Journal of Empirical Finance 5 (1998) 317–345

Journal of EMPIRICAL FINANCE

The quality of market volatility forecasts implied by S&P 100 index option prices

Jeff Fleming [*]

Jones Graduate School of Management, Rice University, Houston, TX 77005, USA

Abstract

This study examines the performance of the S&P 100 implied volatility as a forecast of future stock market volatility. The results indicate that the implied volatility is an upward biased forecast, but also that it contains relevant information regarding future volatility. The implied volatility dominates the historical volatility rate in terms of ex ante forecasting power, and its forecast error is orthogonal to parameters frequently linked to conditional volatility, including those employed in various ARCH specifications. These findings suggest that a linear model which corrects for the implied volatility's bias can provide a useful market-based estimator of conditional volatility. © 1998 Elsevier Science B.V. All rights reserved.

JEL classification: G12; G13

Keywords: Volatility; Options; Forecasting; Implied volatility; Information content

1. Introduction

The fundamental implications between asset risk and return motivate broad research interest in stock market volatility. Unfortunately, conditional volatility is unobservable. A number of techniques have been developed to statistically model this parameter, [1] but despite the success of these efforts, relying on statistical

[*] Corresponding address: Jones Graduate School-MS 531, Rice University, 6100 Main Street, Houston, TX 77005-1892, USA. Tel.: +1-713-527-4677; e-mail: jfleming@rice.edu.

[1] See Bollerslev et al. (1992) for a survey of autoregressive conditional heteroskedasticity (ARCH) models. Pagan and Schwert (1990) provide a discussion of nonparametric methods.

PII: S0927-5398(98)00002-4

forecasts assumes model stability into the forecast horizon, and, perhaps more critically, the likelihood exists that significant conditioning information has been overlooked.

As an alternative, option valuation models such as the Black and Scholes (1973) specification allow us to directly impute a conditional, market-determined volatility forecast. Other than the volatility input, all of the option pricing parameters are objectively available. So, if the option market is efficient and the valuation model is correctly specified, all relevant conditioning information is collapsed into the option price. The implied volatility, then, should represent a superior volatility forecast.

Despite the strength of this implication and widespread use of implied volatility as a proxy for conditional volatility,[2] there exists only limited evidence of support. The first studies of the issue by Latané and Rendleman (1976), Chiras and Manaster (1978), and Beckers (1981) found that the implied volatility indeed contained relevant information regarding future volatility. These studies, however, examined fairly small datasets and focused on the cross-sectional relations within a select group of stocks.

More recent evidence, based on the analysis of overlapping time-series observations, is less supportive. Day and Lewis (1992), using S&P 100 index options, and Lamoureux and Lastrapes (1993), using individual equity options, find that the implied volatility contains useful information in forecasting volatility, but also that time-series models contain information incremental to the implied volatility. Canina and Figlewski (1993) conclude that the S&P 100 implied volatility is such a poor forecast that it is dominated by the historical volatility rate.

On the other hand, Jorion (1995) finds more favorable evidence in the currency markets where the implied volatility outperforms both moving average and GARCH forecasts. He attributes the disparity from the earlier results to implied volatility measurement error and biases in the regression tests of forecast performance. Measurement error is problematic for index options due to infrequent trading effects on the underlying index price, and, if closing prices are used, due to the mismatch between the times that the stock and option markets close. Moreover, as in all markets, converting an option price into an implied volatility incurs error due to bid/ask spreads and noncontinuous prices. The econometric difficulties in measuring forecast performance stem from the unique structure of the option expiration cycle (i.e., the forecast interval) and the persistence of volatility. Implied volatility is not covariance stationary and nearly a unit root process.

[2] Schmalensee and Trippi (1978) and Poterba and Summers (1986), for example, employ the implied volatility in this context. Day and Lewis (1988) study expiration-day effects using the implied volatility as a volatility proxy, and Schwert (1990) examines its behavior during the 1987 market crash as a measure of market volatility expectations. Stein (1989) characterizes the term structure of implied volatility as an indication of option market overreactions.

Therefore, traditional regression analysis is biased and perhaps spurious in small samples.

This study addresses these problems and provides a new examination of the forecast quality of the S&P 100 implied volatility. In estimating the implied volatility series, we employ the Harvey and Whaley (1992a) procedure involving an end-of-day window of option prices. Using a window of transactions minimizes the effects of bid/ask and noncontinuous option prices, and, since the window is centered around the stock market close, the concerns about infrequent trading and timing mismatch are reduced. In addition, our econometric analysis uses overlapping daily observations and specifically addresses the spurious regression problem inherent in volatility time-series. We also develop an estimator which accounts for the covariance nonstationarity of implied volatility forecast errors.

The analysis focuses on whether the implied volatility is an unbiased volatility forecast, and whether its forecast error is orthogonal to parameters often linked to conditional volatility. The unbiasedness results indicate that the implied volatility is a biased forecast which significantly overstates future volatility. Perhaps more importantly, however, we find no evidence of orthogonality rejections. The implied volatility's forecast power dominates that of the historical volatility rate, and none of the information variables frequently used to model conditional volatility can explain the component of volatility that is unexplained by the implied volatility.

The study is organized as follows. Section 2 develops our interpretation of the implied volatility, and then describes the data and the procedure for estimating the implied volatility series. Section 3 outlines the empirical methodology. Section 4 presents the results. Section 5 extends the analysis to monthly and daily forecast horizons, and Section 6 offers a conclusion.

2. The market volatility forecast

2.1. Interpreting the Black / Scholes implied volatility

The Black/Scholes model assumes a known and constant volatility rate. That we consider the forecast quality of its implied volatility, however, suggests volatility is uncertain and time-varies. We can reconcile these issues by developing a relation between the implied volatility and the true time-varying volatility process.

Consider, for example, the Hull and White (1987) model where the asset price and volatility are uncorrelated and the volatility risk premium is zero. The time t value of an option expiring at T is

$$f_t = E[\mathrm{BS}(\sigma_{t;T})|\boldsymbol{\Phi}_t], \tag{1}$$

the expected Black/Scholes value, $BS(\cdot)$, evaluated at the (uncertain) average instantaneous volatility over the option's life, $\sigma_{t;T} = \sqrt{\frac{1}{T-t}\int_t^T \sigma_x^2 dx}$, and conditioned on $\boldsymbol{\Phi}_t$, all information available at time t. Feinstein (1989a) shows that for near-expiration, at-the-money options, the Black/Scholes model is nearly linear in its volatility argument, so

$$E[\mathrm{BS}(\sigma_{t;T})|\boldsymbol{\Phi}_t] \approx \mathrm{BS}(E[\sigma_{t;T}|\boldsymbol{\Phi}_t]), \tag{2}$$

and

$$f_t \approx \mathrm{BS}(E[\sigma_{t;T}|\boldsymbol{\Phi}_t]). \tag{3}$$

If the approximation is exact, then the implied volatility, $\overline{\sigma}_{t;T} = \mathrm{BS}^{-1}(f_t)$, satisfies

$$\overline{\sigma}_{t;T} = E[\sigma_{t;T}|\boldsymbol{\Phi}_t]. \tag{4}$$

In words, the implied volatility should represent an unbiased forecast of the average volatility over the life of the option, and its forecast error should be orthogonal to the time t information set. [3]

The validity of this implication for S&P 100 index options is subject to two important limitations. First, a zero correlation between price and volatility seems unrealistic because it contradicts the well-known leverage effect observed in the stock market. Second, the implication is developed for European options while S&P 100 options are American style. [4] These limitations may cause the implied volatility to deviate from the market's true volatility forecast and induce inaccuracy in Eq. (4). In this respect, the relation is just an approximation for S&P 100 options.

Ideally, we should instead compute an implied volatility based on a more realistic volatility process and including early exercise opportunities. This implied volatility, then, could be more carefully related to the market's volatility expectation. From a practical standpoint, however, this alternative poses additional complications. First, we can never be certain that our assumption about the volatility process is correct. Second, even if the process is adequately specified, additional parameters must be estimated which amplifies the error in recovering the true volatility forecast. [5] As a result, it is far more common to use a

[3] The validity of this relation depends on the approximation in Eq. (2). Since the Black/Scholes function is strictly concave in volatility, Jensen's inequality implies that the Black/Scholes value (using expected volatility) always exceeds the stochastic volatility value. But, for near-expiration, at-the-money options, the magnitude of this bias is small, and its impact on the implied volatility has the opposite sign of the bias found in this study.

[4] This second limitation would not affect S&P 500 index options. Unfortunately, the market for these options has been less active than for S&P 100 options, particularly in the 1980s. As a result, computing implied volatilities using S&P 500 options entails additional estimation error.

[5] The implied tree approach of Rubinstein (1994), for example, relies on a nonparametric approach to estimate the underlying asset price dynamics. Dumas et al. (1998), however, find that the out-of-sample valuation and hedging performance of this approach is less reliable than that of an ad hoc version of the Black/Scholes model.

Black/Scholes-type implied volatility (whether based on European or American valuation) as an instrument in forecasting volatility.

The obvious question, then, concerns the empirical accuracy of using the implied volatility as a forecast. It seems plausible that the limitations in Eq. (4) for S&P 100 options may have only a small effect for nearby expiration, at-the-money options. The effect of a more general volatility process depends on the disparity between constant and stochastic volatility option valuations. Chesney and Scott (1989), using a random variance model, and Duan (1995), using a GARCH option pricing model, find that the disparity is quite small for nearby at-the-money options, averaging less than 4% of option value. Similarly, Harvey and Whaley (1992b) and Fleming and Whaley (1994) show that the early exercise premium for these options averages just 2–5% of value. This evidence suggests that the magnitude of bias in the implied volatility as a forecast may be small. Furthermore, for options within a fairly tight band of moneyness and time to expiration, it seems reasonable to suppose the bias induced by volatility and/or early exercise misspecification is relatively stable. To the extent this occurs, the orthogonality implication of Eq. (4) is preserved. In this case, the implied volatility covaries with future volatility, but a constant correction factor is necessary to account for the bias.

These arguments motivate our empirical examination of the forecast quality of the S&P 100 implied volatility. Consistent with Eq. (4), we consider both its unbiasedness and orthogonality properties. However, given the limitations described above, we rely on Eq. (4) only as a guide for our empirical specifications and our results should not be viewed as a test of market efficiency. Nonetheless, the investigation addresses an important issue. If the implied volatility represents a poor forecast, then this evidence challenges its widespread use in this context. Evidence to the contrary, on the other hand, tends to validate the implied volatility as a useful forecasting instrument.

2.2. Implied volatility estimation

Suppose the average future volatility rate is known by option market participants. The measured forecast quality of the implied volatility may still be imperfect if we cannot identify the market's volatility expectation (i.e., the 'true' implied volatility). Two sources of error can affect implied volatility estimates. Specification error exists to the extent the market prices options differently than the assumed valuation model; and, estimation error exists when bid/ask price effects and infrequent trading among index stocks cause the observed option price to differ from its theoretical value. Accurately assessing the forecast quality of the implied volatility depends critically on minimizing these two sources of error.

