Available online at www.sciencedirect.com
ScienceDirect
Journal of Empirical Finance 14 (2007) 333 – 354
Journal of EMPIRICAL FINANCE
www.elsevier.com/locate/jempfin

The implied volatility term structure of stock index options ☆

Scott Mixon *

Bates White, LLC, 1300 Eye Street NW, Suite 600, Washington, DC 20005, United States

Accepted 19 June 2006
Available online 30 January 2007

Abstract

This paper tests the expectations hypothesis of the term structure of implied volatility for several national stock market indexes. The tests indicate that the slope of at-the-money implied volatility over different maturities has predictive ability for future short-dated implied volatility, although not to the extent predicted by the expectations hypothesis. The low forecast power may be due to failure to control for a risk premium in the prices of the options. Evidence is presented that a time-varying risk premium proportional to the level of market volatility is consistent with the results.

JEL classification: G13
Keywords: Implied volatility; Volatility term structure; Expectations hypothesis; Volatility risk premium

What does the term structure of implied volatility mean? Under the benchmark Black–Scholes model, it shouldn't exist, or more precisely, it should be a flat, uninteresting line. In practice, the term structure often slopes upward, sometimes downward, but the underlying causes are not obvious. Behavioral economists frame the issue in terms of overreaction and mispricing of assets in predictable ways. Option theorists resort to complicated stochastic volatility models to describe the data. Market participants lean on "supply and demand" to rationalize the phenomenon and leave it at that.

☆ I would like to thank participants of the Eastern Finance Association conference and an anonymous referee for the helpful comments. This paper represents opinions of the author and does not necessarily represent the opinions of Bates White, LLC.
* Tel.: +1 202 747 5973.
E-mail address: scott.mixon@bateswhite.com.

0927-5398/$ - see front matter
doi:10.1016/j.jempfin.2006.06.003

This paper investigates whether a simple, rational economic explanation — the expectations hypothesis — can explain the behavior of the implied volatility term structure for at-the-money stock index options. The intuition and tests are the same as for the expectations hypothesis of the term structure of interest rates. The central question is whether the slope of the term structure predicts future changes in implied volatility. That is, does an upward (downward) sloping term structure predict that volatility will increase (decrease)? If the theory doesn't work well, can it be rehabilitated by accounting for the risk neutral nature of the volatility forecasts embedded in option prices?

This paper represents an improvement on the existing literature in the following ways. First, the major contribution of this paper is to focus on the source of deviations from the expectations hypothesis; I do not simply stop after rejecting the expectations hypothesis. The crucial insight is that the implied volatility term structure is based on the risk neutral measure, whereas the real-world evolution of (implied and realized) volatility occurs under the objective measure. I find that after accounting for the slope of the term structure, current volatility is highly significant in forecasting regressions; this is strong evidence that implied volatility forecasts deviate substantially from objective forecasts. The behavior of the implied volatility term structure is consistent with the volatility risk premium model of Heston (1993) and does not necessarily represent investor "misreaction." Second, the paper investigates the expectations hypothesis using a rich dataset for several national market indexes, rather than solely for the S&P 500. Finally, the data represent over-the-counter volatility quotes for equity index options with constant 1, 3, 6, and 12 month maturities over a longer span of data than most research. This unique data set gives a much more precise test than most earlier work, which focused on exchange traded options with fixed expiration dates only 1 or 2 months away.

Careful examination of the expectations hypothesis is important because the results provide a natural benchmark for evaluating models of volatility dynamics. Dai and Singleton (2002), for example, use the stylized facts from the expectations hypothesis to discriminate among interest rate term structure models. The analysis yields insights into the formation of market expectations for volatility and the price of volatility risk. By using option data, stock market volatility dynamics are explored using a more comprehensive information set than time series models utilize. For example, the well known gap between implied volatility and future sample realized volatility (Fleming, 1998) can be explained by a volatility risk premium or by the expectation of a large crash that does not occur in the sample, but the competing hypotheses cannot be distinguished using historical stock price data alone. Using implied volatility for a single maturity cannot shed much light on the origin of the gap. However, the two explanations have different implications for the dynamics of the implied volatility term structure.

Previous research has also examined the behavior of the implied volatility term structure, although no consensus has emerged. Poterba and Summers (1986) found that the volatility implicit in long-dated equity index options did not move much due to volatility shocks. Stein (1989), using one- and two-month S&P 100 options from 1983–1987, concluded that long-dated options overreact to volatility shocks. Diz and Finucane (1993) examined a similar S&P 100 data set for the 1985–1988 period and reached the opposite conclusion. Campa and Chang (1995) investigated the term structure of implied volatility for foreign exchange options during the 1989–1992 period; they were unable to reject the expectations hypothesis in most instances. Poteshman (2001) used 1988–1997 S&P 500 option data and extended Stein's results. Byoun, Kwok, and Park (2003) conclude that the expectations hypothesis gives a poor description of the volatility term structure for both the S&P 500 and foreign currency options. (Campbell and Shiller, 1991 review the empirical evidence for the term structure of interest rates, which is generally negative.)

Conclusions from the empirical work are summarized as follows. First, the expectations hypothesis fails. The slope of the volatility term structure has significant predictive ability for future short-dated implied volatility, although not to the extent predicted by the expectations hypothesis. The parameter estimates are inconsistent with the hypothesis and the explained variation of implied volatility is extremely low. Monte Carlo tests which explicitly account for small sample bias confirm this rejection is not due to reliance on asymptotic inference. Second, basic tests of the expectations hypothesis omit the instantaneous volatility, which Heston-type models require. When the current level of realized volatility is included in the regressions, the results conform to the predictions of the model. The coefficient on instantaneous volatility is negative and significant: the implied volatility term structure tends to be upwardly biased as a predictor of future implied volatility, and the bias is a function of realized volatility. The model also predicts that this overprediction bias tends to increase as the forecast horizon lengthens, and I also confirm this finding in the data.

The paper is structured as follows. Section 1 reviews the theory behind tests of the expectations hypothesis for implied volatility. Section 2 describes the data used in the tests. Section 3 presents the empirical evidence on the expectations hypothesis of the term structure. Section 4 goes on to examine the forecasting performance of forward implied volatility. Section 5 contains concluding remarks.

1. Theory

In this section, I show how the expectations hypothesis can be applied to the implied volatility term structure. The first subsection describes the theory in a benchmark world with stochastic volatility that is uncorrelated with the underlying asset. The second subsection evaluates the applicability of the theory for stock index options. The final subsection describes how the theory must be modified to allow for a volatility risk premium in option prices.

1.1. Deriving the expectations hypothesis

Campa and Chang (1995) explicitly tested the expectations hypothesis for foreign exchange options and could not reject it. They assume that shocks to the asset and the volatility are uncorrelated and that there is no volatility risk premium. Using the results due to Hull and White (1987), they derive the expectations hypothesis relations among implied volatilities of different maturities. They conclude that the slope of the volatility term structure has significant predictive ability for future implied volatility and that there is little evidence for overreactions of the long rate.

