Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Journal of Banking & Finance 27 (2003) 2151–2175
Journal of BANKING & FINANCE
www.elsevier.com/locate/econbase

The predictive power of implied volatility: Evidence from 35 futures markets

Andrew Szakmary [a,1], Evren Ors [b,2], Jin Kyoung Kim [c,3], Wallace N. Davidson III [b,*]

[a] *Finance Department, Robins School of Business, University of Richmond, Richmond, VA 23173, USA*
[b] *Finance Department, Southern Illinois University, Carbondale, IL 62901, USA*
[c] *Hyup Sung Industrial Co., Ltd., 447-5, Chobu-Ri, MoHyun-Myon YongIn-Si, KyongGi-Do, South Korea*

Received 23 November 2000; accepted 19 April 2002

Abstract

Using data from 35 futures options markets from eight separate exchanges, we test how well the implied volatilities (IVs) embedded in option prices predict subsequently realized volatility (RV) in the underlying futures. We find that for this broad array of futures options, IV performs well in a relative sense. For a large majority of the commodities studied, the implieds outperform historical volatility (HV) as a predictor of the subsequently RV in the underlying futures prices over the remaining life of the option. Indeed, in most markets examined, regardless of whether it is modeled as a simple moving average or in a GARCH framework, HV contains no economically significant predictive information beyond what is already incorporated in IV. These findings add to previous research that has focused on currency and crude oil futures by extending the analysis into a very broad array of contracts and exchanges. Our results are consistent with the hypothesis that futures options markets in general, with their minimal trading frictions, are efficient.

JEL classification: G13; G14
Keywords: Volatility forecasts; Futures options; Option market efficiency; GARCH models; Implied volatility

* Corresponding author. Tel.: +1-618-453-1429; fax: +1-618-453-7961.
E-mail address: davidson@cba.siu.edu (W.N. Davidson III).
[1] Tel.: +1-804-259-8251.
[2] Tel.: +1-618-453-2459.
[3] Tel.: +1-82-335-33-4221.

0378-4266/$ - see front matter
doi:10.1016/S0378-4266(02)00323-0

1. Introduction

How well does implied volatility (IV) predict future realized volatility (RV)? Research so far has failed to provide a definitive answer as results on this empirical question have been mixed. The purpose of this study is to re-examine the predictive power of IVs, relative to historical and GARCH-based volatility estimates. We accomplish this using a battery of tests with extensive data on 35 futures options covering a wide variety of asset classes and exchanges.

Solving an option pricing model backwards using an observed option price provides an estimate of the IV of returns on the underlying asset. The expected value of future RV should be equal to IV if (1) options are priced correctly and (2) the option pricing formula is correct. We find some support for this joint-hypothesis: Implieds are biased yet better predictors of RV than alternative historical volatility (HV) estimates.

Early research on the predictive content of IV found that IV explains variation in future volatilities better than HV. Latané and Rendleman (1976), Chiras and Manaster (1978), Schmalensee and Trippi (1978), and Beckers (1981) utilized stock options and the basic Black and Scholes (1973) option pricing model without considering dividends, the possibility of early exercise, different closing times in stock and option markets, and the various options' terms to maturity, to arrive at this conclusion.

In subsequent research, Kumar and Shastri (1990), Randolph et al. (1990), Day and Lewis (1992), Lamoureux and Lastrapes (1993), and Canina and Figlewski (1993), all of which examine either options on individual stocks or options on the cash S&P 100 index, use more sophisticated methodologies with more careful treatment of the data. They generally find that IV is a poor forecast of the subsequently RV over the remaining life of the option. For example, Canina and Figlewski (1993) use a regression approach to examine the predictive content of S&P 100 index options and find virtually *no* relation between the implied and subsequently RV over the remaining life of the option, despite the fact that the HVs and future RVs are related. Some of these studies also find that IV has little power to predict short-run changes in the volatility of the underlying asset, e.g. over a one-week horizon, compared to predictions that could be derived from time-series models. Specifically, Day and Lewis (1992) and Lamoureux and Lastrapes (1993) analyze the predictive power of IV within the context of GARCH-type models. They find that IV has some predictive power, but that GARCH and/or HV improve this predictive power.

In contrast to this later stream of research, Christensen and Prabhala (1988), using a monthly sampling frequency with non-overlapping data, find that IV is a good predictor for RV. Their results also demonstrate that the predictive ability of IV improved after the structural pricing shift in the OEX market that followed the 1987 stock market crash. Given the equivocal results and conclusions in stock market options alone, it is clear that further research on the predictive power of IV is warranted.

Only two studies of which we are aware have examined the performance of IV as a predictor outside of equity markets (Day and Lewis, 1993; Jorion, 1995). Both studies find that the IVs embedded in crude oil and foreign currency futures options, respectively, *do* provide a reasonably good forecast of the ultimately RVs in the underlying futures; for example, both studies find that the information content of the implieds for predicting volatility in the underlying asset over short time horizons is superior to the information content of GARCH models conditioned on HV. Jorion (1995) also demonstrates, using the Canina and Figlewski (1993) regression approach, that while IVs are not completely unbiased estimates of RVs, in currency futures the implieds outperform HV as a predictor of RV over the remaining life of the option. [4]

Our study is motivated by two major shortcomings in the previous literature. First, as noted above, most previous work has focused on options on individual stocks or on cash market stock indices. These options do not trade on the same exchange as their underlying asset, and the exchanges do not have the same daily closing time. The closing prices used by these studies would not be synchronous. Worse, as Fleming et al. (1996) show, trading costs in cash stock markets are relatively high, rendering arbitrage transactions involving both options and their underlying stocks more difficult. Thus, one major contribution of our study is that we use data from futures markets, where the options on futures and the underlying futures contracts trade on the same floor. These markets have trading costs that are orders of magnitude lower than those involving cash market transactions. Lower trading costs increase the feasibility of arbitrage, making the assumption of unlimited arbitrage inherent to option pricing models more tenable. Furthermore, while the stock market closes fifteen minutes prior to the closing of the options market (Harvey and Whaley, 1992) the futures and options markets we analyze close simultaneously. Studies based on stock options may not obtain consistent estimates of IV because the option prices, underlying security price, and information about the underlying asset are not observed simultaneously. This would introduce measurement error as simultaneous option and underlying asset prices are required to estimate IV. [5]

Another shortcoming of the previous literature, which again arises from the predominant focus on equity markets, is incompleteness in terms of the types of assets examined. One possible reason why Day and Lewis (1993) and Jorion (1995) obtain very different results than most studies that examine equity markets is that that they use futures options, in which trading frictions and measurement errors are much lower. However, an equally plausible explanation is that the time series properties of various asset classes differ, and IV may be a better predictor for some assets

[4] In addition to examining different markets, another key difference among the studies is that Jorion (1995) and Day and Lewis (1993) focus on near-the-money options, while Canina and Figlewski (1993) include those that are deep in or out-of-the-money as well.

[5] Some studies have controlled for the non-synchronous closing. For example, Lamoureux and Lastrapes (1993) used intraday data to match closing prices in the options and stock markets.

and options (e.g., foreign exchange, crude oil) than for others (equity markets). [6] Thus, the most important contribution of our study is that we examine futures options for a wide variety of asset classes, specifically, 35 different markets traded on eight different exchanges. In addition to futures options on an equity index (the S&P 500), currencies and crude oil, we also examine those on short-term and long-term interest rates, agricultural commodities, livestock, metals, refined petroleum products, and natural gas. By examining the predictive content of IV over such a wide range of futures options and markets, and by seeing if the predictive content varies across different categories of asset classes trading on different exchanges, we hope to determine whether Jorion's (1995) conjecture that IV is likely to be a good predictor for futures options *in general* is correct, or alternatively, if Jorion's results are specific to currency futures options only.

The balance of the paper is organized as follows: We describe our data and discuss related issues in Section 2. We present models, develop hypotheses, and present results for regression tests of implied and HV as predictors of RV in Section 3. Section 4 presents tests of the predictive power of IV vis-à-vis forecasts provided by a frequently-used GARCH model. Section 5 concludes the paper.

2. Data

In this study we compute and test IV data from eight separate exchanges and 35 futures options. Table 1 contains a list of the markets analyzed in this study. These options include equity-index, financial, energy, industrial, and agricultural futures options.

We obtain our data from Bridge. The Bridge data contain daily time series of the IV of futures options as well as data on the underlying futures such as open, high, low and settlement prices. Bridge calculates IV using the Black (1976) model with daily settlements for futures options and for futures, using the risk free rate obtained from the treasury bill futures contract. The data provider calculates IV on the two calls and the two puts with strike prices nearest to the underlying futures prices and reports an unweighted average of these four IVs. In options on futures contracts, unlike in options on cash instruments, transactions and financing costs are relatively small. Consequently, the IVs of call and put options are almost identical.