We begin by specifying an appropriate valuation model. Valuing S&P 100 options must account for early exercise opportunities as well as discrete index dividends and embedded wildcard options. [6] To incorporate these features, we use the modified binomial model of Fleming and Whaley (1994). The model provides a theoretical option value, g, that, in the absence of specification error, equals the market price, f,

$$f = g(S, X, \tau, r, d, \sigma) \tag{5}$$

where S is the S&P 100 index level; X and τ are the option's exercise price and time to expiration (t to T); r is the riskless interest rate; d is the amount (and timing) of dividends during τ; and, σ is the index volatility rate.

Having specified a valuation model, we can now invert the model to estimate the implied volatility. The precision of our estimate, however, is potentially influenced by microstructural effects which can distort observed option prices. To minimize these effects, we employ an end-of-day estimation window as developed by Harvey and Whaley (1992a). The implied volatility, $\bar{\sigma}_\tau$, is estimated from all option transactions within a 10-min window centered around the stock market close. Using an interval allows an 'averaging' technique which minimizes the effects of bid/ask and noncontinuous prices. In addition, by using end-of-day prices, the effects of infrequent trading become less significant. [7]

The daily call (or put) implied volatility is obtained from the nonlinear regression

$$f_i = g(S_i, X_i, \tau, r_\tau, d_\tau, \bar{\sigma}_\tau) + \varepsilon_i, \tag{6}$$

where f_i denotes the market price of the i-th call (put) option in the 10-min interval. Given our discussion of Eq. (4), we use only near-expiration, at-the-money options in the regression. Using at-the-money options also further mitigates the bid/ask and noncontinuous price effects because the implied volatilities for these options are fairly insensitive to small changes in option price.

2.3. Data

The primary data source is the S&P 100 option transaction price history provided by the Chicago Board Options Exchange. In addition to the option price, each record in this dataset contains the contemporaneous S&P 100 index level. Computing the implied volatility also requires the riskless interest rate and expected index dividends. To proxy for the riskless rate, we use the effective yield on the Treasury bill whose maturity most closely matches the option expiration, but is at least 30 days. The effective yield is computed from the average of the 3:00 CST bid and ask discounts reported in *The Wall Street Journal*. To proxy for

[6] S&P 100 options embed an end-of-day wildcard exercise feature. Each day, the option's settlement price is determined at 3:00 CST, but the exercise decision can be deferred until 3:20. As a result, an adverse index move during these 20 min (which would otherwise reduce option value) can be 'recovered' by exercising at the previously-established settlement price.

[7] Infrequent trading typically induces positive autocorrelation in returns. The close-to-close index returns during the period of this study, however, exhibit an insignificant negative autocorrelation.

expected dividends, we use the actual dividends paid by S&P 100 firms. [8] Prior to June 1988, the aggregate index dividend series was not publicly reported. We obtained these data from Harvey and Whaley (1992b), and the dividends after this date were collected from the *Standard & Poor's 100 Information Bulletin*.

To limit the effect of option expirations, the options used in our analysis are the nearest, but with at least 15 days, to expiration. S&P 100 options expire on a monthly cycle so this selection rule provides implied volatilities that range from 15 to 47 days to expiration and have an average horizon of about 30 calendar days.

Our sample period is from October 1985 through April 1992. Within this period, we eliminate all observations that overlap the October 1987 stock market crash. [9] This sample encompasses 80 contract months and has 1664 daily observations. Ideally, we would prefer to retain the crash data in the analysis, however, including these data in the limited sample overstates the true likelihood of this event. Moreover, although including the crash would dramatically impact our parameter estimates, it does not alter our conclusion that the implied volatility is a biased forecast. Including the crash actually increases the bias since it induced a subsequent long-term increase in implied volatility which more than offsets the understatement of volatility on October 19.

3. Methodology

3.1. The implied volatility hypothesis

This section outlines our methodology for evaluating the forecast performance of the implied volatility. The underlying hypothesis is that the implied volatility represents an unbiased forecast and that its forecast error is orthogonal to the market's information set. This hypothesis is consistent with Eq. (4), but its validity for S&P 100 options may be influenced by misspecification of the 'true' stochastic volatility process and the effect of early exercise opportunities. As a result, the empirical analysis can be viewed as an assessment of whether these influences are small enough that the implied volatility adequately proxies for expected volatility.

According to our hypothesis,

$$\sigma_{t;T} = \overline{\sigma}_{t;T} + e_{t;T}, \tag{7}$$

where the implied volatility forecast error, $e_{t;T}$, should be mean zero and orthogo-

[8] Using the actual series to proxy for expected dividends is not unreasonable for the options used in this study. The options have an average time to expiration of a month, and most large firms declare dividends at least a month prior to payment.

[9] For most of our analysis, the crash period includes all observations from the November/December 1987 option contract months, October 5 through December 4.

nal to the conditioning vector, $\boldsymbol{\Phi}_t$. Section 2.2 described the estimation procedure for the implied volatility, $\bar{\sigma}_{t;T}$. We now must estimate $\sigma_{t;T}$, the (unobservable) mean instantaneous volatility realized over the option's life. The sample variance of returns,

$$\widehat{\sigma^2_{t;T}}$$

is frequently used to measure variability, but in our analysis, $\sigma_{t;T}$ (rather than $\sigma^2_{t;T}$) is the parameter of interest. Therefore, we define $\hat{\sigma}_{t;T} =$

$$(\widehat{\sigma^2_{t;T}})^{\frac{1}{2}}$$

as our measure of realized volatility. In small samples, $\hat{\sigma}_{t;T}$ is generally a biased estimator of $\sigma_{t;T}$, and the correction factor necessary for unbiasedness depends on the distribution of returns. For the forecast horizons considered in this study, however, the size of the correction factor would be very small. [10]

Substituting this measure of realized volatility into Eq. (7),

$$\hat{\sigma}_{t;T} = \bar{\sigma}_{t;T} + u_{t;T}, \tag{8}$$

where u captures the forecast error of (7) and the estimation error of the realized volatility. If $\hat{\sigma}$ is unbiased for σ, then u is mean zero, and, if the estimation error is unpredictable, u will possess the orthogonality properties of e.

The specification in Eq. (8) can be evaluated using the moments' (GMM) method of Hansen (1982) by estimating α and β in the moment vector,

$$\boldsymbol{g}_T(\alpha,\beta) = \frac{1}{NK}\sum_{t=1}^{NK}(\hat{\sigma}_t - \alpha - \beta\hat{\sigma}_t)\boldsymbol{Z}_t, \tag{9}$$

where the term NK is defined as the number of observations, and $\boldsymbol{Z}_t$ represents a vector of instruments. Under our hypothesis, the estimates of α and β should be indistinguishable from zero and one, respectively, and $\epsilon_t = \hat{\sigma}_t - \alpha - \beta\bar{\sigma}_t$ should be orthogonal to every subset $\boldsymbol{Z}_t$ of the full information set $\boldsymbol{\Phi}_t$. [11]

[10] For example, under the assumption of normally distributed returns, the correction factor is

$$((T-t-1)/2))^{\frac{1}{2}}\Gamma((T-t-1)/2)/\Gamma((T-t)/2),$$

where $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}$, $\mathrm{d}x$ is the gamma function. For the sample used in this study, the period from t to T is always at least nine trading days which implies a maximum correction of just 3.166%. As a result, using this correction has no meaningful effect on any of the empirical results.

[11] Our regressions do not correct for the errors-in-the-variables problem induced by mismeasurement of the implied volatility. Two factors motivate this decision. First, instrumenting with either the S&P 500 futures option implied volatility or the lagged implied volatility does not noticeably affect the slope estimates. Second, because the true implied volatility is unobservable, empirical interest is in the forecast quality of the estimated implied volatility.

Defining $\boldsymbol{\delta} = (\alpha, \beta)'$, GMM estimation of (9) minimizes the criterion function,

$$J_T(\boldsymbol{\delta}) = \boldsymbol{g}_T(\boldsymbol{\delta})' \boldsymbol{\Omega}_T^{-1} \boldsymbol{g}_T(\boldsymbol{\delta}), \tag{10}$$

where $\boldsymbol{\Omega}_T$ is a consistent estimator for the covariance matrix,

$$\boldsymbol{\Omega} = \sum_{j=-\infty}^{+\infty} E\left[\left(\hat{\sigma}_t - \left(1, \bar{\sigma}_t\right) \boldsymbol{\delta}_0 \right) \left(\hat{\sigma}_{t-j} - \left(1, \bar{\sigma}_{t-j}\right) \boldsymbol{\delta}_0 \right) \boldsymbol{Z}_t \boldsymbol{Z}'_{t-j} \right]. \tag{11}$$

Eq. (10) is exactly identified when $\boldsymbol{Z}_t = (1, \bar{\sigma}_t)'$, and the GMM estimate of $\boldsymbol{\delta}$ identically equals the ordinary least squares estimate. The GMM approach, however, can account for the overlapping error structure induced by the option expiration cycle, and is robust to conditional heteroskedasticity and serially correlated errors. Hansen (1982) demonstrates the asymptotic distribution of the GMM estimator,

$$\sqrt{NK}\,(\boldsymbol{\delta}_T - \boldsymbol{\delta}_0) \overset{a}{\sim} N\left(0, (\boldsymbol{D}' \boldsymbol{\Omega}^{-1} \boldsymbol{D})^{-1}\right), \tag{12}$$

where

$$\boldsymbol{D} = E\left[\boldsymbol{Z}_t \frac{\partial\left(\hat{\sigma}_t - \left(1, \bar{\sigma}_t\right) \boldsymbol{\delta} \right)}{\partial \boldsymbol{\delta}} \right]. \tag{13}$$

Eqs. (9) and (10) also facilitate a test of the orthogonality restrictions in Eq. (4). Specifically, the forecast errors embedded in Eq. (9) should be orthogonal to the elements of $\boldsymbol{\Phi}_t$. Let $\boldsymbol{z}_t$ represent a row vector of selected instruments from $\boldsymbol{\Phi}_t$. For $\boldsymbol{Z}_t = (1, \bar{\sigma}_t, \boldsymbol{z}_t)'$, the system in (10) is over-identified, and the closeness of the covariance-weighted moment conditions to zero measures the orthogonality of the forecast errors with the information set. In particular,

$$NK \boldsymbol{g}'_T(\boldsymbol{\delta}_T) \boldsymbol{\Omega}_T^{-1} \boldsymbol{g}_T(\boldsymbol{\delta}_T) \overset{a}{\sim} \chi_n^2, \tag{14}$$

where n is the dimension of $\boldsymbol{z}_t$.

While the GMM approach can measure the forecast quality of the implied volatility, a (sufficient) condition for consistent estimation is the stationarity and ergodicity of the error vector. Time-series volatility data, however, have a high degree of serial correlation and this raises a concern about spurious results. In addition, the implied volatility forecast errors are not covariance stationary due to the unique overlap structure caused by the option expiration cycle. These topics are addressed below.