Campa and Chang derive the relation for current and future expected at-the-money implied volatilities. Define m as the number of months until expiration for the short-dated option, k as the number of periods of length m, and $\sigma^2_{t,T}$ as the (squared) implied standard deviation, quoted at time t for an option expiring at time T. They show that the long-dated implied variance equals the average expected value of the sequence of short-dated implied variances spanning the same time until expiration. More precisely, they show that

$$\sigma^2_{0,km} = (1/k)E_0\left[\sum_{i=0}^{k-1}\sigma^2_{im,(i+1)m}\right](\theta_{km}/\theta_m)^2. \tag{1}$$

The constant $(\theta_{km}/\theta_m)^2 < 1$ enters because the Black–Scholes option price (for options struck at the forward) is nearly linear in volatility, yet slightly concave. To see this, recall that Hull and

White show that the appropriate option price for an option quoted at time 0 and expiring at time t (under the conditions described above) equals the expected Black–Scholes price, evaluated at the average variance over the life of the option: $C^{\mathrm{HW}} = E[C^{\mathrm{BS}}(\bar{\sigma}^2_{0,t})]$. The concavity of the Black–Scholes formula in σ implies that, for at-the-money options, $C^{\mathrm{HW}} = \theta_t C^{\mathrm{BS}}(E[\bar{\sigma}^2_{0,t}])$, where $\theta_t < 1$ decreases as the time to expiration for the option grows. The Black–Scholes model evaluated at the average variance overprices ATM options, and this overpricing increases with time to expiration; θ_t corrects for this. The concavity is small for at-the-money options, which leads to the approximation $\theta_t C^{\mathrm{BS}}(E[\bar{\sigma}^2_{0,t}]) \approx C^{\mathrm{BS}}(\theta_t^2 E[\bar{\sigma}^2_{0,t}])$. This expression relates the average variance over the life of the option to the implied Black–Scholes variance. With this result, one can compare short-dated and longer dated Black–Scholes option volatilities and show that Eq. (1) holds. This result is generally true when shocks to volatility and the underlying are independent and there is no volatility risk premium.

Initially, it is assumed that the ratio (θ_{km}/θ_m) equals 1; the nature of the resulting bias is described later. Hence, the basic relationship among volatility quotes is:

$$\sigma^2_{0,km} = (1/k)E_0 \sum_{i=0}^{k-1} [\sigma^2_{im,(i+1)m}]. \tag{2}$$

In Eq. (2), the current volatility quote equals the average of current and expected future short-dated volatility quotes. For example, the current 6 month rate (quoted now) is the average of the current 3 month rate (quoted now) and the expected 3 month rate (quoted in 3 months). This corresponds to the traditional expectations hypothesis of interest rates, save for a term premium. The implication is that the slope of the term structure is informative about where the market believes implied volatility will be in the future.

1.2. Appropriateness of the theory

There are issues about using the implications of Eq. (2) for stock index options. First, the relationship was derived assuming zero correlation between shocks to the asset's price and its volatility. This assumption is not true for stock indexes, but the simulation evidence by Lamoureaux and Lastrapes (1993) and Heynen, Kemna, and Vorst (1994) indicates that the non-zero correlation may not be a significant problem. Using realistic parameter values, the authors conclude that the average expected volatility and the implied volatility for at-the-money options are virtually indistinguishable, even when the asset's price and volatility are correlated.

Second, the derivation is appropriate for options struck at-the-money forward,[1] but the data available for this study consists of options struck at-the-money spot. While the true variance forecast is not at-the-money volatility, it is closely related, even in the presence of arbitrary volatility skews. To understand how adequate the approximation is, I compare the VXO volatility index with the VIX volatility index. The VXO was the original VIX index introduced by the Chicago Board Options Exchange in 1993. It represents a hypothetical, one-month at-

[1] The fact that the derivation relies on options struck at the forward price is intuitive, given that options are priced under the risk-neutral measure. Under this measure, the asset grows at the cost of carry, and the expected value equals the forward price. Volatility is, of course, measured relative to the expected value of the random variable. Technically, the derivation relies on the expected value of the Black–Scholes option price, and the approximate linearity of the Black–Scholes function for options struck at the forward enables one to reverse the order of the expectation operator and option pricing function.

the-money implied volatility on the S&P 100 index. The VIX, launched in 2003, represents the volatility strike for a 1 month variance swap (the risk neutral expected variance) on the S&P 500.

If the risk neutral expected volatility is well proxied by at-the-money implied volatility, the VIX and VXO should track each other quite closely. Values of both series exist for the 1990–2003 sample period, and the correlation between the two series is greater than 0.98. Using this sample, I regress the weekly (Friday) value of the VIX minus the VXO on a constant and the VXO.[2] The constant is significant with a value of 1.78 vol points and the slope coefficient has a value of 0.003 (t-statistic = 0.37). The R^2 of the regression is 0.03%. Informal comparisons of OTC variance swaps with at-the-money volatility for a variety of expirations provide similar results, with similar values for the constant term. The regression provides little evidence the approximation is systematically incorrect, aside from a roughly constant arithmetic scaling. Nonetheless, as more actual market observations on variance swaps becomes available, they might provide a cleaner test of the expectations hypothesis.[3]

Third, the analysis presented here assumes that the shape of the implied volatility term structure depends on dynamics in the volatility process. Another approach is to allow for interesting dynamics in the drift of the underlying asset price process rather than the diffusion part. Lo and Wang (1995) show how, given an unconditional variance of returns, the appropriate implied volatility for use in option prices depends on the particular drift process. Black (1989) suggests that mean reversion in stock prices would lead to a downward sloping volatility term structure (i.e., that root — T scaling of volatility would overstate the correct volatility). Given the typical upward slope of the implied volatility term structure and the difficulty in estimating the correct drift of stock index prices, I interpret the results in terms of volatility dynamics.

Finally, the assumption is made that volatility risk is not priced in equilibrium. Because volatility is negatively correlated with stock indexes, which account for a substantial portion of nominal wealth, this assumption is implausible. There is indirect and direct evidence that option prices incorporate premiums for risk. Fleming (1998) finds that index option volatility systematically overstates future realized volatility. Bakshi and Kapadia (2002), Buraschi and Jackwerth (2001), Coval and Shumway (2001), Fleming (1999), and Pan (2002) each use different approaches and all conclude that additional risk factors, such as stochastic volatility, are priced in index option markets. These findings indicate that the expectations in (1) and (2) should be interpreted as expectations under the risk neutral measure, not the objective measure. It also suggests that the more general theory should be examined, which is done next.

1.3. The volatility term structure under the Heston model

The empirical plausibility of the expectations hypothesis under the risk neutral measure can be evaluated by considering a parametric risk adjustment. Consider the dynamics of instantaneous

[2] I also rescaled the VXO by a factor of $\sqrt{22}/\sqrt{30}$ to convert index values into a per-calendar day annual rate, which is the typical quoting convention. Both series are expressed in volatility percentage points.

[3] Using out-of-the-money options could be a promising way to extend this research. This would likely require developing theory that describes the joint evolution of volatility and crash expectations more fully than is considered here. Far out-of-the-money options are much more sensitive to jump expectations (or model dynamics) than at-the-money options are. Using at-the-money options focuses the analysis squarely on volatility dynamics, whereas analysis incorporating out-of-the-money options moves the research closer to ambitious parametric modeling of the entire asset price process.

squared volatility, σ^2. A popular specification of risk neutral stochastic volatility, described by Heston (1993), is

$$d\sigma^2 = (k\theta - k\sigma^2 - \lambda\sigma^2)dt + \gamma\sqrt{\sigma^2}dz, \tag{3}$$

where $k>0$ is a mean reversion parameter, $\lambda<0$ is the price of volatility risk, and θ is the long-run mean for squared volatility. Innovations to the Wiener process dz are scaled by the product of the parameter γ and the current level of volatility. As discussed above, it is plausible to ignore the complications caused by the negative correlation between volatility and returns when discussing at-the-money options. In that case, while the objective volatility process is mean reverting, the risk neutral process is mean reverting at a slower rate, or is possibly non-stationary if $\lambda \leq -\mathrm{k}$. The implication is that the option implied volatility (approximately the average expected risk neutral volatility) is systematically biased away from the expected implied volatility if $\lambda \neq 0$.

The risk adjustment causes the relationship in Eq. (2) to fail in particular directions. Fig. 1 displays the consequences of a non-zero risk premium. The figure shows the deterministic behavior in the square root of average variance due to the drift in Eq. (3) using plausible values. In the example, $k=0.12$, $\lambda=-0.035$, and $\theta=(17.25\%)^2/52$ (time is measured in weeks). Although the square root of objective variance tends toward 17.25%, the risk neutral process, on which options are priced, tends toward $k\theta/(k+\lambda)=(20.5\%)^2/52$, or 20.5% on an annualized basis. Because of this, the slope of the observed term structure will not accurately predict the change in variance if a risk premium exists.