We do make one adjustment to the Bridge IVs: Bridge calculates the implieds based on the number of calendar days until maturity. However, several studies, e.g. French (1984), Fleming et al. (1995), Davidson et al. (2001) have shown that

[6] In addition, the sample period used by Day and Lewis (1993), which included the Iraqi invasion of Kuwait and the subsequent Gulf War period, may not be representative of normal conditions in the crude oil or other markets. A priori, we would expect that when sudden, unexpected political and military developments occur with unusual frequency, a forward-looking estimator such as IV would do especially well compared with alternate techniques that extrapolate historical data. It is difficult to draw general inferences regarding the efficiency of futures options markets from the results for this particular commodity during this period.

Table 1
List of futures options analyzed

Classification	Options contract	Sample beginning and ending dates	Active months of futures contract[a]	Exchange[b]
Equity index	S&P 500 index	01/28/1983–02/05/2001	H,M,U,Z	CME
Interest rate	Treasury bonds	09/01/1988–05/07/1999	H,M,U,Z	CBOT
	Treasury notes 10 year	09/01/1988–05/07/1999	H,M,U,Z	CBOT
	Treasury notes 5 year	05/24/1990–04/15/1999	H,M,U,Z	CBOT
	Eurodollar (ED)	03/20/1985–02/05/2001	H,M,U,Z	CME
	Treasury bill	01/03/1989–07/24/1996	H,M,U,Z	CME
	Long UK gilt	08/23/1990–01/28/1998	H,M,U,Z	LIFFE
	German Bond	08/23/1990–06/16/1999	H,M,U,Z	LIFFE
	Short Sterling	08/15/1990–04/03/1998	H,M,U,Z	LIFFE
	Notional Bond, Euro, 10 year	03/11/1991–11/02/1998	H,M,U,Z	MATIF
Currency	Japanese Yen	03/05/1986–02/05/2001	H,M,U,Z	CME
	Deutsche Mark	01/24/1984–02/05/2001	H,M,U,Z	CME
	Canadian Dollar	06/18/1986–02/05/2001	H,M,U,Z	CME
	British Pound	02/25/1985–02/05/2001	H,M,U,Z	CME
	Swiss Franc	02/25/1985–02/05/2001	H,M,U,Z	CME
Energy	Crude oil	01/11/1989–02/05/2001	F–Z	NYM
	Heating oil no. 2	02/01/1989–02/05/2001	F–Z	NYM
	Gasoline unleaded	04/17/1989–02/05/2001	F–Z	NYM
	Natural gas	10/02/1992–02/05/2001	F–Z	NYM
Metals	Copper	04/26/1990–02/05/2001	F,H,K,N,U,Z	COMEX
	Gold	01/03/1989–02/05/2001	G,J,M,Q,V,Z	COMEX
	Silver	03/06/1990–12/31/1996	H,K,N,U,Z	COMEX
Agriculture	Corn	02/24/1989–02/05/2001	H,K,N,U,Z	CBOT
	Soybeans	02/24/1989–02/05/2001	F,H,K,N,Q,U,X	CBOT
	Soybean meal	02/24/1989–02/05/2001	F,H,K,N,Q,U,V,Z	CBOT
	Soybean oil	02/24/1989–02/05/2001	F,H,K,N,Q,U,V,Z	CBOT
	Wheat	02/24/1989–02/05/2001	H,K,N,U,Z	CBOT
	Coffee	03/05/1990–02/05/2001	H,K,N,U,Z	CSCE
	Sugar	03/06/1990–02/05/2001	H,K,N,V	CSCE
	Cocoa	03/06/1990–02/05/2001	H,K,N,U,Z	CSCE
	Cotton	01/30/1990–02/05/2001	H,K,N,V,Z	NYCE
	Orange juice	03/02/1990–01/15/1999	F,H,K,N,U,X	NYCE
Livestock	Feeder cattle	01/09/1987–02/05/2001	F,H,J,K,Q,U,V,X	CME
	Live cattle	10/30/1984–02/05/2001	G,J,M,Q,V,Z	CME
	Lean hogs	02/01/1985–02/05/2001	G,J,M,N,Q,V,Z	CME

[a] F = January, G = February, H = March, J = April, K = May, M = June, N = July, Q = August, U = September, V = October, X = November, and Z = December.

[b] Chicago Board of Trade (CBOT); Chicago Mercantile Exchange (CME); Coffee, Sugar, Coca Exchange (CSCE); Commodity Exchange (COMEX); London International Financial Futures Exchange (LIFFE); Marché à Terme International De France (MATIF); New York Mercantile (NYM); New York Cotton Exchange (NYCE).

in virtually all markets, volatility is much higher on trading days than on non-trading days. Thus, the implieds are more properly associated with trading days left until option maturity. Following Ederington and Lee (1996), we adjust the Bridge implieds as follows:

$$\mathrm{IV}_t = \mathrm{IV}_t^{\mathrm{Bridge}} \times \sqrt{\frac{T_{\mathrm{C}}}{T_{\mathrm{M}}}}$$

where IV_t and $\mathrm{IV}_t^{\mathrm{Bridge}}$ are, respectively, the adjusted and the Bridge IVs on day t, and T_{C} and T_{M} are, respectively, the number of calendar days and trading days left until the date of expiration of the option.

From the Bridge data as adjusted above, we create a single, daily IV series for each contract. In this process, we use the option contracts with the closest expiration date prior to the maturity date of the underlying futures contract. For example, the S&P 500 options are available with expirations for every month, but the S&P 500 futures are available only with March, June, September, and December expirations. Thus, for the S&P 500 options, we use only the options contracts with the closest expiration date prior to the expiration of the March, June, September, and December futures contracts. This approach ensures the best match between the IVs (which are computed from current option and underlying futures prices) and the RVs calculated from subsequent futures prices.

We use IV from the options until ten trading days prior to option expiration and then switch to the next deferred option series. This procedure allows us to: (1) avoid distortion effects arising from estimating volatility with too short of an option maturity, (2) reduce distortion effects associated with the use of Black's model for too long of an option maturity, and (3) provide uniformity in volatility estimation across the contracts we examine. [7] To eliminate any effect at the time of rollover, we compute daily futures returns using only price data from the identical contract. Therefore, on the day of rollover, we gather futures prices for both the nearby and first-deferred contracts, so that the daily return on the day after rollover is measured with the same contract month.

Another issue in using IV is the moneyness of the options. It is well known that for options of equivalent maturity, IV tends to be higher for deep in the money or deep out of the money options than for those that are near the money (see, for example, Rubinstein, 1985; Ignacio et al., 1999). If actual return distributions have fat tails compared with the log-normal distribution assumed in option pricing models, then this "smile" effect in IVs is more likely to be a result of model inadequacy than irrational behavior by investors (Jackwerth and Rubinstein, 1996). The IVs computed from at the money options are least affected by the non-normal distribution of returns; Beckers (1981) and Canina and Figlewski (1993) find that the IVs of near the money options are better predictors of future RV than the IVs of deep in or out

[7] We repeated most of the tests examined in this study using rollover periods of 5 and 15 days prior to option expiration, and the results were not materially different.

of the money options. One advantage of the data used in this study is that Bridge uses only nearest to the money calls and puts to calculate IVs on each day.

3. Regression tests of implied volatility as a predictor of realized volatility

3.1. Research hypotheses

IV has been regarded as an unbiased expectation of the RV under the assumption that the market is informationally efficient and the option pricing model is specified correctly. If, as previous research indicates, the pricing model is accurate for the short-term, near the money options used in this study, then rejection of IV as an unbiased and efficient predictor would imply that the options market does not efficiently establish prices.

Consistent with the existing literature, to test whether the IV has a significant amount of information over the HV, we examine the following three hypotheses:

H1. IV is an unbiased estimate of the future RV.

H2. IV has more explanatory power than the HV in forecasting RV.

H3. IV efficiently incorporates all information regarding future volatility; HV contains no information beyond what is already included in IV.

The statistical procedures we use to test these hypotheses are described in the next section.

3.2. Statistical procedures

3.2.1. Stationarity tests

In studying time series data, examining stationarity is the most important prerequisite to further analysis. A non-stationary time series has an integrated process in which a given shock to the series does not die out. Sims et al. (1990) demonstrate that an econometric model with an integrated time series will generally be misspecified.