3.2. Spurious regressions

Eq. (9) underlies recent tests of the implied volatility as a conditional volatility forecast. [12] The series in this regression are specified in levels and each series has

[12] See, for example, Feinstein (1989b) and Canina and Figlewski (1993).

a high serial correlation. One source of the serial correlation is volatility persistence, and a second source occurs using data sampled more finely (daily) than the forecast interval (monthly). To the extent these series are nonstationary, Eq. (9) represents a spurious regression. Granger and Newbold (1974) and Phillips (1986) demonstrate the invalidation of conventional inference under these conditions.

Using Dickey and Fuller (1979, 1981) tests, we can reject the nonstationarity of each of the volatility series, however, the spurious regression problem may still affect inference based on small samples. The high serial correlations in the data induce residual autocorrelation and this can yield inefficient slope estimates and spurious explanatory power. Moreover, tests of competing explanatory variables will be biased in favor of detecting a stronger relation for the variable with the largest serial correlation. These concerns gradually disappear with sample size and they are very minimal with the large sample used in this study.

Nonetheless, the spurious regression problem is a key feature of time-series tests of the implied volatility hypothesis using overlapping data. Therefore, for our unbiasedness tests, we rely on an alternative specification which is similar to that commonly used to evaluate the unbiasedness of forward prices (e.g., Fama, 1984). To develop this alternative, we note that $\bar{\sigma}_{t-1;T}$ is available in the time t information set, so we can express Eq. (4) as

$$\bar{\sigma}_{t;T} - \bar{\sigma}_{t-1;T} = E\left[\hat{\sigma}_{t;T} - \bar{\sigma}_{t-1;T} | \boldsymbol{\Phi}_t\right]. \tag{15}$$

Now, if $\bar{\sigma}_{t;T}$ is unbiased for $\hat{\sigma}_{t;T}$, then α and β should be indistinguishable from zero and one in estimating the moment vector,

$$\boldsymbol{g}_T(\alpha,\beta) = \frac{1}{NK}\sum_{t=1}^{NK}\left(\hat{\sigma}_t - \bar{\sigma}_{t-1} - \alpha - \beta(\bar{\sigma}_t - \bar{\sigma}_{t-1})\right)\boldsymbol{Z}_t, \tag{16}$$

where $\boldsymbol{Z}_t = (1, \bar{\sigma}_t - \bar{\sigma}_{t-1})'$. To the extent $\bar{\sigma}_{t;T}$ follows a random walk, its first differences will be serially uncorrelated, and Eq. (16) is free of the spurious regression problem.

While the form of (16) is convenient for unbiasedness tests, it complicates orthogonality testing. Only under the null (i.e., $\alpha = 0$, $\beta = 1$) are the errors embedded in (16) equal to the true implied volatility forecast errors. Therefore, if unbiasedness is rejected, orthogonality tests using Eq. (16) would be misspecified. Fortunately, the χ^2 test given by Eq. (14) is not biased by the spurious regression problem.[13] So, to test orthogonality, we use the original specification in Eq. (9).

3.3. GMM covariance matrix

Time-series tests of Eq. (4) also must account for the covariance nonstationarity of the implied volatility forecast error. Within an option contract month the

[13] This result was demonstrated by Fleming (1993).

intervals represented by $t;T$ and $t+1;T$ overlap. So, if volatility is forecast incorrectly 30 days from expiration, a substantial portion of this error will be replicated by the forecast 29 days from expiration. Hansen and Hodrick (1980) outline a method for consistent GMM estimation in the context of overlapping forecast intervals; but, their method is inappropriate for evaluating the implied volatility hypothesis. Here, the forecast interval coincides with the option expiration cycle which is telescoping rather than, as in Hansen and Hodrick, a moving interval of fixed length. As a result, the errors are highly correlated within a contract month, but the correlation drops off across contract months. In addition, within a contract month, the correlation is not constant as the overlap between consecutive forecasts further from expiration represents a greater portion of the forecast interval.

Appendix A develops consistent GMM estimators for this telescoping time-series structure,

$$\boldsymbol{\Omega}_T = \frac{1}{NK}\sum_{t=1}^{NK}\left(\sum_{l=-K+1}^{K-1}\phi_{t,t-l}\left(\hat{\epsilon}_t,\hat{\epsilon}_t\bar{\sigma}_t\right)'\left(\hat{\epsilon}_{t-l},\hat{\epsilon}_{t-l}\bar{\sigma}_{t-l}\right)\right), \tag{17}$$

$$\boldsymbol{D}_T = \frac{1}{NK}\sum_{t=1}^{NK}\left(1,\bar{\sigma}_t\right)'\left(1,\bar{\sigma}_t\right), \tag{18}$$

where $\hat{\epsilon}$ is the GMM residual from Eq. (9) and $\phi_{t,t-l}$ is a dummy variable equal to one when the contract months represented by observations t and $t-l$ overlap. The intuition behind $\boldsymbol{\Omega}_T$ is that the covariance structure is constant from one contract month to the next, and this estimator represents the mean covariance matrix within each contract month. In addition, this covariance matrix is robust to conditional heteroskedasticity and residual autocorrelation within the contract month.

4. Empirical results

4.1. Unbiasedness tests

This section examines the implied volatility hypothesis developed in Section 2. Specifically, we evaluate whether the implied volatility represents an unbiased forecast of the average volatility over the life of the option, and whether its forecast error is orthogonal to the market's information set. As a standard for comparison, we also consider the forecast quality of the 28-day historical volatility.[14] The historical volatility rate has been commonly used to index the forecast

[14] A 28-day interval approximates the average life of the options in the sample and it avoids day-of-the-week effects in returns by using an equal number of returns from each day of the week.

quality of the implied volatility. While it may appear to be a naive alternative, our regressions allow for both mean and slope coefficients, effectively modeling monthly volatility as an AR(1) process, consistent with Poterba and Summers (1986) and Stein and Stein (1991). Because the historical volatility rate is available in the market's information set, the implied volatility should incorporate the information conveyed by the historical volatility, as well as any additional relevant information.

A simple regression using the historical volatility to forecast realized volatility suffers from an errors-in-the-variables problem due to measurement error in the observed volatilities. This problem is more serious than for the implied volatility regressions because the historical volatility rate is a lagged value of realized volatility. As a result, the measurement error induces an MA(1) process in the observed forecast errors. In order to obtain consistent regression estimates, we apply the instrumental variable approach. The instrument we use is the Parkinson (1980) extreme-value estimator of historical volatility,

$$\bar{\sigma}^*_{h,t} = \sqrt{\frac{252}{n}\sum_{i=0}^{n-1} 0.3607 \ln(H_{t-i}/L_{t-i})^2} \tag{19}$$

where H_t and L_t, respectively, denote the high and low index levels during the close-to-close interval ending at time t. We construct this estimator using S&P 500 futures prices over the preceding $n = 28$-day period.[15] $\bar{\sigma}^*_{h,t}$ is an effective instrument because it is correlated with $\bar{\sigma}_{ht}$, but their measurement errors are independent.

We now assess the forecast quality of the implied volatility, focusing first on its unbiasedness. Recall that as a precaution against the spurious regression problem, we test unbiasedness by conducting the minimization in Eq. (10) using moment vector (16). Eqs. (17) and (18) provide the estimates of $\boldsymbol{\Omega}_T$ and $\boldsymbol{D}_T$, respectively. We estimate α and β separately for the call option and put option implied volatilities, as well as the historical volatility, using the two-step procedure of Hansen and Singleton (1982).

The unbiasedness regression results are reported in Table 1. Comparing the $\bar{R}^2$-statistics across regressions, the call option implied volatility has slightly more explanatory power than the put option implied volatility, but each of these measures exhibits a much stronger relation with realized volatility than does the historical volatility. Nonetheless, the negative intercepts for the implied volatilities

[15] For a stock index portfolio, the Parkinson estimator would be a biased volatility measure because infrequent trading misrepresents the true extreme values. Index futures prices, however, are not influenced by infrequent trading, and Wiggins (1992) demonstrates the consistency of the Parkinson estimator in this context.

Table 1
Unbiasedness tests of S&P 100 volatility proxies

f	$\hat{\alpha}$	t_{α}	$\hat{\beta}$	$t_{1-\beta}$	$\bar{R}^2$	F-stat	CS^2
Call implied	−0.0191	−3.63	0.5727	9.30	0.0273	46.35	167.13
Put implied	−0.0249	−4.58	0.6438	6.88	0.0247	42.03	90.01
28-day historical	0.0005	0.09	0.4526	2.61	0.0114	19.73	8.39

GMM estimation results are reported for the moment vector,

$$\boldsymbol{g}_{fT}(\alpha_f,\beta_f)=\frac{1}{NK}\sum_{t=1}^{NK}\left(\left(\hat{\sigma}_t-\bar{\sigma}_{f,t-1}\right)-\alpha_f-\beta_f\left(\bar{\sigma}_{f,t}-\bar{\sigma}_{f,t-1}\right)\right)\left(1,\bar{\sigma}^{*}_{f,t}-\bar{\sigma}_{f,t-1}\right)',$$

where f = S&P 100 call and put option implied and 28-day historical volatilities. $\hat{\sigma}_t$ is the annualized S&P 100 index volatility realized over the life of the option observed at t. $\bar{\sigma}^*$ is an instrumental variable (Parkinson estimator) for the historical volatility regression, otherwise $\bar{\sigma}^*=\bar{\sigma}$. K is the maximum time to option expiration and N is the number of contract months in the sample. The t-statistics reported in the table measure the significance of departures from zero and one, respectively, for $\hat{\alpha}$ and $\hat{\beta}$. $\bar{R}^2$ is the adjusted coefficient of determination. CS^2 is an asymptotic Wald test of the joint null that $\alpha=0$ and $\beta=1$ for a particular forecast. The sample is the noncrash period October 1985–April 1992 and has 1619 daily observations.

suggest that the S&P 100 option implied volatility, on average, overstates realized volatility. [16] The historical volatility appears to be less biased.

A formal unbiasedness test can be conducted using the distribution given in Eq. (12) for the GMM estimate, δ_T. $\hat{\alpha}$ and $\hat{\beta}$ should equal to zero and one under the hypothesis that $\bar{\sigma}$ is unbiased for $\hat{\sigma}$. The validity of this joint null can be assessed with a Wald test, using Eq. (12),

$$CS^2=NK(\boldsymbol{\delta}_T-\boldsymbol{\delta}_0)'\left(\boldsymbol{D}'_T\boldsymbol{\Omega}_T^{-1}\boldsymbol{D}_T\right)(\boldsymbol{\delta}_T-\boldsymbol{\delta}_0)\overset{a}{\sim}\chi^2_2,\tag{20}$$

where $\boldsymbol{\delta}_0=(0,1)'$. The CS^2-statistics reported in Table 1 indicate strong unbiasedness rejections for each volatility forecast. For the historical volatility, this finding is not unexpected. It is frequently argued that volatility follows an autoregressive process and therefore the slope coefficient should be less than one.