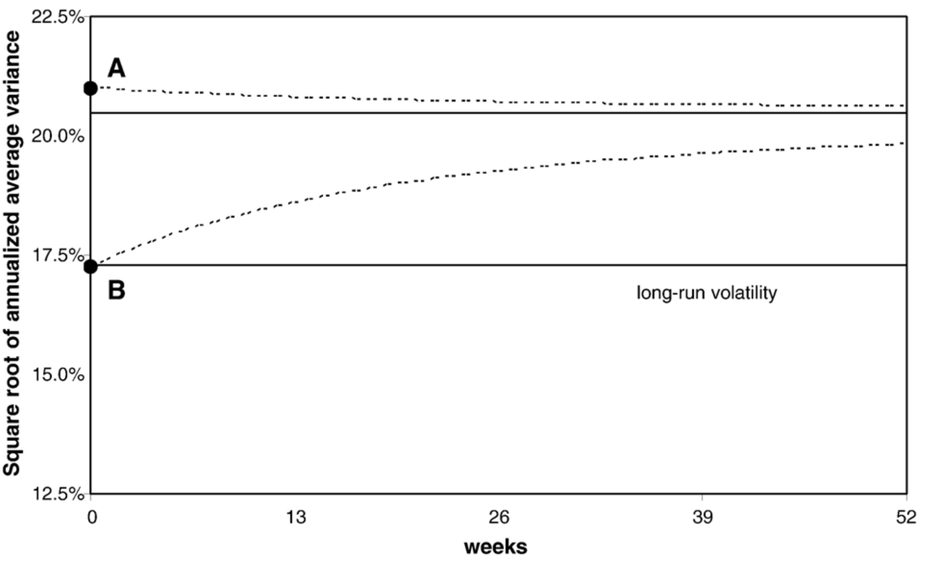

Fig. 1. Hypothetical risk neutral average volatility path. Hypothetical profile of expected average volatility for risk neutral volatility with a long run mean of 20.5%. The deterministic drift examined is $d\sigma^2=(k\theta-k\sigma^2-\lambda\sigma^2)dt$, where $k=0.12$, $\lambda=-0.035$. The dashed lines are the expected risk-neutral paths for average volatility after starting at initial point A (21%) and initial point B (17.25%). The square root of the long-run mean for variance under the objective measure is 17.25%; the square root of the long-run mean under the risk-neutral measure is 20.5%.

Fig. 1 is worth discussing in more detail. It shows the future path of the square root of average risk neutral variance over time (i.e., the implied volatility term structure) for two different initial points. From point A, a 21% volatility rate, the resulting implied term structure is quite flat, even though the model suggests that the square root of variance will actually revert to the long-run mean of 17.25%. From point B, a 17.25% volatility rate, the resulting term structure is substantially upward sloped, even though variance is already at its long-run steady state value. The predicted change in the short-dated variance under the risk neutral measure will typically be upwardly biased as a prediction of the expected change in the objective variance.

It is possible to derive more precisely how the expectations hypothesis fails. Using a simplified version of the Heston model, I examine the case where the long-dated option expires at time T and the short-dated options of interest expire at time $T/2$. This case directly applies to the situation where the long-dated options expire in 6 months, for example, and the short-dated options expire in 3 months. For this case, the following proposition shows the exact form of the omitted variable in the expectations hypothesis regressions (i.e., Eq. (4)) under this implementation of the Heston model.

Proposition 1. *Assume that variance evolves under the risk neutral measure according to* Eq. (3)*, with $k>0$, $\lambda<0$, $k+\lambda\neq 0$. Define implied variance $\sigma^2_{t,T}=\frac{1}{T-t}E^*_t\left[\int_t^T\sigma^2_s ds\right]$, where the expectation is taken with respect to the risk neutral distribution. Consider the case where, starting at time 0, the implied variance for the options expiring at time T/2 and time T are available. Define the expected residual function $g\left(\sigma^2_0,\lambda,\theta,k,T\right)=E_0\left[\sigma^2_{T/2,T}\right]-\left(\sigma^2_{0,T}-\frac{1}{2}\sigma^2_{0,T/2}\right)$. Then under the expectations hypothesis, $g\ (\sigma^2_0,0,\theta,k,T)=0$, and under the Heston model, $g\ (\sigma^2_0,\lambda,\theta,k,T)=a+b\sigma^2_0$, where $b<0$.*

Proof. See Appendix.

Hence, under a model like Eq. (3), the expectations hypothesis is misspecified and omits the function $g\ (\sigma^2_0, \lambda, \theta, k, T)$. In practice, this means that expectations hypothesis regressions incorrectly omit the instantaneous variance, which should appear with a negative coefficient.

This upward bias can often be an increasing function of the forecast horizon. Intuitively, note that at the limit as the expiration of the option approaches zero, the expected average variance under the risk neutral distribution is very close to the expected average variance under the actual distribution. As the forecasting horizon lengthens, the difference between the long-run means under the physical and risk neutral distributions asserts itself. Proposition 2 makes this intuition precise; it shows how extending the forecast horizon from Proposition 1 affects the omitted variable in the standard regressions.

Proposition 2. *Define parameters as in* Proposition 1*. Consider the case where, starting at time 0, the implied variance for the options expiring at time T/2 and 3T/2 are available. Define the expected residual function $g'\left(\sigma^2_0,\lambda,\theta,k,T\right)=\frac{1}{3}E_0\left[\sigma^2_{T/2,T}\right]+\frac{1}{3}E_0\left[\sigma^2_{T,3T/2}\right]-\left(\sigma^2_{0,3T/2}-\frac{1}{3}\sigma^2_{0,T/2}\right)$. Then under the expectations hypothesis, $g'\ (\sigma^2_0,0,\theta,k,T)=0$, and under the Heston model, $g'\ (\sigma^2_0,0,\theta,k,T)=a'+b'\sigma^2_0$. Further, $b'<b$ for a range of parameters where λ is relatively large in absolute magnitude compared to k (b is defined in* Proposition 1*).*

Proof. See Appendix.

The probability of a significant crash in the index may also cause the expectations hypothesis to be misspecified. However, a constant jump probability would cause the risk neutral and objective variances to deviate by an amount not dependent on the instantaneous variance. The dependence of the forecasting bias as a function of the instantaneous variance yields testable implications which are considered in the empirical sections. Specifically, the risk premium interpretation requires that the coefficient on the instantaneous variance be negative and, for a given short-dated option, increasing in absolute magnitude as the forecast interval increases.

2. Data

The data are the bid side volatilities for at-the-money calls, collected from traders of over-the-counter stock index options at a major investment bank.[4] The raw data are end of week values and span the period 9 May 1994 to 12 October 2001. Monthly observations are proxied by taking every fourth weekly observation, yielding 97 monthly observations. The data represent the Black–Scholes implied volatility for European options. Unlike volatilities derived from historical data on exchange traded options, the values are the trading desk's indication of the market's Black–Scholes implied volatility at the time of collection. Consequently, there is no direct need to invert the pricing formula using potentially asynchronous index values or mismatched interest rates or dividends to obtain the implied volatility. Harvey and Whaley (1991) discuss the estimation errors induced when the model must be inverted to obtain the Black–Scholes volatility for index options. Informal comparison of the data with interpolated volatilities computed from exchange traded contracts reveals no systematic or peculiar deviations between the two.