To test stationarity, we employ the augmented Dickey–Fuller (ADF) method (Dickey and Fuller, 1979). We perform the ADF test by running ordinary linear regressions of the form

$$\Delta X_t = \phi_0 + \phi_1 X_{t-1} + \sum_{i=1}^{N} \varphi_i \Delta X_{t-i} + e_t, \tag{1}$$

where X_t is the series being tested for unit roots; $\Delta X_t = X_t - X_{t-1}$; ΔX_{t-i} is ΔX_t with the ith lag; and N is the lag length decided by Schwert's (1989) procedure. The null hypothesis of Eq. (1) is that X_t contains a unit root. The test statistic is the t-statistic for ϕ_1; critical values appear in Fuller (1976). When the test statistic is significant the null hypothesis is rejected. However, an insignificant test statistic does not necessarily imply a unit root; in some cases it may simply mean there is insufficient evidence to reject the null. [8]

3.2.2. Tests of implied volatility as a predictor of realized volatility

Canina and Figlewski (1993) developed simple tests to measure market efficiency. IV is regarded as a forecast of the market's RV:

$$\mathrm{IV} = E_{\mathrm{MKT}}[\mathrm{RV}], \tag{2}$$

where E_{MKT} is the market's expectation of the RV.

Following Canina and Figlewski, we calculate the actual RV as the annualized standard deviation of the continuously compounded return:

$$\mathrm{RV} = \left(\frac{260}{T_{\mathrm{M}} - 1} \sum_{t=1}^{T_{\mathrm{M}}} (R_t - \bar{R})^2 \right)^{1/2}, \tag{3}$$

where T_{M} is the number of trading days until the option expiration; $R_t = \ln(P_t/P_{t-1})$; and $\overline{R}$ is the sample mean of R_t. If the market is efficient then the estimation errors should be a zero-mean random variable. We employ the procedure in Canina and Figlewski (1993) by estimating the following regressions:

$$\mathrm{RV}_t = \alpha + \beta \mathrm{IV}_t + e_t, \tag{4}$$

$$\mathrm{RV}_t = \alpha' + \beta' \mathrm{HV}_t + e_t, \tag{5}$$

$$\mathrm{RV}_t = \alpha + \beta \mathrm{IV}_t + \beta' \mathrm{HV}_t + e_t. \tag{6}$$

If, as our hypothesis H1 earlier stated, IV is an unbiased predictor of the RV, we should expect $\alpha = 0$ and $\beta = 1$ in regression (4). If, in accordance with hypothesis H2, IV includes more information (i.e., current market information) than HV, then IV should have greater explanatory power than HV, and we would expect a higher R^2 from regression (4) than regression (5). Finally, if hypothesis H3 is correct, then when IV and HV appear in the same regression, as in (6), we would expect $\beta' = 0$ since HV should have no explanatory power beyond that already contained in IV. [9]

---

[8] Barnhart and Szakmary (1991) found that, in testing market efficiency, results are sensitive to correct model specification. They used an error correction model for the non-stationary and cointegrated series. The stationarity issue is not mentioned in most of the recent related studies. Therefore some caution might be appropriate in interpreting their results. Jorion (1995) mentioned this point but did not report the unit root test results for the series in his study.

[9] Regressions (4)–(6) are appropriately specified only if the variables are stationary. If a series is not stationary, the appropriate procedure is to estimate error-correction models to draw inferences concerning the coefficients.

In the above regressions, we use a 30 day window to calculate HV. [10] Since we use daily data but implied and RVs are measured over horizons that vary from 10 to approximately 70 trading days (depending on the specific contract) the regressions are estimated with overlapping observations. Following Canina and Figlewski (1993) and Jorion (1995), we use the Hansen (1982) correction to adjust the standard errors of the regression coefficients to properly reflect serial correlation of varying lengths in the residuals of regressions (4)–(6).

3.3. Empirical results

3.3.1. General results

The ADF unit root test results for each series are reported in Table 2. From these results we reject the null hypothesis of a unit root, and find strong evidence that the IV is stationary for most of the series under examination, with the possible exception of options on sugar futures. Indeed, the ADF test statistic is significant at the 1% level for 28 of the 35 markets. The stationarity results for RV are similar: All of the ADF test statistics are significant at least at 5%, with 32 of the 35 significant at 1%. We find slightly more potential non-stationarity among the HV estimates, although even here 28 of 35 markets exhibit significant test statistics. All told, we believe our stationarity results, especially for the IVs and RVs, support the conjecture that these series are stationary, because even for those few series for which stationarity is not clearly indicated, the test statistics are negative and are close to being significant. Thus, failure to reject the null hypothesis of a unit root for IV for sugar futures options is likely due to the low power of the ADF test rather than genuine non-stationarity, and in the subsequent empirical tests we treat all of the series as stationary.

3.3.2. Results for forecasting realized volatility

The test results for the relative predictive power of IV and HV in forecasting the RV appear in Table 3. We report the regression results for Eqs. (4)–(6) after correcting the estimated standard errors of the coefficients for heteroskedasticity and serial correlation in the residuals. [11] For each of the three regressions and 35 commodities, we report the coefficients, associated *t*-statistics, and adjusted-R^2.

[10] We also estimated regression (5) with HV measured over 60 and 90 day windows, and chose the 30 day window because it has the largest adjusted-R^2, on average, when estimating regression (5). Thus, the 30 day window provides the most competitive alternative to IV in forecasting RV. Furthermore, HV measured over 30 days exhibits greater stationarity, overall, than HV with a 60 or 90 day window, which should improve the consistency of the estimates and the inferences that we make.

[11] In general, we follow Canina and Figlewski (1993), but as we mentioned in the data description, our data sets have different characteristics. Instead of using all the IV available and running simple cross-sectional regressions, we construct our data in a single time series by selecting only those futures options with the closest expiration date prior to the maturity date of the underlying futures contracts. When the nearby futures option has 10 trading days before expiration, we roll over to the first-deferred futures option contract. This approach provides continuous time series data without frequently switching to the next nearby options.

Table 2
ADF unit root tests

Contract	RV	IV	HV30	GFOR
S&P 500 Index	−5.48725**	−4.99233**	−4.95925**	−12.12282**
Treasury bond	−5.22145**	−5.19074**	−4.03584**	−5.53483**
Treasury notes 10 year	−4.92470**	−3.40018*	−3.91628**	−5.61982**
Treasury notes 5 year	−4.60333**	−4.46622**	−3.75975**	−4.72677**
Eurodollar	−4.52345**	−3.69233**	−3.29640*	−7.17840**
Treasury Bill	−4.63884**	−5.25645**	−4.23938**	−10.88068**
Long UK Gilt	−4.51139**	−5.31155**	−2.87404*	−4.67915**
German Bond	−3.85044**	−3.28857*	−2.98409*	−5.50639**
Short Sterling	−3.09317*	−3.41396*	−2.53331	−7.99496**
Notional Bond, Euro, 10 year	−3.01774*	−3.00877*	−2.07926	−3.14602*
Japanese Yen	−5.91587**	−5.23244**	−3.73168**	−8.53299**
Deutsche Mark	−5.71105**	−5.67473**	−4.61561**	−8.21738**
Canadian Dollar	−5.19874**	−5.32468**	−3.89250**	−7.27108**
British Pound	−4.98601**	−4.80138**	−3.72225**	−4.70768**
Swiss Franc	−6.43938**	−6.17871**	−5.03810**	−7.30693**
Crude oil	−4.89471**	−5.46472**	−3.55631**	−13.04258**
Heating oil no. 2	−4.22308**	−3.66816**	−3.05282*	−5.47333**
Gasoline unleaded	−5.67456**	−4.80158**	−3.44403**	−7.02647**
Natural gas	−5.29682**	−4.57741**	−3.55647**	−5.60948**
Copper	−4.74780**	−5.24482**	−3.59325**	−3.75451**
Gold	−4.85776**	−3.08414*	−2.55930	−5.15613**
Silver	−3.53383**	−3.74836**	−2.36738	−3.55612**
Corn	−4.48224**	−5.66623**	−4.27984**	−6.24048**
Soybeans	−6.18183**	−6.38133**	−5.82267**	−7.84573**
Soybean meal	−5.77966**	−6.05843**	−3.35173*	−6.45405**
Soybean oil	−6.48294**	−7.19254**	−5.87992**	−10.15955**
Wheat	−4.96576**	−5.45061**	−3.98041**	−6.30320**
Coffee	−4.28892**	−4.48163**	−4.00599**	−6.50264**
Sugar	−3.38079*	−2.51678	−2.77039	−2.87248*
Cocoa	−4.39154**	−3.16364*	−3.12594*	−4.52058**
Cotton	−4.77727**	−4.47188**	−3.29148*	−4.09197**
Orange juice	−4.80347**	−4.23748**	−3.79602**	−6.31088**
Feeder cattle	−4.43207**	−4.99828**	−4.28982**	−5.28568**
Live cattle	−4.78898**	−4.61840**	−2.80336	−4.25771**
Lean hogs	−4.89855**	−4.79941**	−4.11589**	−4.37335**

For each series we estimate the following regression:

$$\Delta X_t = \phi_0 + \phi_1 X_{t-1} + \sum_{i=1}^{N} \varphi_i \Delta X_{t-i} + e_t$$

where X_t is the series being tested for unit roots, $X_{t-i} = X_t$ with the ith lag, $X_{t-i} = X_t - X_{t-1}$, and N is the lag length. For each series, the lag length is chosen to eliminate any significant autocorrelation in the residuals. The test statistic is the t-statistic on the coefficient ϕ_1; however, it does not have a standard $t-1$ distribution. The critical values are given in Fuller (1976).