The finding that the call and put option implied volatilities overstate realized volatility is consistent with the findings of Jorion (1995) using foreign currency options. These results should not seem surprising given the limitations of Eq. (4) described in Section 2.1. In particular, unbiasedness may be influenced by

[16] In general, the intercept equals $\bar{Y}-\hat{\beta}\bar{X}$ where $\bar{X}$ and $\bar{Y}$, respectively, are the averages of the independent and dependent variables. For the implied volatility first differences in Eq. (16), $\bar{X}\approx 0.0001$. As a result, the large negative intercepts indicate $\bar{Y}<0$, meaning that the implied volatility, on average, exceeds realized volatility.

misspecification of the volatility process and/or the existence of early exercise opportunities. A volatility risk premium, for example, would cause the implied volatility to overstate the market's volatility expectation and, in turn, overstate future volatility.

The influence of early exercise is, perhaps, more difficult to predict. The prospect of early exercise effectively shortens the forecast horizon underlying the implied volatility. It seems implausible though that the volatility over this shorter horizon systematically exceeds the volatility over the life of the option. A more realistic explanation follows by realizing that periods of lower volatility correspond with lower option time values and therefore greater probability of early exercise. As a result, the possibilities of higher volatility receive more weight in determining the option's value, and this may cause the implied volatility to overstate realized volatility.

The remaining question is whether the implied volatility's forecast bias stems purely from mismeasurement of the market's true volatility forecast or whether it implies that options are mispriced. If the bias signals mispricing, it suggests that trading strategies which sell options (i.e., volatility) should earn abnormal profits. Fleming (1993) finds that during this period selling near-expiration, at-the-money S&P 100 options did earn large positive profits *in the absence of transaction costs*. The finding suggests the bias is truly a function of option market prices rather than entirely attributable to model misspecification. The profits for these positions, however, disappear after imposing bid/ask transaction costs. Therefore, the implied volatility bias does not seem to signal the existence of abnormal profit opportunities.

Despite the evidence of forecast bias, the regression results reported in Table 1 also suggest that both the implied and historical volatilities contain information regarding future volatility. In particular, the F-statistics indicate a significant relation between the implied volatility and realized volatility. So, despite the limitations of Eq. (4), the S&P 100 implied volatility appears to have value in forecasting future stock market volatility. The efficiency and orthogonality tests below provide a more detailed examination of this issue.

4.2. Efficiency and orthogonality tests

In addition to the unbiasedness implication, Eq. (4) suggests the implied volatility forecast errors should be orthogonal to the information set, $\boldsymbol{\Phi}_t$. They should, therefore, be orthogonal to any subset $\boldsymbol{z}_t$ of $\boldsymbol{\Phi}_t$. We evaluate this hypothesis by simply adding selected instruments from $\boldsymbol{\Phi}_t$ to $\boldsymbol{Z}_t$ in the GMM system. While it is impossible to specify every element of $\boldsymbol{\Phi}_t$ to provide a complete set of tests, the instruments selected include variables suggested in the literature to explain time-variation in volatility. In this respect, the selection criteria increases the likelihood of orthogonality rejection.

We divide the orthogonality tests into two sets. First, we examine the efficiency of $\bar{\sigma}_t$, where efficiency refers to informational efficiency relative to past forecast errors. In other words, if we let $\mu_t = \hat{\sigma}_t - \bar{\sigma}_t$ denote the forecast error at time t, $\bar{\sigma}_t$ is efficient if $E[\mu_t \mu_{t-l}] = 0$, where μ_{t-l} is available in the time t information set. The second set of tests, referred to as 'orthogonality tests,' will include notions of orthogonality other than efficiency.

The efficiency and orthogonality tests are both conducted by estimating the GMM system in Eqs. (9) and (10), where $\boldsymbol{Z}_t = (1, \bar{\sigma}_t, \boldsymbol{z}_t)'$ overidentifies the system. $\boldsymbol{z}_t$ contains the instruments of interest. If $E[\boldsymbol{z}_t \epsilon_t] \neq 0$, the elements of $\boldsymbol{z}_t$ supplement the information in $\bar{\sigma}_t$ to provide a more precise estimator of $\hat{\sigma}_t$. We test this inequality with the Hansen (1982) test of overidentifying restrictions,

$$\mathrm{OI}_n^2 = NK\boldsymbol{g}_T'(\boldsymbol{\delta}_T)\boldsymbol{\Omega}_T^{-1}\boldsymbol{g}_T(\boldsymbol{\delta}_T) \overset{a}{\sim} \chi_n^2, \tag{21}$$

where n is the dimension of $\boldsymbol{z}_t$.[17]

For the efficiency tests, we define $\boldsymbol{z}_t = \boldsymbol{\mu}_t$, where $\boldsymbol{\mu}_t$ contains the past forecast errors. As noted earlier, error correlation should exist due to the forecast horizon overlap, but the efficiency notion implicit in relation (4) restricts the order of serial correlation. Specifically, any forecast error that is entirely realized prior to t should be orthogonal to the time t forecast error.

Table 2 provides the regression results and OI^2-statistics for two monthly lagged forecast errors, μ_{t-K} and μ_{t-2K}. μ_t^l represents the forecast error using the $\hat{\alpha}$ and $\hat{\beta}$ estimates provided on the first line of the table for each forecast. Neither of the lags (for μ or μ^l) violates the hypothesized efficiency of the implied volatility. This evidence is consistent with Eq. (4). For the historical volatility, efficiency is rejected when we use μ as the instrument; but, rejections do not occur using μ^l, suggesting that additional lags seem unwarranted for the AR representation of volatility. This finding generally supports a monthly AR(1) and is consistent with the findings of Poterba and Summers (1986).

To conduct the orthogonality tests, we must first select the elements of $\boldsymbol{z}_t$. Several instruments are motivated by the various specifications used in ARCH models. In the GARCH(1,1) by Bollerslev (1986), for example, σ_{t+1}^2 is modeled as a linear function of σ_t^2 and $\psi_t^2\sigma_t^2$, where $\psi_t = (r_t - E_{t-1}[r_t])/\sigma_t$ and $E_{t-1}[r_t]$ is the time $t-1$ expectation of next period's stock market return. The GARCH-M by Engle and Bollerslev (1986) models expected returns as a function of σ_t^2 which, in the variance equation, relates σ_{t+1} and σ_t^2. The EGARCH by Nelson

[17] It may appear that since $\boldsymbol{\delta}_T$ changes when we add restrictions, OI_n^2 might be lower than if $\boldsymbol{\delta}_T$ were determined independent of $\boldsymbol{z}_t$. Alternative methods, such as those by Newey (1985) and Eichenbaum et al. (1988), rely on 'omitted variables' tests using parameter estimates isolated from $\boldsymbol{z}_t$. However, for the special case of (9) where the omission of $\boldsymbol{z}_t$ leaves an exactly identified linear system, Newey (1985) demonstrates that each of these tests is numerically equivalent.

Table 2
Efficiency tests of S&P 100 volatility proxies

	Call option implied volatility							Put option implied volatility							Historical volatility						
μ_f	$\hat{\alpha}$	t_α	$\hat{\beta}$	$t_{1-\beta}$	$\bar{R}^2$	CS^2	OI^2	$\hat{\alpha}$	t_α	$\hat{\beta}$	$t_{1-\beta}$	$\bar{R}^2$	CS^2	OI^2	$\hat{\alpha}$	t_α	$\hat{\beta}$	$t_{1-\beta}$	$\bar{R}^2$	CS^2	OI^2
Empty	0.043	1.41	0.640	2.01	0.25	16.4	–	0.048	1.51	0.598	2.27	0.23	24.3	–	0.099	3.82	0.366	3.67	0.10	14.7	–
μ_{t-K}	0.290	1.14	0.722	1.87	0.24	15.5	0.6	0.035	1.12	0.665	1.89	0.23	23.7	3.2	0.083	3.28	0.459	3.17	0.08	10.8	4.8
μ_{t-K}, μ_{t-2K}	0.059	3.07	0.537	4.33	0.24	37.0	3.5	0.077	4.17	0.417	6.24	0.20	88.1	5.4	0.098	5.45	0.347	5.84	0.09	34.3	6.6
μ^l_{t-K}	0.036	1.50	0.685	2.29	0.24	17.3	0.1	0.032	1.14	0.686	1.97	0.22	22.6	0.9	0.099	4.57	0.367	4.53	0.10	21.1	0.0
$\mu^l_{t-K}, \mu^l_{t-2K}$	0.051	2.53	0.588	3.63	0.24	29.1	1.6	0.080	4.30	0.394	6.47	0.20	101.0	5.5	0.097	6.16	0.380	6.70	0.10	45.0	0.0

The efficiency of ex ante S&P 100 volatility measures is assessed with GMM estimation of the vector

$$\boldsymbol{g}_{fT}(\alpha_{f,}\beta_f) = \frac{1}{NK}\sum_{t=1}^{NK}\left(\hat{\sigma}_t - \alpha_f - \beta_f\bar{\sigma}_{ft}\right)\left(1, \bar{\sigma}^*_{ft}, \boldsymbol{\mu}_{ft}\right)'.$$

$\bar{\sigma}_f$ represents either the daily S&P 100 call or put option implied volatility or the annualized 28-day historical volatility rate. $\hat{\sigma}$ is the annualized S&P 100 return volatility over the life of the option. K is the maximum time to option expiration and N is the number of contract months in the sample. $\bar{\sigma}^*$ is an instrumental variable (Parkinson estimator) for the historical volatility regression, otherwise $\bar{\sigma}^* = \bar{\sigma}$. $\boldsymbol{\mu}_{ft}$ contains the lagged forecast errors which are computed as either $\mu_t = \hat{\sigma}_t - \bar{\sigma}_t$, or $\mu^l_t = \hat{\sigma}_t - \hat{\alpha} - \hat{\beta}\bar{\sigma}_t$. The t-statistics for $\hat{\alpha}$ and $\hat{\beta}$, respectively, report the significance of departures from zero and one. CS^2 is an asymptotic Wald test of the joint null that $\alpha = 0$ and $\beta = 1$. OI^2 is the Hansen (1982) test statistic of overidentifying restrictions and is asymptotically distributed χ^2_n, where n is the dimension of $\boldsymbol{\mu}$. The sample is the noncrash period October 1985–April 1992.

(1991) captures asymmetry in the relation between volatility and returns, linking $|\psi_t| - E[\psi_t]$ to ln σ^2_{t+1}. These models, then, suggest several instruments that covary with volatility and, under Eq. (4), should be incorporated into the implied volatility. To focus on these parameters, r_t, $|r_t|$, and $|r_t - E_{t-1}[r_t]|$ are included in $\mathbf{z}_t$, as well as ψ_t, $|\psi_t|$, and σ_t^2. By including these instruments, a byproduct of the orthogonality tests is an indirect test of the conditional volatility unexplained by the implied volatility but explainable with ARCH.