Fig. 2 displays the 1 month and 12 month implied volatility for the S&P 500 index. It also graphs the difference between the two series, which measures the slope of the term structure of volatility. The slope was predominately positive during the low volatility 1994–1996 period, and stayed near zero during much of 1997. When volatility rose dramatically during late 1997, the slope initially went negative and then positive before plummeting again during the volatility of Autumn 1998.[5] Following this turbulent period, the slope stayed positive as long-dated volatility remained high during much of 1999 and 2000. By 2001, short-dated volatility had again risen to the level of longer dated volatility. Graphs of implied volatility for the other indexes are very similar.

Table 1 displays summary statistics of the volatility quotes. The values are available for options with a time to expiration of 1, 3, 6, and 12 months; they are quoted in annualized terms (% per year). The underlying indexes are the S&P 500 (US), FTSE 100 (United Kingdom), DAX 30 (Germany), CAC 40 (France), and Nikkei 225 (Japan). For the Nikkei, quotes are not regularly available for 1 month options until October 1996. Hence, summary statistics and tests that require 1 month quotes are excluded for this index.

One of the major features of the data is the decreasing variability of the volatility quotes as the time to expiration increases. Also, each of the series appears markedly skewed to the right, as might be expected. Finally, we observe that the mean and median values of the volatilities tend to

[4] The OTC index option market is large and liquid. According to the Statistical Annex of the December 2001 Bank for International Settlements *Quarterly Review*, the notional value outstanding in June 2001 for OTC options on US equities was $242 billion, while the notional value outstanding of index options on North America exchanges was $1,129 billion.

[5] A referee points out that the term structure slope is usually negative when volatility is high. Even under the volatility risk premium model in Eq. (3), which usually imparts an upward bias to the term structure as a forecast of future volatility, a downward sloping term structure predicts a decline in future volatility. Even to the casual observer without a model, a downward sloping term structure appears to be a strong signal of declining future volatility.

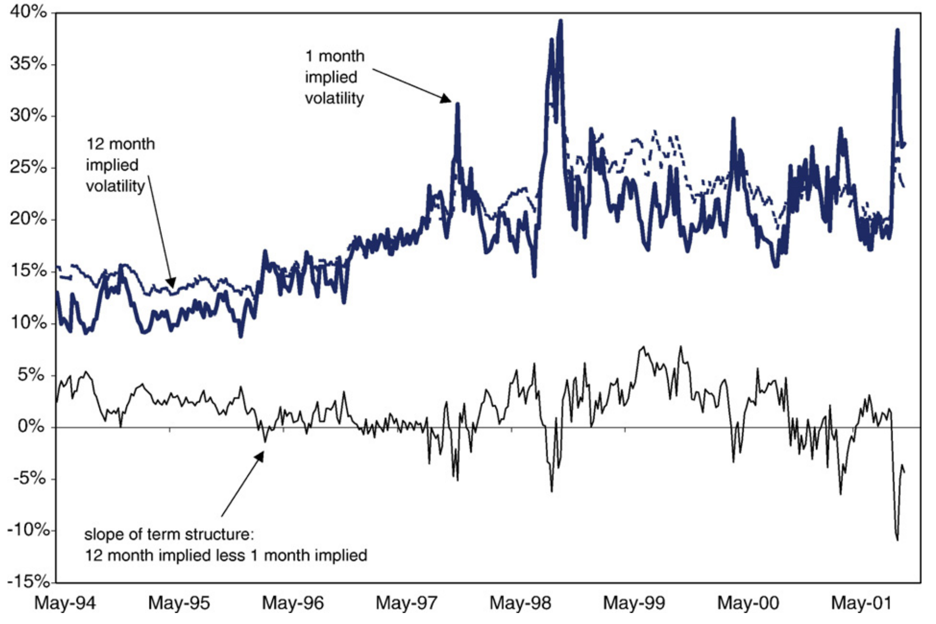

Fig. 2. Time series of S&P 500 implied volatility. Time series of implied volatility from 1 month and 12 month S&P 500 options from May 9, 1994 to October 12, 2001, based on OTC indications. The time series of the difference between the 12 month volatility and the 1 month volatility is also plotted.

describe an upward sloping term structure for each of the series except for the Nikkei. This property of options data appears to be general, as it is documented for currency options in Campa and Chang (1995); theoretical reasons for this characteristic are discussed in Backus et al. (1997). A further explanation may be due to a volatility risk premium, discussed earlier.

3. Term structure regressions

This section describes the results of empirical tests of the expectations hypothesis. The first sub-section describes a test of the traditional expectations hypothesis as modeled in Eq. (2). The conclusion is that the model has some explanatory ability (the slope of the implied volatility term structure has a bit of predictive ability), but overall, the hypothesis fails. The second subsection details Monte Carlo simulations that account for small sample bias that could be influencing the results. The hypothesis is again rejected as an adequate description of the data. The last subsection tests the expectations hypothesis as it appears under a Heston-type volatility risk premium model (Eq. (3)). Parameter estimates conform to the predictions of the model, and the explanatory power of the regressions is dramatically enhanced.

3.1. The expectations hypothesis

The basic hypothesis is tested by regressing the future change in implied variance on the slope of the implied variance term structure and testing whether the constant is zero and, more importantly, whether the slope coefficient equals unity. The test examines whether the spread

Table 1
Descriptive statistics of implied volatility quotes

		Mean	SD	Median	Skewness	Kurtosis	Min	Max
SPX	1 M	18.27	5.61	18.39	0.60	3.79	8.79	39.23
	3 M	18.92	5.36	19.57	0.36	3.15	9.92	39.38
	6 M	19.43	5.13	20.12	0.30	2.84	10.91	38.66
	12 M	19.98	4.83	20.54	0.26	2.33	12.35	36.39
FTSE	1 M	18.66	6.62	18.00	1.26	5.15	9.70	45.00
	3 M	19.67	6.30	18.89	0.99	4.04	11.30	43.47
	6 M	20.16	6.05	19.31	0.94	3.81	12.50	42.70
	12 M	20.67	5.76	20.19	0.86	3.49	13.50	41.10
DAX	1 M	20.80	7.66	19.25	1.51	6.63	9.50	58.00
	3 M	22.25	7.44	21.25	1.08	4.67	11.00	53.50
	6 M	22.70	7.07	22.00	0.95	4.19	11.50	51.50
	12 M	22.89	6.49	23.00	0.86	3.78	13.75	47.50
CAC	1 M	22.07	6.27	21.00	1.43	6.19	11.75	52.00
	3 M	22.70	5.70	22.00	1.25	5.53	13.25	48.50
	6 M	22.98	5.41	22.10	1.06	4.65	14.00	45.00
	12 M	23.20	5.20	22.50	1.07	4.49	15.00	44.00
NKY	1 M	–	–	–	–	–	–	–
	3 M	22.80	5.49	22.50	0.51	3.51	10.40	40.39
	6 M	22.27	4.93	22.13	0.41	3.30	11.00	37.54
	12 M	22.12	4.66	22.15	0.30	2.83	11.80	36.33

Values are quoted in annualized terms (% per year). The data are for options on the S&P 500, FTSE 100, DAX 30, CAC 40, and Nikkei 225 stock indexes. The quotes are for 1, 3, 6, and 12 month at-the-money European options spanning the period May 9, 1994 to October 12, 2001 (389 observations).

adequately predicts changes in short-dated variance. In regression notation, the testable relationship is

$$(1/k)\sum_{i=1}^{k-1}[\sigma^2_{im,(i+1)m}-\sigma^2_{0,m}] = \alpha_0 + \beta_0[\sigma^2_{0,km}-\sigma^2_{0,m}] + \sum_{i=1}^{k-1}\varepsilon_{0im}. \tag{4}$$

The relation is obtained by subtracting σ^2 from both sides of Eq. (2). Regression results for Eq. (4) are displayed in Table 2. The results for each index option are presented for six times-to-expiration pairs. Column headings display the pair of option maturities used for the regressions in that column. For example, the column labeled "3, 1" shows the results for regressions using the difference between 3 month and 1 month variances to predict the subsequent change in the 1 month implied variance, and so on.