* and ** denote significance at the 5% and 1% levels, respectively.

All slope coefficients for IV in Eq. (4) are positive and significant at the 1% level. These findings are consistent with H1. This hypothesis, however, is more restrictive, in that it predicts that the slope coefficients should be one, and the constant

Table 3
Forecasting RV with IV and HV

	(4) $\mathrm{RV}_t = \alpha + \beta \mathrm{IV}_t + e_t$			(5) $\mathrm{RV}_t = \alpha' + \beta' \mathrm{HV}_t + e_t$			(6) $\mathrm{RV}_t = \alpha + \beta \mathrm{IV}_t + \beta' \mathrm{HV}_t + e_t$			
	α	β	Adjusted-R^2	α'	β'	Adjusted-R^2	α	β	β'	Adjusted-R^2
S&P 500 Index	2.902**	0.614**	0.231	10.558**	0.306**	0.132	2.596*	0.657**	−0.036	0.231
	2.633	10.241		6.907	5.072		2.059	6.129	−0.794	
Treasury bond	3.550*	0.470**	0.125	6.086**	0.320**	0.108	3.720*	0.322*	0.172	0.144
	2.273	3.468		6.001	2.934		2.550	2.217	1.351	
Treasury notes 10 year	2.130*	0.472**	0.159	3.661**	0.378**	0.144	2.282*	0.308*	0.196	0.178
	2.128	3.781		5.197	3.388		2.385	2.508	1.773	
Treasury notes 5 year	1.424**	0.459**	0.211	2.262**	0.408**	0.183	1.452**	0.314*	0.189	0.229
	2.695	4.437		5.277	3.939		2.861	2.293	1.345	
Eurodollar	4.542**	0.584**	0.399	7.408**	0.505**	0.278	4.488**	0.526**	0.077	0.401
	4.866	10.598		6.128	6.776		4.771	5.699	0.797	
Treasury bill	10.291**	0.371**	0.106	14.867**	0.214**	0.046	10.289**	0.392**	−0.026	0.106
	6.075	5.608		8.098	3.508		6.142	2.881	−0.200	
Long UK Gilt	2.968**	0.460**	0.285	4.768**	0.348**	0.189	2.954**	0.394*	0.085	0.290
	2.644	4.009		4.994	2.767		2.879	2.235	0.518	
German Bond	0.756	0.656**	0.456	2.378**	0.483**	0.314	0.655	0.749**	−0.095	0.458
	1.872	10.600		4.982	5.253		1.683	6.177	−0.773	
Short Sterling	4.110**	0.624**	0.345	9.513**	0.304**	0.191	4.187**	0.548**	0.079*	0.353
	3.096	6.983		6.836	5.960		3.365	6.261	2.126	
Notional Bond, Euro 10 year	0.770	0.575**	0.476	1.642**	0.609**	0.458	0.919**	0.346**	0.284	0.500
	1.846	7.798		5.406	7.771		2.640	2.704	1.620	
Japanese Yen	4.010**	0.548**	0.248	7.141**	0.364**	0.163	4.043**	0.481**	0.075	0.250
	3.280	5.793		9.696	5.925		3.443	3.577	0.854	
Deutsche Mark	4.000**	0.520**	0.254	6.989**	0.374**	0.172	4.053**	0.457**	0.073	0.256
	3.914	7.266		7.832	5.018		4.272	4.792	0.697	
Canadian Dollar	1.646**	0.488**	0.285	3.017**	0.375**	0.150	1.649**	0.498**	−0.014	0.285
	3.205	6.434		7.516	5.411		3.219	4.330	−0.144	

Table 3 (*continued*)

	(4) $\mathrm{RV}_t = \alpha + \beta \mathrm{IV}_t + e_t$			(5) $\mathrm{RV}_t = \alpha' + \beta' \mathrm{HV}_t + e_t$			(6) $\mathrm{RV}_t = \alpha + \beta \mathrm{IV}_t + \beta' \mathrm{HV}_t + e_t$			
	α	β	Adjusted-R^2	α'	β'	Adjusted-R^2	α	β	β'	Adjusted-R^2
British Pound	2.572*	0.609**	0.383	4.983**	0.498**	0.320	2.710**	0.454**	0.173	0.396
	2.426	7.136		6.775	7.411		2.874	3.803	1.648	
Swiss Franc	5.436**	0.474**	0.182	7.805**	0.360**	0.154	5.484**	0.333**	0.161	0.196
	3.771	4.649		7.156	4.034		4.141	3.050	1.511	
Crude oil	−1.109	0.759**	0.724	11.854**	0.548**	0.357	−0.976	0.842**	−0.112	0.730
	−0.425	10.696		3.505	4.303		−0.427	6.354	−1.168	
Heating oil no. 2	4.201	0.657**	0.574	15.924**	0.421**	0.195	5.097	0.744**	−0.140	0.585
	1.152	6.119		5.033	3.565		1.657	4.251	−1.207	
Gasoline unleaded	3.288	0.668**	0.610	14.037**	0.475**	0.238	4.237*	0.780**	−0.175*	0.625
	1.301	9.259		3.774	3.433		2.022	6.602	−2.034	
Natural gas	5.765	0.618**	0.430	22.567**	0.437**	0.201	5.971	0.694**	−0.105	0.435
	1.459	8.635		4.852	4.159		1.505	6.309	−0.901	
Copper	9.609**	0.403**	0.177	12.602**	0.395**	0.157	9.284**	0.269**	0.192	0.195
	3.890	5.107		5.506	3.705		3.654	2.774	1.551	
Gold	1.897	0.596**	0.368	6.121**	0.456**	0.253	1.971	0.569**	0.031	0.368
	1.897	10.866		7.473	7.615		1.916	6.453	0.447	
Silver	4.521	0.633**	0.365	11.417**	0.510**	0.257	4.558	0.538**	0.122	0.371
	1.562	6.486		5.383	5.298		1.607	4.938	1.300	
Corn	1.088	0.662**	0.513	9.208**	0.476**	0.245	1.031	0.643**	0.029	0.513
	0.622	8.781		5.712	4.776		0.591	7.264	0.423	
Soybeans	2.652	0.637**	0.421	9.929**	0.436**	0.192	2.485	0.607**	0.048	0.422
	1.560	8.438		5.798	4.497		1.434	7.377	0.759	
Soybean meal	4.545*	0.618**	0.383	9.611**	0.498**	0.239	4.313*	0.555**	0.089	0.386
	2.526	8.180		6.421	6.426		2.464	5.632	1.207	
Soybean oil	7.954**	0.446**	0.266	12.912**	0.315**	0.091	8.145**	0.460**	−0.028	0.266
	4.580	6.294		9.173	4.280		5.057	4.948	−0.370	
Wheat	4.630*	0.606**	0.366	10.928**	0.466**	0.211	4.619*	0.598**	0.011	0.366
	2.370	8.136		6.110	5.523		2.374	5.594	0.114	

Coffee	5.214	0.698**	0.340	22.049**	0.437**	0.185	5.243	0.689**	0.010	0.340
	1.218	7.899		5.786	4.648		1.225	6.032	0.117	
Sugar	14.748**	0.351**	0.308	9.626**	0.637**	0.458	8.413**	0.127	0.515**	0.481
	3.378	3.008		4.701	9.368		3.935	1.740	5.748	
Cocoa	5.791*	0.582**	0.374	12.492**	0.510**	0.299	5.745*	0.435**	0.193*	0.393
	2.140	7.388		6.306	8.069		2.331	4.287	2.501	
Cotton	5.565*	0.580**	0.388	9.796**	0.507**	0.232	5.009*	0.512**	0.112	0.394
	2.540	7.147		4.073	4.427		2.047	7.013	1.058	
Orange juice	13.822**	0.382**	0.367	17.440**	0.397**	0.164	12.421**	0.348**	0.092	0.373
	6.720	8.330		7.389	5.640		5.933	5.856	1.107	
Feeder cattle	2.795**	0.588**	0.380	4.793**	0.555**	0.335	2.690**	0.399**	0.250**	0.408
	3.323	9.174		8.061	10.145		3.224	4.831	3.467	
Live cattle	2.099*	0.638**	0.541	4.168**	0.667**	0.468	2.123*	0.520**	0.150	0.546
	2.375	12.544		4.875	10.691		2.429	7.395	1.951	
Lean hogs	4.410**	0.622**	0.495	0.922**	0.617**	0.418	3.967**	0.451**	0.234*	0.518
	3.423	12.527		4.980	7.652		2.870	5.709	2.302	

We test whether the IV forecasts the RV better than the HV with regressions (4)–(6) where RV_t is the annualized RV from day t to the maturity of the futures option, IV_t is the IV, and HV_t is the 30 day HV. We use the Hansen (1982) correction to adjust the coefficient standard errors to reflect varying lengths of overlap in the data.