To focus more squarely on the comparison with ARCH forecasts, we also consider orthogonality tests using a specific GARCH(1,1) model. The GARCH(1,1) variance equation is specified (using daily data) as

$$\tilde{\sigma}^2_{t+1} = a_0 + a_1 R_t^2 + b_1 \tilde{\sigma}_t^2, \tag{22}$$

where R_t is the S&P 100 return, demeaned and adjusted for first-order autocorrelation. Fitting the model across our sample period yields the estimates $\hat{a}_0 =$ 0.000006, $\hat{a}_1 = 0.05$, and $\hat{b}_1 = 0.89$. We now use the fitted model to generate volatility forecasts. Let $\tilde{\sigma}^2_{t+j|t}$ represent the forecast at time t of the variance j days into the future.[18] We define the GARCH(1,1) forecast of the average volatility over the life of the option as $\tilde{\sigma}_{t+1;T} = \sqrt{\frac{1}{T-t-1}\sum_{j=1}^{T-t}\tilde{\sigma}^2_{t+j|t}}.$[19] By including this instrument in the orthogonality tests, we incorporate an additional feature of the ARCH class of models. In particular, because the j-step-ahead forecast depends on the current level of volatility and its speed of adjustment toward the long-term mean, this dynamic is captured in the GARCH(1,1) forecast of the average volatility.

A number of other orthogonality instruments, in addition to the ARCH subset, are motivated in the literature. Shanken (1990), for example, finds that short-term interest rates and a January dummy variable predict stock market volatility. We include both of these parameters in the instrument vector, $\mathbf{z}_t$. Two interest rate parameters are considered, the effective yield on the Treasury bill nearest 30 days to maturity, and the 30-day rate scaled by the volatility level. We set the January dummy equal to one when $\hat{\sigma}$ is determined primarily from returns realized in January. In addition, Gallant et al. (1992) find evidence that conditional volatility is related to lagged trading volume, so we also include detrended stock market volume (constructed from the S&P's *Security Price Index Record*) in the orthogonality tests. Both levels and logarithms of volume are considered.

Table 3 summarizes the orthogonality regression results. Overall, few of the OI^2-statistics approach their critical values for either the historical or implied volatilities. At first, this result may seem to indicate the lack of a relation between

[18] The algorithm used to compute j-step-ahead forecasts follows Hamilton (1994), Chap. 21.

[19] This definition conforms with the instrument used by Jorion (1995).

Table 3
Orthogonality tests of S&P 100 volatility proxies

	Call option implied volatility							Put option implied volatility							Historical volatility						
$\boldsymbol{z}_f$	$\hat{\alpha}$	t_α	$\hat{\beta}$	$t_{1-\beta}$	$\overline{R}^2$	CS^2	OI^2	$\hat{\alpha}$	t_α	$\hat{\beta}$	$t_{1-\beta}$	$\overline{R}^2$	CS^2	OI^2	$\hat{\alpha}$	t_α	$\hat{\beta}$	$t_{1-\beta}$	$\overline{R}^2$	CS^2	OI^2
Empty	0.047	1.60	0.622	2.20	0.25	16.0	–	0.048	1.56	0.597	2.31	0.23	24.5	–	0.099	3.91	0.369	3.72	0.10	15.4	–
σ^0_{-f}	0.030	1.26	0.717	1.97	0.24	14.7	0.8	0.031	1.17	0.689	2.03	0.23	23.0	1.0	0.101	3.98	0.322	3.99	0.10	16.0	7.9
$\sigma^0_{-f},\sigma^K_{-f}$	0.052	2.79	0.573	4.12	0.24	40.6	3.4	0.048	2.23	0.581	3.57	0.23	45.3	2.7	0.094	3.59	0.384	3.53	0.10	12.9	9.2
$\tilde{\sigma}$	0.038	1.33	0.659	2.00	0.24	19.5	2.1	0.041	1.33	0.627	2.14	0.23	28.3	2.4	0.111	4.98	0.270	5.11	0.10	26.2	1.1
r^{28}	0.053	1.99	0.588	2.58	0.25	16.0	0.3	0.053	1.91	0.569	2.67	0.23	24.3	0.2	0.100	4.17	0.361	3.89	0.10	18.1	0.0
$\|r^{28}\|$	0.046	1.59	0.622	2.20	0.25	19.1	0.1	0.048	1.56	0.589	2.35	0.23	30.6	0.4	0.105	4.45	0.330	4.14	0.10	20.7	0.6
$r^{28},\|r^{28}\|$	0.053	1.99	0.582	2.62	0.24	19.6	0.4	0.055	1.96	0.555	2.78	0.23	30.4	0.6	0.105	4.57	0.332	4.19	0.10	22.3	0.6
$\|r^{28}-\bar{r}^{28}\|$	0.050	1.73	0.596	2.39	0.25	20.3	0.6	0.054	1.80	0.556	2.62	0.23	29.6	0.8	0.101	4.08	0.355	3.87	0.10	16.9	0.2
r^{30}_f	0.050	1.99	0.606	2.67	0.25	17.3	0.0	0.051	1.96	0.581	2.82	0.23	25.7	0.0	0.101	4.22	0.354	4.05	0.10	17.9	0.1
$r^{30}_f/\overline{\sigma}_c$	0.047	1.51	0.615	2.09	0.25	17.1	0.6	0.045	1.40	0.611	2.12	0.23	25.7	0.3	0.100	3.77	0.360	3.61	0.10	14.4	2.0
$r^{30}_f/\overline{\sigma}_h$	0.039	1.17	0.662	1.70	0.24	15.6	0.8	0.040	1.14	0.635	1.81	0.23	24.7	0.8	0.095	3.66	0.385	3.47	0.10	13.6	1.0
Jan	0.051	1.85	0.599	2.46	0.25	16.1	1.6	0.058	2.08	0.541	2.92	0.23	27.1	1.7	0.114	5.53	0.258	5.63	0.10	31.8	1.0
V^{28}	0.049	1.81	0.610	2.46	0.25	17.2	0.1	0.051	1.78	0.582	2.61	0.23	26.2	0.2	0.101	3.62	0.355	3.44	0.10	13.4	0.9
$\ln V^{28}$	0.048	1.69	0.613	2.29	0.25	16.0	0.1	0.050	1.63	0.588	2.38	0.23	24.5	0.1	0.101	4.06	0.354	3.84	0.10	16.8	0.4
ψ	0.049	1.75	0.611	2.31	0.25	15.4	0.0	0.048	1.68	0.597	2.44	0.23	25.5	0.0	0.097	4.02	0.376	3.77	0.10	16.8	0.1
$\|\psi\|$	0.049	1.74	0.605	2.40	0.25	19.5	0.1	0.053	1.77	0.567	2.60	0.23	29.3	0.3	0.106	4.38	0.321	4.13	0.10	19.6	0.7
$\psi,\|\psi\|$	0.051	1.84	0.598	2.46	0.25	19.0	0.1	0.051	1.84	0.573	2.67	0.23	31.8	0.3	0.105	4.77	0.326	4.37	0.10	24.5	0.8
$\overline{\sigma}^2_h$	0.040	1.76	0.664	2.55	0.24	17.3	0.1	0.040	1.60	0.641	2.55	0.23	26.3	0.2	0.098	3.92	0.373	3.75	0.10	15.5	0.0

The orthogonality of ex ante S&P 100 volatility measures is assessed with GMM estimation of the vector

$$\boldsymbol{g}_{fT}(\alpha_f,\beta_f)=\frac{1}{NK}\sum_{t=1}^{NK}\left(\hat{\sigma}_t-\alpha_f-\beta_f\overline{\sigma}_{ft}\right)\left(1,\overline{\sigma}^*_{ft},\boldsymbol{z}_{ft}\right)'.$$

$\boldsymbol{z}_{ft}$ contains the various orthogonality instruments: $\sigma^d_{-f,t}$, observed at $t-d$, is the alternative volatility measure ($\overline{\sigma}_c$ in the case of historical volatility); $\tilde{\sigma}_t$ is the GARCH(1,1) volatility forecast; r^{28}_t is the S&P 100 return for the 28-day period ending at t; $r^{30}_{f,t}$ is the effective yield on the T-bill closest to 30 days from maturity; Jan_t is a January dummy variable; V^{28}_t is the total (detrended) trading volume over the previous 28 calendar days; ψ_t equals $(r^{28}_t-\bar{r})/\overline{\sigma}_{ht}$, where $\bar{r}$ is the mean index return during the sample; and, $\overline{\sigma}^2_{ht}$ is the squared historical volatility. The t-statistics for $\hat{\alpha}$ and $\hat{\beta}$ indicate the significance of departures from zero and one, respectively. CS^2 is an asymptotic test of the joint null that $\alpha=0$ and $\beta=1$ for a particular forecast. The orthogonality statistics report the Hansen (1982) test of overidentifying restrictions and are asymptotically distributed χ^2_n, where n is the dimension of z. The sample is the noncrash period October 1985–April 1992.

our selected instruments and S&P 100 volatility. Recall, however, that Eq. (9) evaluates forecast quality over the life of the option. As a result, the forecast horizon is not a constant interval, but is telescoping with the option expiration cycle. This variation in the forecast horizon likely reduces the informativeness of our instrument set. [20] Indeed, the significance of these parameters increases when we consider forecast quality over constant (and shorter) forecast horizons in Section 5.

The most interesting feature of Table 3 concerns the implied and historical volatilities as instruments. When we include the historical volatility rate in the orthogonality tests for the call and put implied volatilities, the OI^2-statistics indicate nonrejections. Conversely, including the implied volatility in the historical volatility regressions induces strong orthogonality rejections. This evidence suggests that the implied volatility contains the information conveyed by the historical volatility, as well as additional information that is relevant for predicting volatility. In this sense, the implied volatility represents the dominant forecast.

It is also interesting to consider the GARCH(1,1) forecasts. Although not reported in the table, the GARCH forecasts display less bias than either the implied or historical volatilities. In fact, the CS^2-statistic is less than 2.0 which indicates unbiasedness is not rejected. Its explanatory power, however, is comparable to the historical volatility, and, in Table 3, it fails to generate orthogonality rejections in the implied and historical volatility regressions. When these tests are reversed (not shown in the table), including the call option implied volatility as an instrument in the GARCH regression produces a strong orthogonality rejection ($OI^2 = 7.2$), but instrumenting with the historical volatility does not yield rejection ($OI^2 = 1.2$). So, while the GARCH(1,1) and historical volatility comparison is inconclusive, the implied volatility subsumes the information contained in the GARCH(1,1).

At this point, we can summarize the validity of Eq. (4) for S&P 100 options. In Section 4.1, we found that the S&P 100 implied volatility systematically overstates future stock market volatility. In this section, however, we find that, despite its bias, the implied volatility covaries with future volatility and its forecast error is orthogonal to parameters often used to forecast volatility. It appears, then, that the limitations in applying Eq. (4) to S&P 100 options do not negate the implied volatility's predictive power or its orthogonality properties. This seems reasonable provided that the misspecification of the volatility process and early exercise opportunities have a fairly stable effect on the implied volatility.