The reported standard errors for this table, and each of the other tables, are computed using the method of Newey and West (1987). In order to increase the precision of the estimates, each maturity pair was estimated jointly across markets. Under the expectations hypothesis, $\alpha_0=0$ and $\beta_0=1$, and those are the relevant null hypotheses. Many of the β estimates are significantly greater than zero, but virtually all of them are significantly less than unity. Additionally, the adjusted R^2 values are around 10% or less throughout most of the table, despite the autocorrelation induced by overlapping data.

The regression results imply that long-dated volatility moves too much to be justified by the model's predicted changes in short-dated volatility. For example, suppose 3 month SPX implied volatility is 18% and 6 month implied volatility is 20%. The expectations hypothesis predicts that the 3 month volatility in 3 months would be 21.8%. The regression slope coefficient is

Table 2
Predicting changes in short-dated volatility

		3,1	6,1	12,1	6,3	12,3	12,6
S&P 500	α_0	−0.001	−0.002	−0.001	0	0.001	0.001
	(s.e)	(0.001)	(0.002)	(0.003)	(0.001)	(0.002)	(0.001)
	β_0	0.51*	0.665	0.603	0.515	0.283*	0.107*
	(s.e)	(0.116)	(0.173)	(0.233)	(0.309)	(0.244)	(0.241)
	R^2	5.5	15.8	9.1	6.4	5.0	2.4
FTSE 100	α_0	−0.002	−0.004	−0.004	0	−0.001	−0.001
	(s.e.)	(0.002)	(0.004)	(0.006)	(0.002)	(0.004)	(0.003)
	β_0	0.56*	0.658*	0.628*	0.313*	0.37*	0.325*
	(s.e.)	(0.118)	(0.109)	(0.091)	(0.121)	(0.086)	(0.118)
	R^2	0.0	5.8	7.7	2.9	5.8	4.7
DAX	α_0	−0.001	−0.005	−0.005	0	0	0
	(s.e.)	(0.002)	(0.004)	(0.007)	(0.003)	(0.006)	(0.004)
	β_0	0.279*	0.538*	0.63*	0.372*	0.477*	0.299*
	(s.e.)	(0.119)	(0.094)	(0.1)	(0.223)	(0.156)	(0.2)
	R^2	7.4	16.8	24.9	4.4	11.2	6.1
CAC	α_0	−0.001	−0.003	−0.003	0	−0.001	−0.001
	(s.e.)	(0.002)	(0.004)	(0.006)	(0.002)	(0.005)	(0.005)
	β_0	0.528*	0.611*	0.67*	0.466*	0.555*	0.767
	(s.e.)	(0.085)	(0.049)	(0.052)	(0.14)	(0.05)	(0.192)
	R^2	7.9	18.3	23.4	8.2	12.3	13.9
Nikkei	α_0	–	–	–	0.003	0.006	0.001
	(s.e.)	–	–	–	(0.002)	(0.004)	(0.002)
	β_0	–	–	–	0.699	0.791	0.415
	(s.e.)	–	–	–	(0.175)	(0.149)	(0.318)
	R^2	–	–	–	10.7	10.3	0.7
Joint test	χ^2	239.2	93.6	177.9	72.1	299.3	89.7

* indicates rejection of the null hypothesis $\alpha=0$ or $\beta=1$ at the 5% level.
Regression of the change in the implied variance of short-maturity options on the spread between long- and short-dated implied variance. The regressions for each column are estimated jointly.
Column headings refer to the pair of long- and short-dated option maturities compared in that column. Newey and West (1987) standard errors are reported below parameter estimates. The symbol $\sigma^2_{t,T}$ refers to the implicit option variance quoted at time t for an option expiring at time T. The symbol m refers to the number of months until expiration for short-dated options; km refers to the number of months until expiration of the long-dated option. The row "joint test" provides the chi-square statistic (with 8 or 10 degrees of freedom) testing the joint hypothesis $\alpha=0$ and $\beta=1$ for all regressions in that column.

$$(1/k)\sum_{i=1}^{k-1}\left[\sigma^2_{im,(i+1)m}-\sigma^2_{0,m}\right]=\alpha_0+\beta_0[\sigma^2_{0,km}-\sigma^2_{0,m}]+\sum_{i=1}^{k-1}\varepsilon_{0im}$$

approximately 0.5, and the regression predicts that the future 3 month implied volatility would be 20%: the slope of the term structure exaggerated the expected change in implied volatility. It is tempting to conclude that this regularity might justify an option strategy similar to "riding the yield curve." The conclusion is not warranted.

Examination of the ideal case, where the regression perfectly predicts volatility changes, makes this clear. Suppose dividends and interest rates are constant at zero and the investor knows with certainty that the index will be unchanged at the end of 3 months. He may choose to sell a 3 month ATM call at 18% volatility, or he may sell a 6 month ATM call and buy it back in 3 months. In this case, the 6 month option would have to be priced at a volatility greater than 25.5% in order to break even relative to selling the 3 month call. A term structure slope of this magnitude is typically not observed, so "riding the volatility term structure" in this fashion is not typically justified, even for omniscient investors.

One impediment to interpreting the results is that the estimates for β_0 are biased due to the omission of the θ terms required to derive Eq. (2). Eq. (3) omits the term $((\theta_m/\theta_{km})^2-1)\sigma^2{}_{0,\ km}$ from the right hand side. Treating this as a standard omitted variables problem, the resulting expected value of β_0 is well known:

$$E[\beta_0] = \beta_0 + \left((\theta_m/\theta_{km})^2-1\right)\mathrm{cov}((\sigma^2_{0,km}-\sigma^2_{0,m}), \sigma^2_{0,km})/\mathrm{var}(\sigma^2_{0,km})$$

Because $\theta_{km} < \theta_m$ for $k > 1$, the bias has the same sign as the covariance term in the bias equation. Empirically, this covariance is often negative, which implies that the estimated values of β_0 are lower than the true values. This negative correlation is consistent with mean reversion in the volatility quotes. This serves to mitigate some of the positive results for the expectations hypothesis in Table 2. However, the estimated correlations are positive for some of the option pairs without statistically significant results. For the S&P 500 series, the correlation is positive for the 12 month, 1 month quote pair and for the 6 month, 1 month pair. For the FTSE 100, the 12 month, 1 month option pair exhibits a positive correlation. The approximation implies that the β_0 coefficients are biased upward for these tests. The regression results overstate the true value for these pairs.

3.2. Addressing small sample bias

Bekaert, Hodrick, and Marshall (1997) conclude that another upwards bias is likely to occur in regressions such as Eq. (4). The well-known downward bias in estimating autocorrelations essentially drives this result. They also find that the slope estimates exhibit extreme small sample dispersion. While the cross sectional pooling of the data mitigates these effects, small sample biases may still exist for persistent series. It is necessary to examine the small sample distribution carefully to gauge the reliability of the results. We follow the procedure outlined by the authors in that paper to correct the bias and bootstrap the small sample distribution.

The procedure is in two steps. For each market, first estimate the joint process for the implied variances using a bias corrected VAR(1) model. Then, a large number of experiments are simulated from the VAR, each with 97 observations. By construction, the expectations hypothesis holds in the population moments of the simulated data. However, estimation of Eq. (4) on the artificial data yields parameters with sampling variation in each experiment. These estimated regression coefficients generate the small sample distribution for testing.

Formally, let $y(t) = [\sigma^2_{t,1}, (\sigma^2_{t,3}-\sigma^2_{t,1}), (\sigma^2_{t,6}-\sigma^2_{t,1}), (\sigma^2_{t,12}-\sigma^2_{t,1})]'$ for a given market and assume a VAR(1) describes the data:

$$y(t) = \mu + Ay(t-1) + \varepsilon(t). \tag{5}$$

While a VAR(1) generates acceptable residual diagnostics, the VAR parameters are likely to be biased in small samples. Prior to generating the residuals, the parameters are bias-adjusted using the following procedure. First, estimate the VAR and simulate new data for the short-dated implied variance and variance spreads using the bootstrapped residuals. In each experiment, generate 197 values of each variable, discard the first 100 observations to remove start-up effects, and rerun the VAR on the simulated data. The differences between the original OLS point estimates and the averages of the OLS estimates from 5000 experiments are added to the original point estimates in order to bias-adjust them.