* and ** represent 5% and 1% significance levels, respectively.

terms should be zero. The magnitude of the slope coefficients for IV ranges from 0.351 for sugar to 0.759 for crude oil; thus, all are less than 1. This evidence does not support H1 in its restrictive form. Furthermore, we find that 34 out of the 35 futures options show a positive constant term, and 23 of these are significantly positive at the 5% level or better. Again, this finding does not fully accord with H1. Thus, consistent with Jorion (1995) and other studies, our results demonstrate that IV is a biased forecast of the RV: The positive constant terms and slopes below unity imply that when IV is relatively high, the subsequent RV tends to be lower, and vice versa. [12]

To test the second hypothesis that IV is a better predictor of RV than is HV, we compare the adjusted-R^2 of regressions (4) and (5). As shown in Table 3, the predictive power of IV is superior to that of HV in 34 of the 35 futures options markets. Indeed, in 27 out of the 34 markets in which the IV regression exhibits a higher R^2, the R^2 from regression (4) is at least 5 percentage points (0.05) greater than the R^2 from regression (5). In seven markets, while the predictive power of IV remains greater, the margin is narrower. The only exception to the superior predictive power of IV is sugar, where the R^2 from regression (5) using HV as the independent variable is considerably greater than the R^2 from regression (4), which uses IV.

Due to the large number of separate markets we examine, we can formally assess the probability that the observed dominance of implied over HV is due to chance. If we assume that the 35 markets are independent of each other, then by any reasonable measure this probability is extremely small. The cumulative binomial probability function for X or fewer successes in N independent trials is given by

$$\mathrm{CP}(X) = \sum_{X=0}^{N} \frac{N!P^X(1-P)^{N-X}}{X!(N-X)!}, \tag{7}$$

where, P is the probability of success in each trial. Here, $N = 35$. If the null hypothesis is that IV and HV have equal predictive power, then the probability that IV predicts better than HV under the null is 0.5, and $P = 0.5$. If we define a "successful" outcome as one in which the IV regression has a higher R^2, then 34 markets meet this criterion, and we need to use Eq. (7) to compute the probability of observing 33 or fewer successes. Since CP(33) is equal to 0.999999999, this implies that the probability of 34 or more successes is less than 0.0001, even when we double the probability to allow for a two-tailed test. Using a more stringent standard, we can define a successful outcome as one in which the R^2 from regression (4) exceeds the R^2 from regression (5) by at least 0.05. However, as discussed earlier, even this higher hurdle is met by 27 of the 35 markets. Since CP(26) is 0.99906, the probability of

[12] As Jorion (1995) explains, even though futures options are less impacted by stale prices and high transactions costs in their underlying instruments, they are still affected by measurement error in the form of bid–ask spreads. These measurement errors have the effect of biasing the slope coefficients in regression (4) downwards. However, an extensive simulation analysis in Jorion's study (which used fewer observations than are used in our study for most commodities) showed that these biases are not large enough to account for the degree to which the estimated slope coefficients are below unity, unless one assumes implausibly large bid–ask spreads.

observing 27 or more successes, using a two-tailed test, is ≈ 0.0019. Thus, at any conventional significance level, the null hypothesis can be rejected in favor of the alternative: IV predicts RV better than HV across the wide number of contracts that we test.

When we examine the results from regression (6) in which IV and HV are both included in the regression, all but one of the coefficients for IV are significant (again, sugar is the only exception), while only 6 of the 35 coefficients for HV are significant. Of these six, one (for unleaded gasoline) is significantly *negative*. In those five cases in which both the coefficient for IV and HV are statistically significant for the same contract, the coefficients for IV are closer to one (nominally) and have greater levels of significance. [13]

Since the coefficient for HV in regression (6) is significant, at the 5% level or better, in only six of 35 markets we examine, it is interesting to use the binomial probability distribution presented earlier to explore whether these findings concerning HV could be due purely to chance. If our null hypothesis is that HV *never* adds information that is not already contained in IV, then the probability of nevertheless obtaining a significant HV coefficient in regression (6) is 0.05. Using $N = 35$, $X = 5$ and $P = 0.05$ in Eq. (7) reveals that the cumulative probability of obtaining 5 or fewer significant HV coefficients is 0.99275. This indicates that the probability of 6 or more significant coefficients (this time using a one-tailed test) is ≈ 0.007, and this is low enough to reject the null. Alternatively, we note that 26 out of 35 of the coefficients on HV in regression (6) are positive, and ask what is the chance of getting such a result if, in fact, the true probability of obtaining a positive β' coefficient in each regression is 0.50. The binomial model tells us that the probability of obtaining 26 or more positive coefficients is only 0.003. Thus, no matter how we look at our results, we are forced to conclude that H3 is rejected, and HV adds predictive information in at least some markets. We note, however, that the average improvement in adjusted-R^2 when we move from regression (4) to regression (6) is only 1.3%. Thus, while H3 is rejected statistically, the additional information in HV does not appear to be economically meaningful. [14]

[13] Multicollinearity in regression (6) could potentially increase the standard errors and affect our results. The coefficient of correlation (ρ) between IV and HV is positive and statistically significant at the 1% level for all contracts: ρ ranges from a low of 0.56 for the orange juice contract to a high of 0.87 for the German Bond contract. To determine whether insignificant β' coefficients are due to multicollinearity we did the following test. We regressed residuals from regression (4) on HV. If the coefficient of HV is significant, this would suggest that HV contains information that helps explain RV over and above (i) information that is common to both IV and HV and (ii) information that is unique to IV. We found that only in four out of 35 contracts HV added information beyond what was contained in IV. We, then, regressed residuals from regression (5) on IV. We found that for 30 out of 35 contracts IV added information beyond that contained (i) commonly in IV and HV, and (ii) uniquely in HV. We concluded that multicollinearity alone cannot explain our results.

[14] Measurement error in IV will also tend to bias HV's coefficient up slightly. If the true IV fully contains HV but the measured IV differs from the true IV, then both the measured IV and HV have information about the true IV (that is, they are proxies). Thus this explanation for why the IV coefficient is less than 1.0 in regression (6) may explain why HV or GFOR coefficients are usually positive. We thank the anonymous referee for this explanation.

Our findings for most futures options markets qualitatively agree with Jorion's (1995) results for currency futures options, but not with Canina and Figlewski (1993) and other studies which have examined cash stock index options, or options on individual stocks. While the likely explanation is that futures options are more efficient because trading frictions are lower, alternative possibilities exist. First, it is possible that the findings may be influenced by the sample period. Canina and Figlewski's sample period runs from March 1983 to March 1987. Thus, their sample entirely predates the October 1987 crash, whereas for the commodities in our study, all or at least most of the data are drawn from the post-crash period. As an indication of the extent, if any, of the influence of the sample period, we re-ran the regression tests in Table 3 for the pre-crash and post-crash periods for the S&P 500 futures options. We defined the pre-crash period as January 28, 1983–August 24, 1987 (the latter date is 15 trading days prior to the expiration of the September 1987 contract; the RVs for the subsequent December 1987 contract are heavily influenced by the crash). The post-crash period was defined as January 4, 1988–February 5, 2001. These results (not separately reported) showed that the slope coefficient in regression (4) was somewhat higher in the post-crash period. However, it was positive and highly significant in the pre-crash period as well. Moreover, the R^2 of regression (4) exceeded the R^2 of regression (5) by a wide margin both before and after the crash, and in both periods, in regression (6), β was significantly positive while β' was negative. Thus, the sample period does not appear to have much influence and cannot explain the difference in findings for the S&P 500 futures options in this study and the OEX options in Canina and Figlewski (1993).