These results may seemingly contradict earlier research for the S&P 100 options, but the methodology used in this study differs in two important respects. First, the econometric analysis explicitly addresses the spurious regression prob-

[20] Instruments which proxy for calendar effects, such as time to expiration and its square root (expressed in calendar days or trading days), however, do not produce orthogonality rejections.

lem that is induced by volatility persistence and tends to overstate the performance of the historical volatility relative to the implied volatility. Second, the implied volatility in this study is computed from a window of option prices which minimizes the measurement error and, presumably, improves its measured forecast quality. In contrast to the evidence for S&P 100 options, both the bias and orthogonality results are consistent with the evidence of Jorion (1995) for currency options.

5. Constant horizon volatility prediction

5.1. Monthly volatility

We now consider the implied volatility's quality as a forecast over a fixed horizon, independent of the option expiration cycle. We first examine the implied volatility as a forecast of monthly volatility. The implied volatility's performance in this setting is of interest for two reasons. First, volatility predictions over the life of the option are not useful in a context requiring daily estimates over, say, the next month; option contracts expiring in exactly one month would be unavailable as a daily series. One might, however, use daily implied volatilities to proxy such a series. Second, a strong relation exists between the ex post volatility estimate $\hat{\sigma}_{t;T}$, computed over the life of the option, and $\hat{\sigma}_t^{28}$, computed over the next 28 calendar days, because they include largely the same returns observations. Therefore, the implied volatility's forecast performance over a fixed, 1-month horizon should be related to its performance over the life of the option.

To evaluate performance over a 1-month horizon, we simply respecify the orthogonality tests used in Section 4.2. In particular, the realized volatility over the next 28 calendar days, $\hat{\sigma}_t^{28}$, replaces $\hat{\sigma}_{t;T}$ in Eq. (9). Unlike in Section 4.2, the error vector can now be assumed covariance stationary because the forecast horizon is fixed and the overlap is constant rather than telescoping. Therefore, GMM estimation is conducted with a constant 19-period overlap (expressed in trading days).

Table 4 presents the orthogonality results,[21] using the same sequence of instruments as in Table 3. The significance of these instruments is now more apparent than in the earlier context of a telescoping forecast horizon. For the historical volatility, orthogonality approaches rejection with lagged S&P 100 returns (absolute values), the January dummy variable, and ψ (absolute values), the ARCH parameter. In addition, the implied volatility induces even stronger historical volatility rejections than in Section 4.2. Conversely, the relation between

[21] The unbiasedness and efficiency tests are not presented here. The results are similar, however, to those for the volatility over the life of the option in that unbiasedness is strongly rejected and efficiency is not rejected.

Table 4
GMM estimation results for 28-day S&P 100 volatility

Panel A: basic regression ($\boldsymbol{z} = \underline{0}$)

f	$\hat{\alpha}$	t_{α}	$\hat{\beta}$	$t_{1-\beta}$	$\bar{R}^2$	CS^2
Call implied	0.047	1.81	0.607	2.53	0.2914	27.58
Put implied	0.048	1.83	0.580	2.73	0.2716	40.49
28-day historical	0.096	4.31	0.369	4.11	0.1342	18.97

Panel B: orthogonality results

$\boldsymbol{z}_t$	Call implied	Put implied	Historical
σ_{-f}^{0}	0.49	0.50	8.40
σ_{-f}^{0}, σ_{-f}^{28}	2.42	1.57	10.65
$\tilde{\sigma}$	1.46	1.58	0.33
r^{28}	1.12	0.94	0.47
$\lvert r^{28}\rvert$	0.92	0.21	3.73
r^{28},$\lvert r^{28}\rvert$	1.46	0.94	3.76
$\lvert r^{28} - \bar{r}^{28}\rvert$	0.06	0.35	1.28
r_f^{30}	0.18	0.17	0.21
$r_f^{30}/\bar{\sigma}_c$	0.64	0.23	2.90
$r_f^{30}/\bar{\sigma}_h$	0.73	0.70	1.00
Jan	3.43	3.38	2.39
V^{28}	0.12	0.21	1.07
$\ln V^{28}$	0.08	0.10	0.47
ψ	0.71	0.46	0.18
$\lvert\psi\rvert$	0.08	0.00	2.32
ψ,$\lvert\psi\rvert$	1.17	0.51	3.23
$\bar{\sigma}_h^2$	0.02	0.02	0.00

GMM estimation results are reported for the moment vector,

$$\boldsymbol{g}_{fT}(\alpha_{f,}\beta_f) = \frac{1}{NK}\sum_{t=1}^{NK}\left(\hat{\sigma}_t^{28} - \alpha_f - \beta_f\bar{\sigma}_{ft}\right)\left(1,\bar{\sigma}_{ft}^{*},\boldsymbol{z}_{ft}\right)',$$

where $f =$ S&P 100 call and put option implied and 28-day historical volatilities, $\hat{\sigma}_t^{28}$ is the annualized S&P 100 index volatility realized over the 28-day period subsequent to t. The elements of z_{ft} include: the alternative forecast, $\sigma_{-f,t}^{d}$, observed at $t-d$; $\tilde{\sigma}_t$, the GARCH(1,1) forecast; r_t^{28}, the 28-day S&P 100 return ending at t; $r_{f,t}^{30}$, the 30-day riskless rate; Jan_t, a January dummy variable; V_t^{28}, monthly lagged stock market volume; ψ_t, equal to $(r_t^{28} - \bar{r})/\bar{\sigma}_{ht}$; and $\bar{\sigma}_{ht}^2$, the squared historical volatility. The t-statistics for $\hat{\sigma}$ and $\hat{\beta}$ indicate the significance of departures from zero and one, respectively. CS^2 is an asymptotic test of the joint null that $\alpha = 0$ and $\beta = 1$ for a particular forecast. The orthogonality statistics report the Hansen (1982) test of overidentifying restrictions and are asymptotically distributed χ_n^2, where n is the dimension of z. The sample is the noncrash period October 1985–April 1992.

the historical volatility and the implied volatility forecast error is much weaker. For the implied volatility, the January dummy variable is the only parameter to generate test statistics which approach rejection, and these results are attributable to the large implied volatilities observed in January 1988 just after the crash.

Absent this, the implied volatility exhibits orthogonality in the context of a constant, 1-month forecast horizon.

5.2. Daily volatility

The results above can be generalized to even finer time horizons. With a daily forecast horizon, we estimate the subsequent day's S&P 100 volatility rate as

$$\hat{\sigma}_t^1 = \sqrt{252(r_{t+1} - \bar{r})^2}, \tag{23}$$

where r_{t+1} is the index return and $\bar{r}$ is the mean return over the sample. The GMM system used to test orthogonality is again the same, but with $\hat{\sigma}_{t;T}$ replaced by $\hat{\sigma}_t^1$.

Table 5 reports the regression results. The analysis includes a second measure of historical volatility, $\hat{\sigma}_{t-1}^1$, but this 1-day measure fails to demonstrate much explanatory power. This result may seem to contradict studies such as Ding et al. (1993) that find serial correlation in absolute returns. The t-statistic reported in the table (for $1-\beta$), however, does imply that the estimate of β is significantly positive ($t_{\beta=0} > 3$). This finding is consistent with the level of first-order serial correlation in absolute returns for this sample period (5%).

Because the serial correlation in $\hat{\sigma}^1$ is relatively low, the regressions reported in Table 5 are largely unaffected by the spurious regression problem. As a result, the coefficient estimates in Panel A are generally consistent with the unbiasedness results from Section 4.1. Although the implied volatility intercepts are not significantly different from zero, the CS^2-statistics indicate large bias. Comparing the $\bar{R}^2$-statistics across regressions, the implied volatility exhibits a much stronger explanatory power of next-day volatility than does the 28-day or 1-day historical volatility.

In the orthogonality tests (Panel B), we use the same set of instruments as before, but with the volume, the January dummy, and S&P 100 return variables defined by a single day (rather than 28 days). Consistent with the earlier results, orthogonality rejections occur only for the historical volatilities. Most significantly, the implied volatility as an instrument produces very strong rejections for both the 28-day and 1-day historical volatilities. The scaled interest rate and most of the ARCH parameters, including the GARCH(1,1) forecasts, also induce orthogonality rejections. In addition, the volume parameters and lagged (and absolute) returns seem important for the 1-day historical volatility. Conversely, for the call and put option implied volatilities, none of the parameters approach rejection. The implied volatility, then, appears to convey useful information regarding tomorrow's volatility rate.

5.3. Informational time decay

The implied volatility's quality as a forecast of next-day volatility is related to the relatively short-term perspective of the S&P 100 option market. Traders often argue that option prices, rather than conforming to any particular valuation model,

Table 5
GMM estimation results for next-day S&P 100 volatility

Panel A: basic regression ($\boldsymbol{z} = \underline{0}$)

f	$\hat{\alpha}$	t_{α}	$\hat{\beta}$	$t_{1-\beta}$	$\overline{R}^2$	CS^2
Call implied	0.012	0.86	0.594	4.71	0.0557	482.24
Put implied	0.008	0.55	0.598	4.55	0.0576	583.21
28-day historical	0.053	5.10	0.411	8.52	0.0285	223.97
1-day historical	0.102	20.69	0.122	22.16	−0.0060	507.13

Panel B: orthogonality results

$\boldsymbol{z}_t$	Call implied	Put implied	28-day historical	1-day historical
σ_{-f}^{0}	0.08	0.64	21.01	36.12
σ_{-f}^{0}, σ_{-f}^{1}	0.49	1.22	20.97	36.12
$\tilde{\sigma}$	0.61	1.33	2.63	14.02
r^1	0.25	1.50	2.20	1.58
$\lvert r^1 \rvert$	0.65	0.96	0.54	10.05
$r^1, \lvert r^1 \rvert$	1.25	3.60	3.49	10.63
$\lvert r^1 - \bar{r}^1 \rvert$	0.69	0.87	0.45	9.80
r_f^{30}	0.32	0.24	0.33	1.99
$r_f^{30} / \overline{\sigma}_c$	0.38	0.04	7.91	22.52
$r_f^{30} / \overline{\sigma}_h$	0.07	0.26	1.54	9.91
Jan	0.08	0.09	0.22	0.04
V^1	0.04	0.05	0.01	1.25
$\ln V^1$	0.85	0.68	0.78	2.33
ψ	0.34	2.08	3.29	2.98
$\lvert \psi \rvert$	1.29	1.29	0.46	19.24
$\psi, \lvert \psi \rvert$	2.00	4.25	4.30	20.90
$\overline{\sigma}_h^2$	0.00	0.25	2.39	16.84

GMM estimation results are reported for the moment vector,

$$\boldsymbol{g}_{fT}(\alpha_f, \beta_f) = \frac{1}{NK} \sum_{t=1}^{NK} \left(\hat{\sigma}_t^1 - \alpha_f - \beta_f \overline{\sigma}_{ft} \right) \left(1, \overline{\sigma}_{ft}^{*}, \boldsymbol{z}_{ft} \right)',$$

where $f =$ S&P 100 call and put option implied and 28-day and 1-day historical volatilities. $\hat{\sigma}_t^1$ is the annualized volatility during the following day, $\sqrt{252(r_{t+1} - \bar{r})^2}$ where r_t and $\bar{r}$ are the 1-day S&P 100 return and mean return, respectively. The 1-day historical volatility is $\hat{\sigma}_{t-1}^1$. The elements of z_{ft} include: $\sigma_{-f,t}^{d}$, observed at $t-d$; $\tilde{\sigma}_t$, the GARCH(1,1) forecast; r_t^1, the daily lagged S&P 100 return; $r_{f,t}^{30}$, the 30-day riskless rate; Jan_t, a January dummy variable; V_t^1, daily lagged stock market volume; ψ_t, equal to $(r_t^1 - \bar{r})/\overline{\sigma}_{ht}$; and, $\overline{\sigma}_{ht}^2$, the squared historical volatility. The t-statistics for $\hat{\alpha}$ and $\hat{\beta}$ indicate the significance of departures from zero and one, respectively. CS^2 is an asymptotic test of the joint null that $\alpha = 0$ and $\beta = 1$ for a particular forecast. The orthogonality statistics report the Hansen (1982) test of overidentifying restrictions and are asymptotically distributed χ_n^2, where n is the dimension of z. The sample is the noncrash period October 1985–April 1992.

simply reflect the 'current state of the market'. To a degree, this perspective reflects the decreasing accuracy with which the more distant future can be predicted. As a result, we might expect the implied volatility to be most strongly related to today or tomorrow's volatility, and decreasingly related to volatility further into the future.