Step two is to construct simulated long-dated variance data from the VAR under the constraints imposed by the expectations hypothesis. For regressions forecasting 1 month implied variance, let

A' denote the bias-adjusted VAR coefficient matrix, and the forecast implied variances that satisfy the hypothesis are constructed with the formula

$$\sigma^2_{t,n} = \frac{1}{n}\sum_{i=0}^{n-1} e1A'y(t), \tag{6}$$

where $e1$ is an indicator vector with a 1 in the first element and zeros elsewhere for n=3, 6, 12.

Finally, estimate regression (Eq. (4)) on the artificial data for each of the 5000 experiments and count how many times the estimated slope falls below the original estimated parameter to compute the empirical p-value. A similar computation applies to regressions forecasting changes in 3 month and 6 month implied variance.

One further note is in order concerning the parameter estimates whose significance is tested in the simulation. For each market, all possible regressions are estimated jointly on the observed data. This is in contrast to the original estimations presented in Table 2, where the data was pooled cross sectionally for a given maturity pair regression. This facilitates testing using the Monte Carlo standard errors, and it provides an idea of the robustness of the estimates. The estimates are similar regardless of the pooling method, and the results in both cases appear far below those implied by the expectations hypothesis.

Table 3 presents the slope coefficients from the joint regressions and the one-sided empirical p-values from the Monte Carlo experiment. For the S&P 500, FTSE 100, DAX, and CAC, the coefficients are significantly below unity at very high confidence levels. In most cases, the p-values are well below 0.01, confirming that the expectations hypothesis does an extremely

Table 3
Results of Monte Carlo experiment

		3,1	6,1	12,1	6,3	12,3	12,6
S&P 500	β_0	0.357	0.590	0.450	0.633	0.290	0.115
	p-value	0.000	0.003	0.000	0.000	0.000	0.000
FTSE 100	β_0	0.371	0.620	0.556	0.282	0.390	0.561
	p-value	0.000	0.005	0.002	0.000	0.000	0.000
DAX	β_0	0.465	0.582	0.672	0.049	0.423	0.281
	p-value	0.004	0.004	0.012	0.000	0.000	0.000
CAC	β_0	0.308	0.607	0.803	0.481	0.722	0.437
	p-value	0.000	0.003	0.040	0.000	0.000	0.000
Nikkei 225	β_0				0.600	0.388	0.034
	p-value				0.025	0.002	0.000

The slope coefficient from the expectations hypothesis regression is presented, along with the empirical one-sided p-values that the coefficient is equal to unity. The p-values are computed as the proportion of times the Monte Carlo regression coefficient was less than the estimated regression coefficient computed using observed data. In the experiment, the assumed data generating process for the implied variance values is a VAR(1) of the shortest dated implied variance and the spreads between that variance and longer dated implied variances, estimated on the observed data. The data for the shortest time to expiration option variance are bootstrapped from the empirical distribution of VAR errors and the longer dated variances are computed from these to be consistent with the expectations hypothesis. Five thousand replications of regressions with sample size of 97 are simulated. Column headings refer to the pair of long- and short-dated option maturities compared in that column. Constant terms are not reported. The rows labeled R^2 give the adjusted R^2 for the regression in that cell. The symbol $\sigma^2_{t,T}$ refers to the implicit option variance quoted at time t for an option expiring at time T. The symbol m refers to the number of months until expiration for short-dated options; km refers to the number of months until expiration of the long-dated option.

$$(1/k)\sum_{i=1}^{k-1}\left[\sigma^2_{im,(i+1)m}-\sigma^2_{0,m}\right] = \alpha_0 + \beta_0[\sigma^2_{0,km}-\sigma^2_{0,m}] + \sum_{i=1}^{k-1}\varepsilon_{0im}$$

poor forecasting job. The results for the Nikkei 225 are not as uniformly damning, but the results are not especially supportive of the expectations hypothesis. The Monte Carlo tests for the Nikkei are also not as powerful as the tests for the other indexes, which include the 1 month option volatilities. Overall, the results that account for the small sample bias and dispersion are in fact a stronger indictment of the hypothesis than asymptotic standard errors imply.

3.3. The expectations hypothesis with a volatility risk premium

As described in the previous section, the expectations hypothesis will not hold if the risk neutral volatility forecast deviates from the objective forecast. Based on the omitted variable arguments in earlier sections, if the expected change in short-dated volatility can be predicted using an asymmetric GARCH(1,1) variance estimate ($\hat{\sigma}_t^2$) in addition to the slope of the term structure, this is evidence to support the existence of a volatility risk premium. The hypothesis is that the current level of volatility should have a negative coefficient and the coefficient should increase in absolute magnitude as the forecasting horizon lengthens. The risk neutral volatility forecast must be adjusted towards the true long-run mean.

The GARCH specification is a standard model accounting for the volatility clustering in financial data as well as the asymmetric effects of positive and negative price changes on volatility. The specification is

$$r_t = a + v_t, v_t \sim t_k(0, \sigma_t^2)$$

$$\sigma_t^2 = b + c\sigma_{t-1}^2 + dr_{t-1}^2 + e\max[0, -r_t]$$

where the notation t_k refers to a t distribution with k degrees of freedom. The specification allows for a larger response of volatility to negative shocks than positive shocks of equal magnitude if $e>0$. The models are estimated separately using daily data for each return series over the period from 2 January 1994 to 12 October 2001. Multivariate GARCH models simultaneously incorporating all of the markets were attempted, but the estimation did not converge due to the extreme complexity of the models.

Because the GARCH volatility is an estimate of the true volatility, the regressions are estimated using instrumental variables. The instrument for the GARCH volatility is the Parkinson (1980) extreme-value estimator

$$\overline{\sigma}_t = \sqrt{\frac{252}{n}\sum_{i=0}^{n-1}\frac{1}{4\ln(2)}\ln(H_{t-i}/L_{t-i})^2} \tag{7}$$

where H_t and L_t denote the high and low futures prices for date t. This estimator is constructed using the futures prices for the preceding 10 trading days. Following Fleming (1998), this variable is an effective instrument because it is highly correlated with the GARCH volatility estimate, but the error in this variable is uncorrelated with the error in the GARCH estimate. The regression results are not particularly sensitive to the specification of volatility proxy: the results are not significantly altered when rolling standard deviations of returns are used as volatility estimates, or if the regressions use instrumental variables, or if the instrument is a rolling standard deviation of returns.