Another factor which might influence our results is the inverse relationship between time to expiration and the IV reported in Becker and Tucker (1991), Canina and Figlewski (1993), and Park and Sears (1985). To examine whether a term-to-maturity effect may have influenced our results, we re-estimated regression (6) after introducing interactive-indicator variables D_1, D_2 and D_3, which equal 1 when the options have $\leqslant 30$ trading days, 31–49 trading days, and $\geqslant 50$ trading days to maturity, respectively, and 0 otherwise: [15]

$$\begin{aligned}\mathrm{RV}_t = {}& \alpha_1 D_1 + \alpha_2 D_2 + \alpha_3 D_3 + \beta_1(D_1\mathrm{IV}_t) + \beta_2(D_2\mathrm{IV}_t) + \beta_3(D_3\mathrm{IV}_t) \\ & + \beta_1'(D_1\mathrm{HV}_t) + \beta_2'(D_2\mathrm{HV}_t) + \beta_3'(D_3\mathrm{HV}_t) + e_t. \end{aligned} \tag{8}$$

The results, reported in Table 4, seem to show that the predictive power of IV is not reliably influenced by the remaining term to maturity of the option. For example, while the null hypothesis that the coefficient estimates are equal across various periods indicated by the above dummy variables (that is, $\beta_1 = \beta_2 = \beta_3$) is rejected in

[15] We could not run these regressions for crude oil, heating oil, unleaded gasoline, and natural gas because, given the monthly expiration cycle, these options always had 33 or fewer trading days to maturity. For certain commodities, there were very few observations with $\geqslant 50$ trading days, so β_3 and β_3' coefficients could not be reliably estimated; we ran these with only the first two terms and test whether coefficients across these are equal. We used the Hansen (1982) correction to estimate standard errors in the presence of overlapping observations.

Table 4
RV, IV, and HV by maturity characterization

	β_1, β_1'	β_2, β_2'	β_3, β_3'	χ^2 test $\beta_1=\beta_2=\beta_3$, $\beta_1'=\beta_2'=\beta_3'$
S&P Index 500	0.725**	0.533**	0.639**	2.336
	−0.043	−0.013	0.107	0.931
Treasury bond	0.131	0.515*	0.368	4.489
	0.341	0.167	0.075	1.530
Treasury notes 10 year	0.136	0.440*	0.314	2.247
	0.407	0.217	0.057	2.377
Treasury notes 5 year	0.155	0.447*	0.286	2.193
	0.375	0.196	0.083	1.323
Eurodollar	0.668**	0.537**	0.488**	1.051
	−0.061	0.062	0.159	2.464
Treasury bill	0.126	0.666	0.386*	2.040
	0.198	−0.161	−0.052	3.605
Long UK Gilt	0.259	0.479	0.645**	3.376
	0.057	0.088	0.026	0.129
German Bond	0.662**	0.604**	0.886**	1.934
	−0.097	0.044	−0.116	0.425
Short Sterling	0.514**	0.559**	0.561**	0.061
	0.089	0.068	0.089**	0.307
Notional Bond, Euro 10 year	0.624**	0.240	0.161	7.910*
	0.045	0.326	0.576**	4.496
Japanese Yen	0.755**	0.315*	0.417*	7.004*
	−0.095	0.182	0.138	4.933
Deutsche Mark	0.511**	0.336**	0.514**	4.515
	−0.011	0.156	0.083	1.611
Canadian Dollar	0.603**	0.389**	0.529**	2.308
	−0.185	0.046	0.080	1.499
British Pound	0.512**	0.261*	0.585**	9.750**
	0.143	0.256*	0.131	1.509
Swiss Franc	0.487**	0.222*	0.257	4.503
	0.105	0.188	0.227	0.462
Copper	0.205*	0.333*	0.535**	3.208
	0.348	0.043	0.025	3.106
Gold[a]	0.412**	0.800**	–	5.385*
	0.164	−0.122	–	3.408
Silver	0.714**	0.391**	0.429**	2.800
	−0.020	0.210	0.199*	1.239
Corn	0.526**	0.524**	0.661**	1.946
	0.234	0.135	−0.028	3.615
Soybeans[a]	0.677**	0.404**	–	8.354**
	0.079	0.076	–	0.001

(*continued on next page*)

Table 4 (*continued*)

	β_1, β_1'	β_2, β_2'	β_3, β_3'	χ^2 test $\beta_1=\beta_2=\beta_3$, $\beta_1'=\beta_2'=\beta_3'$
Soybean meal[a]	0.617**	0.397**	–	2.790
	0.069	0.166	–	0.668
Soybean oil[a]	0.505**	0.354**	–	1.057
	−0.007	−0.055	–	0.111
Wheat	0.526**	0.593**	0.680**	0.440
	0.114	−0.016	−0.064	0.948
Coffee	0.882**	0.476*	0.632**	4.029
	−0.224	0.295	0.021	7.914*
Sugar	0.035	0.348**	0.642**	62.091**
	0.690**	0.279*	0.128	20.102**
Cocoa	0.397*	0.436**	0.476**	0.233
	0.269*	0.123	0.225	1.663
Cotton	0.591**	0.407**	0.634**	5.343
	−0.014	0.253*	0.031	4.991
Orange juice	0.478**	0.286**	0.350**	7.382**
	−0.064	0.170	0.039	4.772*
Feeder cattle	0.414**	0.399**	0.088	1.750
	0.319**	0.197	0.354	1.437
Live cattle[a]	0.594**	0.465**	–	0.984
	0.025	0.246*	–	2.376
Lean hogs[a]	0.504**	0.368**	–	0.929
	0.223*	0.282	–	0.147

We test whether the relations among RV, IV and HV depend on the option's term to maturity according to Regression (8). We use the Hansen (1982) correction to estimate coefficient standard errors in the presence of overlapping observations. The constant terms and R^2 are not reported above. For certain commodities, there were very few observations with $\geqslant 50$ trading days, so the β_3 and β_3' coefficients could not be reliably estimated. We could not run these regressions for Crude oil, heating oil, unleaded gasoline and natural gas because, given the monthly expiration cycle, these options always had 33 or fewer trading days to maturity.

* and **, respectively, indicate significance at the 5% and 1% levels.

[a] For these commodities, β_3 and β_3' coefficients could not be reliably estimated, and the χ^2 tests are for $\beta_1=\beta_2$ and $\beta_1'=\beta_2'$.

7 out of 31 markets, the results are far from being homogeneous: In the soybean, Japanese Yen, orange juice, and the Notional 10-year Euro Bond contracts, β_1 exceeds both β_2 and (where relevant) β_3, whereas in the sugar and gold markets, the slope coefficients rise monotonically with term to maturity. In the case of the British Pound options, β_2 is low, but β_3 exceeds β_1. Further, in 19 out of 31 markets, the estimated β_2 is less than β_1; based on the binomial probability model in Eq. (7), the null hypothesis that these coefficients are in fact equal would only be rejected (at the 0.05 level) if β_2 were less than β_1 in 22 or more (or 9 or fewer) cases. Similarly, for those 25 markets for which β_3 could be estimated, it is less than β_1 in only 10 cases; in

order to reject the null of no term-to-maturity effect, β_3 would have to be lower 18 or more (7 or fewer times). [16]

4. The GARCH model and implied volatility

4.1. Hypothesis and test procedures

Financial market volatility may be predictable (Engle, 1993) and the conditional variance obtained from an ARCH model, or one of its extensions, might provide good estimates of future volatility using historical data. If financial market volatility is predictable, multi-day variance forecasts from an ARCH-type model might constitute a more sophisticated way of using historical data to estimate future volatility than the Canina and Figlewski (1993) type tests employed in the previous section. [17] In this section, we follow Day and Lewis (1992) and Jorion (1995) and test whether conditional variance forecasts from GARCH models can be used to improve the predictive power of IV. Specifically, we test the following hypothesis using GARCH models:

H4. The GARCH model forecast of future volatility does not contain any additional information beyond what is already contained in IV.

To generate the forecasted GARCH conditional variance series to test this hypothesis, we follow Jorion (1995) and estimate the following GARCH (1,1) model with daily data for each contract: [18]

$$R_t = \mu + \varepsilon_t; \qquad \varepsilon_t \sim N(0, h_t^2); \qquad h_t^2 = \delta + \gamma\varepsilon_{t-1}^2 + \theta h_{t-1}^2, \tag{9}$$

where R_t is the return on underlying futures contract in period t; μ is the mean of the return; ε_t is the shock on series y in period t; h_t^2 is the conditional variance of return. We use all available data for each observation to obtain GARCH model estimates. We then recursively forecast conditional variances until the option's

[16] Yet another reason our results may differ from those in previous studies is related to the moneyness of the options used to calculate IV. The Bridge data used in this study employ at-the-money options, while some previous studies used a weighting scheme which includes options far from the money. However, Canina and Figlewski (1993) find that even for near-the-money options, IV is a poor predictor for the cash OEX options, thus, the moneyness of options used is unlikely to be the main factor driving our results.