Fig. 1 illustrates the time decay of the information content of the implied and historical volatilities. The decay is estimated using the specification from Section 5.2, but with $\hat{\sigma}^1_{t+j}$ as the dependent variable for $j = 0,\dots,9$ days into the future. In the figure, as j increases, the relation between the time t volatility forecast and

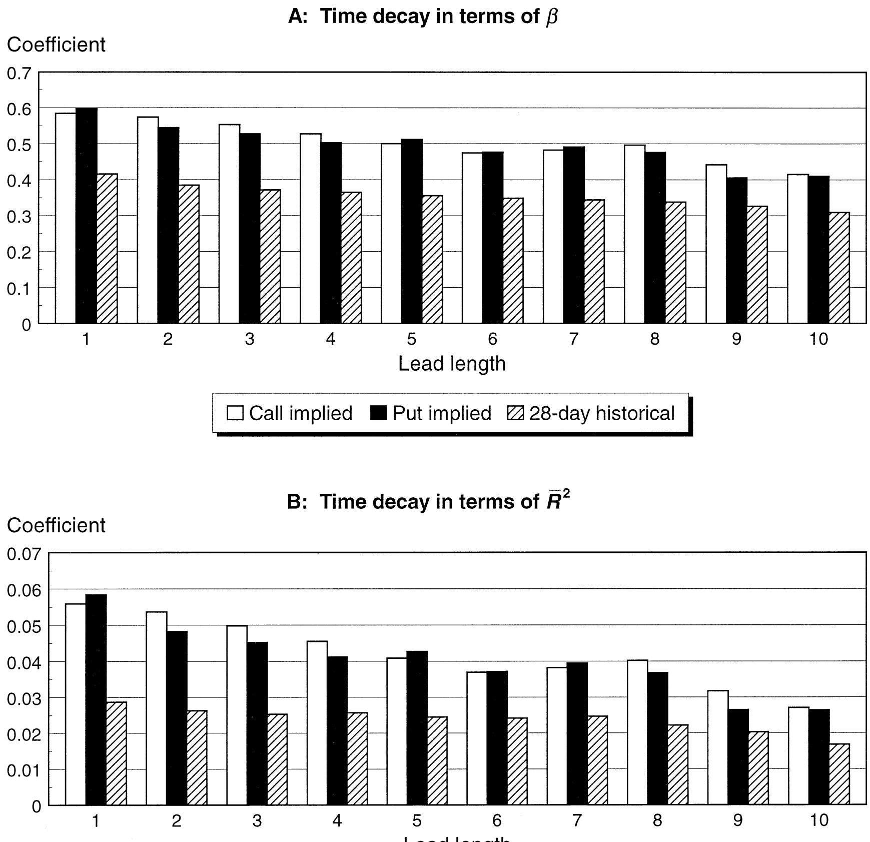

Fig. 1. The information time decay in the S&P 100 option implied volatility. The figure illustrates the variation in the β and $\overline{R}^2$ parameters of the specification,

$$\boldsymbol{g}_{fT}\left(\alpha_{fj},\beta_{fj}\right) = \frac{1}{NK}\sum_{t=1}^{NK}\left(\hat{\sigma}^1_{t+j} - \alpha_{fj} - \beta_{fj}\overline{\sigma}_{ft}\right)\left(1,\overline{\sigma}^*_{ft}\right)',$$

where $f =$ S&P 100 call and put option implied and 28-day historical volatilities. $\hat{\sigma}^1_{t+j}$ annualized S&P 100 index volatility realized over the trading day $j = 0,\dots,9$ days subsequent to t, $\sqrt{252\left(r_{t+j+1} - \bar{r}\right)^2}$, where r_t and $\bar{r}$ represent the 1-day S&P 100 index return and mean return, respectively. $\overline{\sigma}^*$ is an instrumental variable (Parkinson estimator) for the historical volatility regression, otherwise, $\overline{\sigma}^* = \overline{\sigma}$. The sample is the noncrash period October 1985–April 1992.

the volatility realized at $t+j$ slowly decreases. The rate of decay for the historical volatility simply reflects the time-series structure of volatility, and is consistent with the general properties of a GARCH model. For the implied volatility, however, the decay is faster which is consistent with option traders' short-term focus.[22] The implied volatility most strongly reflects the volatility rate expected for the near future and exhibits a weaker relation with more distant volatility.

6. Conclusions

This study investigates whether the S&P 100 index option implied volatility represents an unbiased forecast of stock market volatility and whether its forecast error is orthogonal to the market's information set. Empirically, both the S&P 100 call and put option implied volatilities are biased forecasts. Although this bias may suggest option market inefficiency, it may also stem from misspecification of the volatility process in the option valuation model and/or the existence of early exercise opportunities. The degree of bias, however, does not seem large enough to signal the existence of abnormal trading profits.

Despite the implied volatility's bias, a linear model using only the implied volatility appears to deliver a quality forecast of ex post volatility. The implied volatility is efficient with respect to its past forecast errors, and its forecast errors are orthogonal to parameters often linked to conditional volatility, including the historical volatility rate and parameters embedded in ARCH specifications. None of these parameters can explain the component of volatility that is unexplained by the implied volatility. These results are valid for a forecast horizon equal to the life of the option, as well as for daily and monthly horizons.

Collectively, this evidence supports empirical use of the implied volatility as a proxy for conditional volatility. As a result, the S&P 100 implied volatility may be valuable in at least three forms of research. First, it may be used simply as an index of market sentiment. For example, the CBOE's Market Volatility Index, which is based on S&P 100 implied volatility, provides a real-time measure of expected stock market volatility.[23] Second, because the implied volatility is strongly linked to ex ante market volatility expectations, it may be useful as an alternative method for evaluating asset pricing models. Finally, the implied volatility itself may be related to expected returns. Merton (1980), for example, specifies a relation between conditional volatility and expected returns. So, to the extent the implied volatility adequately proxies volatility, it may be helpful in predicting stock market returns.

[22] This pattern is also consistent with the implications of the GARCH option pricing model of Duan (1995).

[23] See Whaley (1993) and Fleming et al. (1995), respectively, for a description of the index and evidence regarding its empirical properties.

Acknowledgements

This paper is developed from the second chapter of my dissertation. I gratefully acknowledge my committee, Cam Harvey, David Hsieh, Tom Smith, George Tauchen, and Bob Whaley (chairman), as well as Cliff Ball, David Bates, Linda Canina, Ted Day, Fernando Diz, Fred Feinberg, Steve Figlewski, Doug Foster, David Ikenberry, Chris Kirby, Craig Lewis, Barbara Ostdiek, and Mark Rubinstein for their discussions and comments on earlier drafts. In addition, the paper has benefited from the useful suggestions of three anonymous referees, Christian Wolff (the editor), and seminar participants at the University of Colorado, Duke University, Georgia Institute of Technology, University of Illinois, University of Iowa, Northwestern University, University of Pennsylvania, Purdue University, Rice University, University of Texas at Dallas, Vanderbilt University, and the 1993 American Finance Association meetings.

Appendix A. Consistent GMM estimators with telescoping observations

The basic GMM system used in this study is expressed by Eqs. (9)–(13) in the text. This appendix develops consistent GMM estimators for $\boldsymbol{\Omega}$ and $\boldsymbol{D}$ when the data exhibit a telescoping overlap structure.

Consider Eq. (9) where $\boldsymbol{Z}_t=(1,\hat{\sigma}_t)'$ is the vector of instrumental variables. Cumby et al. (1983) demonstrate,

$$\boldsymbol{\Omega}=\lim_{NK\to\infty}\frac{1}{NK}E\{(\mathbf{1},\boldsymbol{X})'\boldsymbol{\epsilon}\boldsymbol{\epsilon}'(\mathbf{1},\boldsymbol{X})\},\tag{A1}$$

$$\boldsymbol{D}=\text{plim}\frac{1}{NK}(\mathbf{1},\boldsymbol{X})'(\mathbf{1},\boldsymbol{X}),\tag{A2}$$

where $\mathbf{1}$, $\boldsymbol{X}$, and $\boldsymbol{\epsilon}$ are $NK\times 1$ vectors with $1,\bar{\sigma}_t$, and $\epsilon_t=\hat{\sigma}_t-\alpha-\beta\bar{\sigma}_t$, respectively, as the t-th elements. We now develop consistent estimators for the case of telescoping overlap by generalizing the approach of Jagannathan (1985) and Hodrick and Srivastava (1987).

Suppose the data consist of non-overlapping option contract months $n=1,\dots,N$, and, within each contract month, we have options $k=K,\dots,1$ days to expiration. In contract month notation, time $t=(n-1)K+k$. Define $\boldsymbol{w}_k^n=\epsilon_{(n-1)K+k}(1,\bar{\sigma}_{(n-1)K+k})$. $\boldsymbol{w}_k^n$ is not covariance stationary due to the telescoping forecast interval, but $\boldsymbol{W}^n=(\boldsymbol{w}_K^{n\prime},\dots,\boldsymbol{w}_1^{n\prime})'$ encompasses the entire contract month and can be reasonably assumed covariance stationary. Now, let $\boldsymbol{\theta}_{i,j}=\mathrm{E}[\boldsymbol{w}_i^n\boldsymbol{w}_j^{n\prime}]$ represent the covariance between the elements of $\boldsymbol{W}$ observed i and j days from expiration.