Table 4 displays the results of the augmented test regressions. The regressions are written as

$$(1/k)\sum_{i=1}^{k-1}[\sigma_{im,(i+1)m}^2 - \sigma_{0,m}^2] = \alpha_0 + \beta_0[\sigma_{0,km}^2 - \sigma_{0,m}^2] + \beta_1(\widehat{\sigma}_t^2) + \sum_{i=1}^{k-1}\varepsilon_{0im}. \tag{8}$$

Table 4
Explaining deviations from the expectations hypothesis

		3,1	6,1	12,1	6,3	12,3	12,6
S&P 500	β_0	0.415*	0.551*	0.504*	0.451*	0.224	0.054
	(s.e.)	(0.089)	(1.122)	(0.161)	(0.255)	(0.212)	(0.278)
	β_1	−0.148*	−0.222*	−0.344*	−0.082*	−0.194*	−0.136*
	(s.e.)	(0.071)	(0.084)	(0.095)	(0.041)	(0.06)	(0.038)
	R^2	19.1	37.4	41.1	19.4	31.3	22.6
FTSE 100	β_0	0.582*	0.696*	0.714*	0.361*	0.441*	0.494*
	(s.e.)	(0.106)	(0.093)	(0.093)	(0.129)	(0.086)	(0.176)
	β_1	−0.233*	−0.361*	−0.538*	−0.161*	−0.297*	−0.18
	(s.e.)	(0.093)	(0.129)	(0.192)	(0.061)	(0.119)	(0.101)
	R^2	13.8	29.8	42.1	14.3	29.6	22.7
DAX	β_0	0.293*	0.48*	0.523*	0.249	0.279	0.135
	(s.e.)	(0.121)	(0.101)	(0.128)	(0.222)	(0.192)	(0.287)
	β_1	−0.174*	−0.297*	−0.409*	−0.114*	−0.247*	−0.152*
	(s.e.)	(0.056)	(0.065)	(0.107)	(0.041)	(0.071)	(0.071)
	R^2	23.3	43.2	52.7	16.9	33.3	24.1
CAC	β_0	0.484*	0.542*	0.542*	0.348*	0.375*	0.502*
	(s.e.)	(0.102)	(0.06)	(0.097)	(0.123)	(0.105)	(0.229)
	β_1	−0.206*	−0.339*	−0.535*	−0.16*	−0.33*	−0.352*
	(s.e.)	(0.081)	(0.098)	(0.157)	(0.06)	(0.1)	(0.156)
	R^2	21.3	38.7	54.6	20.0	38.0	38.9
Nikkei	β_0	–	–	–	0.398	−0.101	0.04
	(s.e.)	–	–	–	(0.551)	(0.403)	(0.381)
	β_1	–	–	–	−0.096	−0.359*	−0.158*
	(s.e.)	–	–	–	(0.069)	(0.1)	(0.052)
	R^2	–	–	–	14.7	22.9	15.3
Joint test	χ^2	268.0	218.2	181.3	109.1	570.2	291.9

* indicates rejection of the null hypothesis $\alpha=0$ or $\beta=1$ at the 5% level.
Regression of the change in the implied variance of short-maturity options on the spread between long- and short-dated implied variance and estimated current variance. The regressions for each column are estimated jointly.
Column headings refer to the pair of long- and short-dated option maturities compared in that column. Newey and West (1987) standard errors are reported below parameter estimates. Constant terms are not reported. The rows labeled R^2 give the adjusted R^2 for the regression in that cell. The symbol $\sigma^2_{t,T}$ refers to the implicit option variance quoted at time t for an option expiring at time T. The symbol m refers to the number of months until expiration for short-dated options; km refers to the number of months until expiration of the long-dated option. The symbol $\widehat{\sigma}^2_t$ refers to the fitted asymmetric-Garch (1,1) variance estimated using daily data. Regressions were estimated using the Parkinson estimator as an instrumental variable for the Garch variance. The row "joint test" provides the chi-square statistic (with 12 or 15 degrees of freedom) testing the joint hypothesis $\alpha=\beta_0=\beta_1=0$ for all regressions in that column.

$$(1/k)\sum_{i=1}^{k-1}\left[\sigma^2_{im,(i+1)m}-\sigma^2_{0,m}\right] = \alpha_0 + \beta_0[\sigma^2_{0,km}-\sigma^2_{0,m}] + \beta_1(\widehat{\sigma}^2_t) + \sum_{i=1}^{k-1}\varepsilon_{0im}$$

As with the previous table, each maturity pair is estimated as a system including all available markets. The results are striking. While many of the slope coefficients (β_0's) are significantly greater than zero, all of the risk variables have the expected negative sign, and 25 of the 27 coefficients (β_1's) are statistically significant. The adjusted R^2 for each market/time to expiration pair is also much higher than the corresponding regression in Table 2, as one would expect. While the basic expectations hypothesis appears partially true, there is strong evidence that deviations from the theory are highly correlated with the level of current variance, as predicted by Proposition 1. Another supportive finding is that the coefficient on the spot variance is larger in absolute magnitude as the forecast horizon lengthens, as predicted by Proposition 2. For example, the β_1 coefficient on the spot variance falls from − 0.148 to − 0.222 to − 0.344 as the S&P 500 forecast horizon changes from 3 to 6 to 12 months (using the 1 month option as the short-dated option). Overall, there is strong evidence that the Heston model greatly improves the forecasting performance of the implied volatility term structure.

4. Forward variance regressions

A direct test of the risk premium hypothesis is to add risk premium proxies to forward variance regressions, as in Buser, Karolyi, and Sanders (1996). If a risk premium exists, the forward variance in the term structure should be biased upward as a forecast of future implied variance. The test is more direct than the expectations hypothesis tests above, although it reflects the same information (Table 5).

It is also closely related to a tradable strategy, given that trades can be conducted with forward starting variance swaps, for example. This section contains the results for two forward volatility regression tests, using the 3 month implied variance and the 3 month forward variance.

Table 5
Forward variance unbiasedness regression

	γ_0 (s.e.)	γ_1 (s.e.)	γ_2 (s.e.)	$\bar{R}^2$
S&P 500	−0.000	0.479*		6.96
	(0.002)	(0.269)		
	0.005*	0.384*	−0.183*	22.02
	(0.002)	(0.222)	(0.051)	
FTSE 100	−0.001	0.401*		3.45
	(0.004)	(0.150)		
	0.005	0.428*	−0.203*	10.48
	(0.004)	(0.163)	(0.094)	
DAX	−0.000	0.462*		5.09
	(0.005)	(0.142)		
	0.006	0.378*	−0.138*	13.76
	(0.005)	(0.134)	(0.057)	
CAC	−0.000	0.549*		9.21
	(0.004)	(0.142)		
	0.007	0.496*	−0.181*	15.31
	(0.005)	(0.136)	(0.073)	
Nikkei 225	0.005	0.898		15.47
	(0.003)	(0.165)		
	0.008	0.827	−0.052	17.21
	(0.005)	(0.195)	(0.074)	

The symbol "*" indicates that γ_0 or γ_2 is statistically different from 0 at the 5% level or that γ_1 is statistically significantly less than 1 at the 5% level.
Regression of the three month change in three month implied variance on the spread between the forward implied variance and three month implied variance. The regressions for each index are estimated jointly.
Newey and West (1987) standard errors are reported below parameter estimates. The symbol IV_t refers to the implicit option variance quoted at time t and $\mathrm{IV}_{t,t+3}$ refers to the three month forward variance quoted at time t. The symbol $\widehat{\sigma}_t^2$ refers to the fitted asymmetric-Garch(1,1) variance estimated using daily data. Regressions were estimated using the Parkinson estimator as an instrumental variable for the Garch variance.

$$\mathrm{IV}_{t+3}-\mathrm{IV}_t = \gamma_0 + \gamma_1(\mathrm{IV}_{t,t+3}-\mathrm{IV}_t) + \gamma_2(\widehat{\sigma}_t^2) + \sum_{i=1}^{3} \varepsilon_i$$

The notation for this section is as follows. Let IV_t be the implied variance observed for 3 month options and $IV_{t,\,t+3}$ be the 3 month, 3 month forward implied variance at time t. Due to the linearity of variance with respect to time, the average of the time t 3 month variance and the 3 month forward variance for a 3 month option equals the 6 month implied variance quoted at time t. Accordingly, the forward implied variance values are constructed using the 3 month and 6 month ATM implied variances.