[17] ARCH modeling began with Engle (1982) and is surveyed in Engle (1993). Bollerslev (1986) introduced the GARCH model that accommodates volatility persistence. The volatility persistence is well documented by French et al. (1987), Bollerslev et al. (1988), Akgiray (1989), Day and Lewis (1992) and Lamoureux and Lastrapes (1993). For many time series models the main criticism concerns the subjective selection of a model. For ARCH models, Nelson (1992) studies the effect of misspecification of the model. He finds that with high-frequency data, in the class of GARCH processes, even misspecified GARCH models provide good estimates of volatility.

[18] Akgiray (1989) finds that GARCH (1,1) provides the best fit for economic variables within the class of GARCH models. Day and Lewis (1992) also find the GARCH (1,1) model has a better fit relative to higher order models.

maturity date using only the information available for that observation. [19] We then average each set of recursive forecasts to obtain the GARCH forecast for the period, GFOR_t. We use actual historical data only through day t (the day on which the IV is observed); the conditional variances for the subsequent days until maturity are estimated recursively via the GARCH model parameters and predicted conditional variances for previous days. The ADF tests for GFOR, presented in Table 2, reject the null hypothesis of a unit root for all GARCH-predicted volatility series.

After we obtain the series of GARCH forecasts we run regressions that are similar to (6), except that we replace HV with the GARCH forecasts (GFOR_t): [20]

$$\mathrm{RV}_t = \alpha + \beta \mathrm{IV}_t + \beta'' \mathrm{GFOR}_t + e_t. \tag{10}$$

Note that GFOR_t are forecasted over the remaining life of the option, and as a result, match the period over which the IVs are estimated. Since, as in Jorion (1995), our GARCH model parameters are estimated with the entire data for each commodity, they still (to a small extent) use information that would not have been available on the date the IV is obtained; therefore, if anything, our procedure may have a slight bias towards finding greater information content in the GARCH forecasts than would actually be present in practice.

4.2. GARCH forecasted volatility results

A comparison of the results of regression (10) reported in Table 5 to those of regression (6) reported in Table 3, reveals that the GARCH forecasts, generally, are not markedly superior as a supplemental predictor of RV. [21] For example, the average R^2 from regression (6) increased from 36.4% in Table 3 to only 36.6% for regression (10) in Table 5. The R^2 from regression (10) exceeds the R^2 from regression (6) by more than 1% point only in five cases (5-year treasury notes, feeder cattle, lean hogs, British Pound, and copper), in one case (sugar) the regression (10) R^2 is 1.2% lower, and for the remaining 29 commodities the R^2's of the two regressions are within 1%. Likewise, in regression (10) just as in regression (6), 34 of the 35 coefficients on IV are positive and significant, with 30 of these significant at the 1%

[19] Specifically, we perform the following procedure to obtain recursive GARCH volatility forecasts, GFOR_t. First GARCH estimates are obtained using all available data for each option. Second, observation on day t is used to obtain one-day-ahead forecast h^2_{t+1}. Third, using h^2_{t+1} we draw an estimate of ε^2_{t+1} from $N(0, h^2_{t+1})$. Fourth, h^2_{t+1} and ε^2_{t+1} are used to obtain one-day-ahead forecast h^2_{t+2}. Third and fourth steps are repeated until the option maturity is reached. Fifth, h^2_{t+1} through h^2_{t+T} are averaged to obtain GFOR_t. We then move to observation $t+1$ and repeat the procedure to obtain GFOR_{t+1}, and so on. Note that while the recursive forecasts (GFORs) rely on coefficient estimates obtained using the whole sample, only day t's information is used in each forecast to obtain GFOR_t.

[20] Day and Lewis (1992) proposed an alternative test that adds the IV estimate to the equation of conditional variance in the GARCH model. A likelihood ratio test is then used to determine whether the addition of the IV term significantly improves the GARCH model. The Day and Lewis (1992) test suffers from the fact that measurement of returns (weekly) and IVs (over option's remaining life) differ. As a result, we do not implement their test in our paper.

[21] To conserve space, the actual GARCH (1,1) models we estimated are not reported. They are available from the authors upon request.

Table 5
Predicting RV with IV and GARCH model forecasts

	α	β	β''	Adjusted-R^2
S&P500 Index	2.972**	0.618**	−0.018	0.230
	2.816	7.319	−0.485	
Treasury bond	0.994	0.296*	0.519	0.150
	0.427	2.077	1.708	
Treasury notes 10 year	0.864	0.295*	0.452**	0.186
	0.716	2.501	2.715	
Treasury notes 5 year	0.499	0.290**	0.458**	0.244
	0.773	2.678	2.637	
Eurodollar	4.454**	0.566**	0.024	0.399
	4.453	7.028	0.319	
Treasury bills	10.519**	0.374**	−0.015	0.105
	3.128	4.242	−0.082	
Long UK Gilt	2.012	0.387*	0.215	0.293
	1.405	2.460	0.742	
German Bond	0.704	0.639**	0.032	0.456
	1.436	8.872	0.279	
Short Sterling	3.242*	0.583**	0.092*	0.351
	2.315	7.230	2.057	
Notional Bond, Euro 10 year	0.062	0.368**	0.438*	0.505
	0.113	4.254	2.091	
Japanese Yen	2.944**	0.469**	0.179	0.253
	2.670	3.949	1.504	
Deutsche Mark	3.373**	0.471**	0.115	0.256
	2.616	6.423	1.015	
Canadian Dollar	1.572**	0.475**	0.033	0.285
	2.994	4.891	0.323	
British Pound	1.660	0.392**	0.347**	0.409
	1.675	3.864	2.689	
Swiss Franc	3.381	0.350**	0.311	0.197
	1.615	3.813	1.762	
Crude oil	−0.048	0.861**	−0.165	0.735
	−0.026	6.559	−1.535	
Heating oil no. 2	5.938*	0.718**	−0.136	0.580
	2.432	4.254	−1.065	
Gasoline unleaded	5.946**	0.803**	−0.264*	0.631
	3.512	5.977	−2.075	
Natural gas	4.675	0.573**	0.085	0.432
	1.145	6.924	1.010	
Copper	6.141	0.217*	0.409*	0.209
	1.873	2.168	1.979	
Gold	1.953	0.557**	0.045	0.369
	1.919	6.129	0.608	

(*continued on next page*)

Table 5 (*continued*)

	α	β	β''	Adjusted-R^2
Silver	4.120	0.603**	0.055	0.365
	1.286	4.768	0.325	
Corn	0.738	0.635**	0.056	0.514
	0.409	8.149	0.665	
Soybeans	1.543	0.588**	0.121	0.425
	0.754	8.620	1.545	
Soybean meal	3.387	0.552**	0.131*	0.388
	1.891	6.878	1.960	
Soybean oil	6.530**	0.425**	0.099	0.267
	3.549	5.519	1.160	
Wheat	4.400*	0.592**	0.029	0.366
	2.287	6.251	0.325	
Coffee	5.159	0.692**	0.008	0.340
	1.226	5.583	0.085	
Sugar	3.868	0.133	0.672**	0.469
	1.480	1.635	4.776	
Cocoa	4.032	0.465**	0.218	0.387
	1.595	4.599	2.061	
Cotton	3.490	0.501**	0.203	0.396
	1.078	6.786	1.276	
Orange juice	14.931**	0.384**	−.040	0.367
	4.191	8.244	−0.379	
Feeder cattle	1.258	0.375**	0.404**	0.424
	1.480	6.007	5.047	
Live cattle	1.699	0.535**	0.171*	0.545
	1.691	9.132	2.448	
Lean hogs	0.598	0.421**	0.430**	0.530
	0.313	6.297	3.164	

We test whether GARCH Model forecasts contain information that is not captured by IV with Regression (10), $\mathrm{RV}_t = \alpha + \beta \mathrm{IV}_t + \beta'' \mathrm{GFOR}_t + e_t$, where RV_t is the annualized RV for T_{M} days (i.e., from day t to the futures option's maturity), IV_t is the IV, GFOR_t is the annualized forecasted volatility from recursive estimates of a GARCH (1,1) model over T_{M} days, using actual returns through day t. We use the Hansen (1982) correction to adjust the coefficient standard errors to reflect varying lengths of overlap in the data. * and ** represent 5% and 1% significance levels, respectively.

level. The only difference between the IV and GFOR results in Table 5 and the IV and HV results in Table 3 is the somewhat greater propensity for GFOR to be significant. For 12 of the 35 commodities, GFOR is positive and significant, indicating that the GARCH model forecasts contain some information regarding RV that is not contained in IV. However, for unleaded gasoline, GFOR is significantly negative (just as HV was in Table 3), and for the remaining 22 commodities we find that GFOR is insignificant.