The forecast interval for observation t overlaps successive observations through option expiration, so correlation exists among terms within the same contract month. Conversely, the forecast errors for preceding contract months are known at time t, so we impose a zero correlation among terms of different contract months,

$$\boldsymbol{\Omega} = \lim_{N\to\infty} \frac{1}{NK} E\left\{ \sum_{n=1}^{N} \sum_{i=1}^{K} \sum_{j=1}^{K} \begin{pmatrix} \epsilon_{(n-1)K+i} \\ \epsilon_{(n-1)K+i}\bar{\sigma}_{(n-1)K+i} \end{pmatrix} \begin{pmatrix} \epsilon_{(n-1)K+j} \\ \epsilon_{(n-1)K+j}\bar{\sigma}_{(n-1)K+j} \end{pmatrix}' \right\}. \tag{A3}$$

In terms of the $\boldsymbol{w}^n$ representation,

$$\boldsymbol{\Omega} = \lim_{N\to\infty} \frac{1}{NK} E\left\{ \sum_{n=1}^{N} \sum_{i=1}^{K} \sum_{j=1}^{K} \boldsymbol{w}_i^n \boldsymbol{w}_j^{n\prime} \right\} \tag{A4}$$

$$= \frac{1}{K} \sum_{i=1}^{K} \sum_{j=1}^{K} \boldsymbol{\theta}_{i,j}, \tag{A5}$$

as K is fixed and N becomes large. To see the intuition of this result, define $\phi_{i,i-l}$ as a dummy variable equal to one for $1 \le i \le K$ and $1 \le i-l \le K$, and equal to zero otherwise. Then, the summation can be extended to an equal 'lag' on either side of observation i,

$$\boldsymbol{\Omega} = \frac{1}{K} \sum_{i=1}^{K} \left(\sum_{l=-K+1}^{K} \phi_{i,i-l} \boldsymbol{\theta}_{i,i-l} \right). \tag{A6}$$

For $l=0$, $\boldsymbol{\Omega}_{i,i-l}$ exclusively represents contemporaneous correlation within $\boldsymbol{W}$. The remainder of the summation $(l = \ K+1,\ldots,K-1;\ l \neq 0)$ incorporates the effect of a $(K-1)$-period overlap between successive observations, as in the approach of Hansen (1982). $\phi_{i,i-l}$, however, limits the overlap in (A6) to observations from the same option contract month. This characterization, then, illustrates that after computing the effect of the overlap, $\boldsymbol{\Omega}$ represents the mean covariance matrix associated with observations throughout the life of the option expiration cycle. For the special case in which the summation over l in Eq. (A6) is limited to $l=0$, the influence of serial correlation within the contract month is eliminated and $\boldsymbol{\Omega}$ is identical to that developed by Hodrick and Srivastava (1987).

To estimate $\boldsymbol{\Omega}$, let $\hat{\epsilon}$ be the GMM residual from Eq. (9) and define

$$\boldsymbol{\Omega}_T = \frac{1}{NK} \sum_{t=1}^{NK} \left(\sum_{l=-K+1}^{K-1} \phi_{t,t-l} \left(\hat{\epsilon}_t, \hat{\epsilon}_t \bar{\sigma}_t \right)' \left(\hat{\epsilon}_{t-l}, \hat{\epsilon}_{t-l} \bar{\sigma}_{t-l} \right) \right). \tag{A7}$$

Given the development of (A6), it follows that if the parameter space is compact and the continuity conditions given in Hansen (1982) are satisfied, then $\hat{\epsilon}_t$ is consistent for ϵ and $\boldsymbol{\Omega}_T$ is consistent for $\boldsymbol{\Omega}$.[24] $\boldsymbol{\Omega}_T$ also is robust to conditional heteroskedasticity and error autocorrelation within the contract month.

[24] In Section 2.2, we constructed the implied volatility series by switching contract months once the nearby month is within 15 days of expiration. For the next 15 days, the forecast errors for the expiring month have not been realized, causing overlap with the errors from the following contract month. To incorporate this pattern, we simply redefine $\phi_{t,t-l}$ in Eq. (A7). Now, $\phi_{t,t-l}$ equals one when the forecast horizons at t and $t-l$ overlap and, otherwise, equals zero.

By similar arguments, it can be shown from Eq. (A2) that

$$\boldsymbol{D}_T = \frac{1}{NK} \sum_{t=1}^{NK} (1, \bar{\sigma}_t)'(1, \bar{\sigma}_t) \tag{A8}$$

is consistent for $\boldsymbol{D}$.

References

Beckers, S., 1981. Standard deviations implied in option prices as predictors of future stock price variability. J. Banking Finance 5, 363–381.

Black, F., Scholes, M.S., 1973. The pricing of options and corporate liabilities. J. Polit. Econ. 81, 637–659.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. J. Econ. 31, 307–327.

Bollerslev, T., Chou, R.Y., Kroner, K.F., 1992. ARCH modeling in finance: a review of the theory and empirical evidence. J. Econ. 52, 1–59.

Canina, L., Figlewski, S., 1993. The information content of implied volatility. Rev. Financial Studies 6, 659–681.

Chesney, M., Scott, L., 1989. Pricing European currency options: a comparison of the modified Black–Scholes model and a random variance model. J. Financial Quant. Anal. 24, 267–284.

Chiras, D.P., Manaster, S., 1978. The information content of option prices and a test of market efficiency. J. Financial Econ. 6, 213–234.

Cumby, R.E., Huizinga, J., Obstfeld, M., 1983. Two-step, two-stage least squares estimation in models with rational expectations. J. Econ. 21, 333–355.

Day, T.E., Lewis, C.M., 1988. The behavior of the volatility implicit in the prices of stock index options. J. Financial Econ. 22, 103–122.

Day, T.E., Lewis, C.M., 1992. Stock market volatility and the information content of stock index options. J. Econ. 52, 267–287.

Dickey, D., Fuller, W., 1979. Distribution of the estimators for autoregressive time series with a unit root. J. Am. Stat. Assoc. 74, 427–431.

Dickey, D., Fuller, W., 1981. Likelihood ratio statistics for autoregressive time series with a unit root. Econometrica 49, 1057–1072.

Ding, Z., Granger, C.W.J., Engle, R.F., 1993. A long memory property of stock market returns and a new model. J. Empirical Finance 1, 83–106.

Duan, J.-C., 1995. The GARCH option pricing model. Math. Finance 5, 13–32.

Dumas, B., Fleming, J., Whaley, R.E., 1998. Implied volatility functions: empirical tests. J. Finance, forthcoming.

Eichenbaum, M.S., Hansen, L.P., Singleton, K.J., 1988. A time series analysis of representative agent models of consumption and leisure choice under uncertainty. Q. J. Econ. 103, 51–78.

Engle, R.F., Bollerslev, T., 1986. Modeling the persistence of conditional variances. Econ. Rev. 5, 1–50.

Fama, E.F., 1984. Forward and spot exchange rates. J. Monetary Econ. 14, 319–338.

Feinstein, S., 1989. The Black–Scholes formula is nearly linear in σ for at-the-money options; therefore implied volatilities from at-the-money options are virtually unbiased. Working paper. Federal Reserve Bank of Atlanta.

Feinstein, S., 1989b. Forecasting stock market volatility using options on index futures. Econ. Rev. (Federal Reserve Bank Atlanta) 97, 115–154.

Fleming, J., 1993. The valuation and information content of S&P 100 index options. PhD dissertation. Duke University, Durham, NC.

Fleming, J., Whaley, R.E., 1994. The value of wildcard options. J. Finance 49, 215–236.
Fleming, J., Ostdiek, B., Whaley, R.E., 1995. Predicting stock market volatility: a new measure. J. Futures Markets 15, 265–302.
Gallant, A.R., Rossi, P.E., Tauchen, G., 1992. Stock prices and volume. Rev. Financial Studies 5, 199–242.
Granger, C.W.J., Newbold, P., 1974. Spurious regressions in econometrics. J. Econ. 2, 111–120.
Hamilton, J.D., 1994. Time Series Analysis. Princeton Univ. Press, Princeton.
Hansen, L.P., 1982. Large sample properties of generalized method of moment estimators. Econometrica 50, 1029–1054.
Hansen, L.P., Hodrick, R.J., 1980. Forward exchange rates as optimal predictors of future spot rates: an econometric analysis. J. Polit. Econ. 88, 829–853.
Hansen, L.P., Singleton, K.J., 1982. Generalized instrumental variables estimators of nonlinear rational expectations models. Econometrica 50, 1269–1286.
Harvey, C.R., Whaley, R.E., 1992a. Market volatility prediction and the efficiency of the S&P 100 index option market. J. Financial Econ. 31, 43–73.
Harvey, C.R., Whaley, R.E., 1992b. Dividends and S&P 100 index option valuation. J. Futures Markets 12, 123–137.
Hodrick, R.J., Srivastava, S., 1987. Foreign currency futures. J. Int. Econ. 22, 1–24.
Hull, J., White, A., 1987. The pricing of options on assets with stochastic volatilities. J. Finance 42, 281–300.
Jagannathan, R., 1985. An investigation of commodity futures prices using the consumption-based intertemporal capital asset pricing model. J. Finance 40, 175–191.
Jorion, P., 1995. Predicting volatility in the foreign exchange market. J. Finance 50, 507–528.
Lamoureux, C.G., Lastrapes, W.D., 1993. Forecasting stock return variance: toward an understanding of stochastic implied volatilities. Rev. Financial Studies 6, 293–326.
Latané, H.A., Rendleman, R.J., 1976. Standard deviations of stock price ratios implied in option prices. J. Finance 31, 369–381.
Merton, R.C., 1980. On estimating the expected return on the market: an exploratory investigation. J. Financial Econ. 8, 323–361.
Nelson, D.B., 1991. Conditional heteroskedasticity in asset returns: a new approach. Econometrica 59, 347–370.
Newey, W.K., 1985. Generalized method of moments specification testing. J. Econ. 29, 229–256.
Pagan, A.R., Schwert, G.W., 1990. Alternative models for conditional stock volatility. J. Econ. 45, 267–290.
Parkinson, M., 1980. The extreme value method for estimating the variance of the rate of return. J. Business 53, 61–66.
Phillips, P.C.B., 1986. Understanding spurious regressions in econometrics. J. Econ. 33, 311–340.
Poterba, J.M., Summers, L.H., 1986. The persistence of volatility and stock market fluctuations. Am. Econ. Rev. 76, 1142–1151.
Rubinstein, M., 1994. Implied binomial trees. J. Finance 49, 771–818.
Schmalensee, R., Trippi, R.R., 1978. Common stock volatility expectations implied by option premia. J. Finance 33, 129–147.
Schwert, G.W., 1990. Stock volatility and the crash of '87. Rev. Financial Studies 3, 77–102.
Shanken, J., 1990. Intertemporal asset pricing: an empirical investigation. J. Econ. 45, 99–120.
Stein, J., 1989. Overreactions in the options market. J. Finance 44, 1011–1023.
Stein, E.M., Stein, J.C., 1991. Stock price distributions with stochastic volatility: an analytic approach. Rev. Financial Studies 4, 727–752.
Whaley, R.E., 1993. Derivatives on market volatility: hedging tools long overdue. J. Derivatives 1, 71–84.
Wiggins, J., 1992. Estimating the volatility of S&P 500 futures prices using the extreme-value method. J. Futures Markets 12, 265–273.