forward put option price and K^* is a sufficiently small positive number such that $\varepsilon(K^*)=2\int_0^{K^*}\frac{C^F(T,\,K)-\max(0,\,F_0-K)}{K^2}\,dK$ is negligible.

References

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

Aït-Sahalia, Y., P. A. Mykland, and L. Zhang, 2003, "How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise," *Review of Financial Studies*, 18, 351–416.

Andersen, T. G., and T. Bollerslev, 1998, "Answering the Critics: Yes, ARCH Models Do Provide Good Volatility Forecasts," *International Economic Review*, 39, 885–905.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and. H. Ebens, 2001, "The Distribution of Realized Stock Return Volatility," *Journal of Financial Economics*, 61, 43–76.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys, 2001, "The Distribution of Realized Exchange Rate Volatility," *Journal of the American Statistical Association*, 96, 42–55.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys, 2003, "Modeling and Forecasting Realized Volatility," *Econometrica*, 71, 579–625.

Barndorff-Nielsen, O. E., and N. Shephard, 2003, "Realized Power Variation and Stochastic Volatility Models," *Bernouilli*, 9, 243–265.

Bakshi, G., C. Cao, and Z. Chen, 1997, "Empirical Performance of Alternative Option Pricing Models," *Journal of Finance*, 52, 2003–2049.

Bakshi, G., C. Cao, and Z. Chen, 2000, "Pricing and Hedging Long-Term Options," *Journal of Econometrics*, 94, 277–318.

Bates, D., 1991, "The Crash of '87: Was it Expected? The Evidence from Options Markets," *Journal of Finance*, 46, 1009–1044.

Black, F., and M. Scholes, 1973, "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy*, 81, 637–659.

Blair, B., S. H. Poon, and S. J. Taylor, 2001, "Forecasting S&P 100 Volatility: The Incremental Information Content of Implied Volatility and High Frequency Index Returns," *Journal of Econometrics*, 105, 5–26.

Breeden, D. T., and R. H. Litzenberger, 1978, "Prices of State-Contingent Claims Implicit in Option Prices," *Journal of Business*, 51, 621–651.

Britten-Jones, M., and A. Neuberger, 2000, "Option Prices, Implied Price Processes, and Stochastic Volatility," *Journal of Finance*, 55, 839–866.

Campa, J. M., K. P. Chang, and R. L. Reider, 1998, "Implied Exchange Rate Distributions: Evidence from OTC Option Markets," *Journal of International Money and Finance*, 17, 117–160.

Canina, L., and S. Figlewski, 1993, "The Informational Content of Implied Volatility," *Review of Financial Studies*, 6, 659–681.

Christensen, B. J., C. S. Hansen, and N. R. Prabhala, 2001, "The Telescoping Overlap Problem in Options Data," Working paper, University of Aarhus and University of Maryland,

Christensen, B. J., and N. R. Prabhala, 1998, "The Relation between Implied and Realized Volatility," *Journal of Financial Economics*, 50, 125–150.

Day, T. E., and C. M. Lewis, 1992, "Stock Market Volatility and the Information Content of Stock Index Options," *Journal of Econometrics*, 52, 267–287.

Derman, E., and I. Kani, 1994, "Riding on a Smile," *Risk*, 7, 32–39.

Derman, E., and I. Kani, 1998, "Stochastic Implied Trees: Arbitrage Pricing with Stochastic Term and Strike Structure of Volatility," *International Journal of Theoretical and Applied Finance*, 1, 61–110.

Derman, E., I. Kani, and N. Chriss, 1996, "Implied Trinomial Trees of the Volatility Smile," *Journal of Derivatives*, 3, 7–22.

Dumas, B., J. Fleming, and R. E. Whaley, 1998, "Implied Volatility Functions: Empirical Tests," *Journal of Finance*, 53, 2059–2106.

Ederington, L. H., and W. Guan, 2002, "Is Implied Volatility an Informationally Efficient and Effective Predictor of Future Volatility?" *Journal of Risk*, 4, 29–46.

Fleming, J., 1998, "The Quality of Market Volatility Forecast Implied by S&P 100 Index Option Prices," *Journal of Empirical Finance*, 5, 317–345.

French, K. R., G. W. Schwert, and R. F. Stambaugh, 1987, "Expected Stock Returns and Volatility," *Journal of Financial Economics*, 19, 3–30.

Hansen, P. R., and A. Lunde, 2004, "Realized Variance and Market Microstructure Noise," forthcoming *Journal of Business and Economic Statistics*.

Harvey, C. R., and R. E. Whaley, 1991, "S&P 100 Index Option Volatility," *Journal of Finance*, 46, 1551–1561.

Harvey, C. R., and R. E. Whaley, 1992a, "Market Volatility Prediction and the Efficiency of the S&P 100 Index Option Market," *Journal of Financial Economics*, 31, 43–73.

Harvey, C. R., and R. E. Whaley, 1992b, "Dividends and S&P 100 Index Option Valuation," *Journal of Futures Markets*, 12, 123–137.

Hausman, J., 1978, "Specification Tests in Econometrics," *Econometrica*, 46, 1251–1271.

Heston, S. L., 1993, "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options," *Review of Financial Studies*, 6, 327–343.

Jackwerth, J. C., 1999, "Option-Implied Risk-Neutral Distributions and Implied Binomial Trees: A Literature Review," *Journal of Derivatives*, 6, 1–17.

Jacod, J., and A. N. Shiryaev, 1987, *Limit Theorems for Stochastic Processes*, Springer-Verlag, Berlin.

Jorion, P., 1995, "Predicting Volatility in the Foreign Exchange Market," *Journal of Finance*, 50, 507–528.

Lamoureux, C. G., and W. D. Lastrapes, 1993, "Forecasting Stock-Return Variance: Toward an Understanding of Stochastic Implied Volatilities," *Review of Financial Studies*, 6, 293–326.

Ledoit, O., and P. Santa-Clara, 1998, "Relative Pricing of Options with Stochastic Volatility," Working paper, University of California at Los Angeles.

Longstaff, F. A., 1995, "Option Pricing and the Martingale Restriction," *Review of Financial Studies*, 8, 1091–1124.

Newey, W. K., and K. D. West, 1987, "A Simple Positive Definite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

Pong, S., M. B. Shackleton, S. J. Taylor, and X. Xu, 2004, "Forecasting Sterling/Dollar Volatility: A Comparison of Implied Volatility and AR(FI)MA Models," *Journal of Banking and Finance*, 28, 2541–2563.

Protter, P., 1990, *Stochastic Integration and Differential Equations: A New Approach*, Springer, Berlin.

Richardson, M., and T. Smith, 1991, "Tests of Financial Models in the Presence of Overlapping Observations," *Review of Financial Studies*, 4, 227–254.

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

Rubinstein, M., 1998, "Edgeworth Binomial Trees," *Journal of Derivatives*, 5, 20–27.

Shimko, D., 1993, "Bounds of Probability," *Risk*, 6, 33–37.

Stoll, H. R., and R. E. Whaley, 1990, "The Dynamics of Stock Index and Stock Index Futures Returns," *Journal of Financial and Quantitative Analysis*, 25, 441–468.

White, H., 1980, "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 817–838.

Zhang, L., P. A. Mykland, and Y. Aït-Sahalia, 2003, "A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data," forthcoming *Journal of the American Statistical Assocation*.

Zhou, B., 1996, "High-Frequency Data and Volatility in Foreign-Exchange Rates," *Journal of Business and Economic Statistics*, 14, 45–52.