Based on binomial probabilities, the probability of finding significant positive coefficients on GFOR at the 5% level in regression (10), for 12 of 35 commodities, if the null hypothesis is that the coefficient on GFOR equals zero for *all markets*, is very close to zero. Similarly, we observe that 29 of the GFOR coefficients are positive. If the probability of obtaining a positive coefficient in each regression were 0.50, then the binomial model indicates that the chance of observing 29 or more positive coefficients (out of a possible 35) is less than 0.0001. Thus, our results with respect to H4 are somewhat mixed. On the one hand, we can confidently state that H4 does not hold for at least some markets. At the same time, however, we must note that GARCH forecasts do not appear to add much predictive power in the majority of cases, and even when they do, the predictive power of IV generally remains strong.

5. Conclusion

Using data from 35 options on futures, we test how well the IVs embedded in options prices predict subsequently RV in the underlying futures. Our tests analyze the unbiasedness of IV as a predictor, as well as the usefulness of IV relative to other alternatives (such as HV or forecasts from a GARCH model) that are commonly employed to predict future volatility. Our study is an improvement over prior research because we examine markets in which the options and their underlying futures trade simultaneously with relatively low transactions costs and minimal trading frictions, and because we survey a much broader variety of underlying asset classes (e.g. stocks, bonds, money market securities, currencies, agricultural commodities, industrial commodities, metals, etc.) than in previous studies.

We find that for this broad array of futures options, IV, though not a completely unbiased predictor of future volatility, performs well in a relative sense. For an overwhelming majority of the 35 commodities studied, IV outperformed HV as a predictor of the subsequently RV in the underlying futures prices over the remaining life of the option. Furthermore, in most markets, HV does not appear to contain any information that is not already incorporated in the IV. These results appear to be robust across differing terms to maturity, and for S&P 500 options, across sample periods as well. When we replace simple 30 day moving average HVs with recursive forecasts from GARCH models, we find little difference in predictive power. Our findings are qualitatively similar to those in Christensen and Prabhala (1988) for equity indices and are consistent with the weak-form efficiency of futures options markets, in that the volatility information embedded in current option prices is a better predictor of future volatility than historical measures of volatility, regardless of how the latter are modeled.

In all of the markets we examine, the futures and options contracts trade on the same exchange, their closing prices are less likely to be subject to non-synchronous trading problems, and transactions costs are relatively low. IV, while not a completely unbiased forecast of the RV, appears to dominate HV in virtually all of these markets. Despite considerable effort, we have been unable to discern any obvious pattern in our results, either by exchange, type of commodity, sample period, or remaining term to maturity. Thus, our findings confirm and extend Jorion's (1995) results for currency futures options. They indicate that futures options markets in general, and by extension, markets in which trading frictions are minimal, do tend to be relatively efficient.

Acknowledgements

The authors would like to express thanks to the anonymous referee and the associate editor who provided valuable insights and suggestions.

References

Akgiray, V., 1989. Conditional heteroscedasticity in time series of stock returns: Evidence and forecasts. Journal of Business 62 (1), 55–80.

Barnhart, S.W., Szakmary, A.C., 1991. Testing the unbiased forward rate hypothesis: Evidence on unit roots, co-integration, and stochastic coefficients. Journal of Financial and Quantitative Analysis 26 (2), 245–268.

Becker, K.G., Tucker, A.L., 1991. Implied index volatilities and intraweek effects in the US equity market. Advances in Futures and Options Research 5 (1), 297–308.

Beckers, S., 1981. Standard deviations implied in option prices as predictors of future stock price variability. Journal of Banking and Finance 5 (3), 363–381.

Black, F., 1976. The pricing of commodity contracts. Journal of Financial Economics 3 (1/2), 167–179.

Black, F., Scholes, M., 1973. The pricing of options and corporate liabilities. Journal of Political Economy 81 (3), 637–654.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31 (3), 307–328.

Bollerslev, T., Engle, R.F., Wooldridge, J.M., 1988. A capital asset pricing model with time-varying covariances. Journal of Political Economy 96 (1), 116–131.

Canina, L., Figlewski, S., 1993. The informational content of implied volatility. Review of Financial Studies 6 (3), 659–681.

Chiras, D.P., Manaster, S., 1978. The information content of option prices and a test of market efficiency. Journal of Financial Economics 6 (2/3), 213–234.

Christensen, B.J., Prabhala, N.R., 1988. The relation between implied and realized volatility. Journal of Financial Economics 50, 125–150.

Davidson, Wallace, Kim, Jin Kyoung, Ors, Evren, Szakmary, Andrew, 2001. Using implied volatility on options to measure the relation between asset returns and variability. Journal of Banking and Finance 25, 1245–1269.

Day, T.E., Lewis, C.M., 1992. Stock market volatility and the information content of stock index options. Journal of Econometrics 52 (1/2), 267–288.

Day, T.E., Lewis, C.M., 1993. Forecasting futures market volatility. Journal of Derivatives, 33–50.

Dickey, D., Fuller, W., 1979. Distribution of estimators for autoregressive time series with a unit root. Journal of the American Statistical Association 74, 427–431.

Ederington, L.H., Lee, J.H., 1996. The creation and resolution of market uncertainty: The impact of information releases on implied volatility. Journal of Financial and Quantitative Analysis 31 (4), 513–539.

Engle, R.F., 1993. Statistical models for financial volatility. Financial Analysts Journal 49 (1), 72–78.

Engle, R.F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflations. Econometrica 50 (4), 987–1008.

Fleming, J., Ostdiek, B., Whaley, R.E., 1995. Predicting stock market volatility: A new measure. Journal of Futures Markets 15 (3), 265–302.

Fleming, J., Ostdiek, B., Whaley, R.E., 1996. Trading costs and the relative rates of price discovery in stock, futures and options markets. Journal of Futures Markets 16 (4), 353–387.

French, D.W., 1984. The weekend effect on the distribution of stock prices: Implications for option pricing. Journal of Financial Economics 13 (4), 547–559.

French, K.R., William Schwert, G., Stambaugh, R.F., 1987. Expected stock returns and volatility. Journal of Financial Economics 19 (1), 3–30.

Fuller, W., 1976. Introduction to Statistical Time Series. New York, Wiley.

Hansen, L.P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50 (4), 1029–1054.

Harvey, C.R., Whaley, R.E., 1992. Market volatility prediction and the efficiency of the S&P 100 index option market. Journal of Financial Economics 31, 43–74.

Ignacio, P., Rubio, G., Serva, G., 1999. Why do we smile? On the determinants of the implied volatility function. Journal of Banking and Finance 23 (8), 1151–1179.

Jackwerth, J.C., Rubinstein, M., 1996. Recovering probability distributions from option prices. Journal of Finance 51 (5), 1611–1631.

Jorion, P., 1995. Predicting volatility in the foreign exchange market. Journal of Finance 50 (2), 507–528.

Kumar, R., Shastri, K., 1990. The predictive ability of stock prices implied in option premia. Advances in Futures and Options Research 4 (1), 165–176.

Lamoureux, C.G., Lastrapes, W.D., 1993. Forecasting stock-return variance: Toward an understanding of stochastic implied volatilities. Review of Financial Studies 6 (2), 293–326.

Latane, H.A., Rendleman Jr., R.J., 1976. Standard deviations of stock price ratios implied in option prices. Journal of Finance 31 (2), 369–381.

Nelson, D.B., 1992. Filtering and forecasting with misspecified arch models I: Getting the right variance with the wrong model. Journal of Econometrics 52 (1/2), 61–90.

Park, H.Y., Stephen Sears, R., 1985. Changing volatility and the pricing of options on stock index futures. Journal of Financial Research 8 (4), 265–274.

Randolph, W.L., Rubin, B.L., Cross, E.M., 1990. The response of implied standard deviations to changing market conditions. Advances in Futures and Options Research 4 (1), 265–280.

Rubinstein, M., 1985. Nonparametric tests of alternative option pricing models using all reported trades and quotes on the 30 most active CBOE option classes from August 23, 1976 through August 31, 1978. Journal of Finance 40 (2), 455–480.

Schmalensee, R., Trippi, R.R., 1978. Common stock volatility expectations implied by option premia. Journal of Finance 33 (1), 129–147.

Schwert, G.W., 1989. Tests roots: A Monte Carlo investigation. Journal of Business and Economic Statistics 7 (2), 147–160.

Sims, C.A., Stock, J.H., Watson, M.W., 1990. Inference in linear time series models with some unit roots. Econometrica 58 (1), 113–144.