The expectations hypothesis implies that the forward implied variance is an unbiased forecast of the future implied variance, or that $E_t\,[IV_{t+3}]=IV_{t,t+3}$. Under the null hypothesis, $\gamma_0=\gamma_2=0$ and $\gamma_1=1$ in the following regression:

$$IV_{t+3}-IV_t=\gamma_0+\gamma_1(IV_{t,t+3}-IV_t)+\gamma_2(\widehat{\sigma}_t)+\sum_{i=1}^{3}\varepsilon_i. \tag{9}$$

The observed change in implied variance is regressed on the predicted change (forward premium) under the expectations hypothesis. Table 6 contains the system estimation results for regressions, both with and without the additional variance estimate as a regressor. As expected, the results mirror those in previous tables. While the forward premium coefficient γ_1 is significantly different from zero in most instances, it is significantly less than unity in all cases but for the Nikkei. Again, in every case but the Nikkei, the coefficient of the instantaneous variance estimate is significantly less than zero. The joint test against the unbiasedness hypothesis has a chi-squared (10) statistic of 71.05 (p-value<0.001); the joint test against unbiasedness and $\gamma_2=0$ has a chi-squared (15) statistic of 126.04, (p-value<0.001). The table also shows that the R^2 values increase, sometimes dramatically, when the current value for variance is added to the specification.

Table 6
Forward variance forecast error

	γ_0 (s.e.)	γ_1 (s.e.)	R^2
S&P 500	0.008*	−0.212*	19.72
	(0.002)	(0.047)	
FTSE 100	0.008*	−0.278*	10.51
	(0.004)	(0.092)	
DAX	0.010*	−0.178*	13.81
	(0.005)	(0.060)	
CAC	0.013*	−0.288*	13.24
	(0.005)	(0.089)	
Nikkei 225	0.013*	−0.212*	13.36
	(0.005)	(0.070)	

The symbol "*" indicates that γ_0 is statistically significantly different from 0 at the 5% level or that γ_1 is statistically significantly less than 0 at the 5% level.
Regression of the future three month implied variance minus the forward implied variance on a constant and estimated current variance. The regressions for each index are estimated jointly.
Newey and West (1987) standard errors are reported below parameter estimates. The symbol IV_t refers to the implicit option variance quoted at time t and $IV_{t,t+3}$ refers to the 3 month forward variance, quoted at time t. The symbol $\widehat{\sigma}_t^2$ refers to the fitted asymmetric-Garch(1,1) variance estimated using daily data. Regressions were estimated using the Parkinson estimator as an instrumental variable for the Garch variance.

$$IV_{t+3}-IV_{t,t+3}=\gamma_0+\gamma_1(\widehat{\sigma}_t^2)+\sum_{i=1}^{3}\varepsilon_i$$

A related test of the forward unbiasedness regression corroborates the results. This test may have different properties than the above regression, although both are true under the null hypothesis (Table 6).

If the forward rate forecast error ($\mathrm{IV}_{t+3} - \mathrm{IV}_{t,t+3}$) is regressed on a constant and the risk adjustment variable $\hat{\sigma}_t^2$, the results are consistent with the omitted variable argument described above in conjunction with the priced volatility risk story. Under the null hypothesis of unbiasedness, the error should have mean zero and not be correlated with any predictor variables. The results of Table 6 show that in all cases, including the Nikkei, the constant is significantly larger than zero and the coefficient on the variance is significantly less than zero. The joint test that all regressors are zero has a chi-square (10) of 35.51 (p-value$=0.001$).

The results of the forward variance regressions provide further, direct evidence in favor of the volatility risk premium explanation for the evolution of the implied volatility term structure. In fact, the right hand side of the forward variance forecast error regressions represents precisely two times the function $g(\sigma^2{}_0, \lambda, \theta, k, T)$ from Proposition 1. Overall, the empirical results strongly corroborate the theoretical predictions derived from the Heston model.

5. Conclusion

This paper asks whether the term structure of volatility implicit in stock index option prices is consistent with the expectations hypothesis. The distinct contributions of the research are both methodological and empirical. Methodologically, it relies on the Heston (1993) model to generate testable implications, rather than to start with the idea that deviations from the Black–Scholes benchmark must be due to irrationality. Empirically, the data is stronger than the data which has typically been used to examine dynamics of the implied volatility term structure. It represents a cross-section of national indexes rather than the S&P 500, and it uses OTC data rather than exchange-listed data.

I note three testable implications of the Heston (1993) model that relate to deviations from benchmark models. The first implication is that the expectations hypothesis fails. The second implication is that the basic specification wrongly omits the instantaneous variance, which should have a negative coefficient in the augmented expectations hypothesis regressions. The third implication is that the coefficient on instantaneous variance should be increasing in absolute magnitude as the forecast horizon lengthens. I find support for each of these ideas; the results are consistent with a volatility risk premium explanation for option pricing.

The results indicate that the slope of the term structure is a significant predictor of future short-dated implied volatility, although the predictions do not match those of the expectations hypothesis. Correcting the expectations hypothesis forecast with a volatility risk premium term significantly improves the model. Economically, this result says that implied volatilities tend to be "overpriced" as forecasts of future implied volatility levels. The "overpricing" is directly related to the current level of market volatility. The effect is relatively small for nearby maturities, but the effect is larger as the forecast horizon lengthens.

The results provide a justification for selling volatility (through variance swaps or similar strategies) to market participants who believe the risk premium interpretation. The implication for them is that the gap between implied and realized volatility may reflect a risk premium to be captured. This interpretation should be tempered by recognition that part of the risk may be due to uncertainty over the volatility process, rather than priced risk associated with an observable random volatility. Further examination using regime-switching volatility models along the lines of Bekaert, Hodrick, and Marshall (2001) may reduce the evidence for a true risk premium.

Clearly, there are other risk factors besides volatility that index option traders must consider. For example, traders and portfolio managers face liquidity risk, if the market for the option and/or the underlying has time-varying liquidity. Scholes (2000) suggests this strongly increased index option prices during the turmoil of 1998–99.

Traders also face dividend risk for longer dated options. Companies may change their dividend policies, or the constituents of the index may be changed such that the future dividend is unpredictable. Along similar lines, traders may fear that index composition and volatility may change dramatically. The growth of technology oriented stocks in the late 1990s led to more technology-oriented and more volatile, lower dividend indexes than would otherwise have existed.

Index committees may also change the index composition significantly in a single move. For example, in April 2000, 30 companies in the Nikkei 225 were deleted from the index with a week's notice (10 companies had been deleted in the previous 50 years for not meeting the size and liquidity criteria) and replaced with "more representative" companies. Due to the price-weighted construction of the Nikkei, the added companies comprised some 40% of the index. At the time of the change, the trailing 6 month historical volatility for the index based on the old composition was 16% per year, and the trailing historical volatility for the index based on the new composition was 25% per year. Using only the time series evidence for the aggregate index, traders could not have anticipated this discrete regime change in volatility, but option prices may have reflected this possibility.

Evidence for a true risk premium may also be reduced if a substantial component of implied volatility dynamics is due to time-varying crash risk, rather than solely due to a volatility risk premium.[6] The results presented here are consistent with a jump probability that is positively correlated with the level of instantaneous volatility. Other authors have concluded that such a model seems consistent with implied volatility dynamics (e.g., Bates, 2000; Mixon, 2002; Pan, 2002).

Despite these and other risks for traders, the evidence presented here suggests that the compensation for risk in the option market is highly correlated with the level of volatility for the market. Even if other risks are priced, the aggregate risk premium for options is highly correlated with volatility. Theoretical predictions from Heston's volatility risk premium model are verified in the data. Evidence emerging from the expectations hypothesis regressions in this paper provides key stylized facts that tightly parameterized models must confront.

Appendix A

Proof of Proposition 1. First, note the following facts:

$$\sigma_{t,T}^2 = \frac{1}{T-t} E_t^* \left[\int_t^T \sigma_s^2 ds \right] = \sigma_t^2 \left[\frac{1-e^{-(k+\lambda)(T-t)}}{(k+\lambda)(T-t)} \right] + \theta \left[\frac{k}{k+\lambda} \right] \left[1 - \frac{1-e^{-(k+\lambda)(T-t)}}{(k+\lambda)(T-t)} \right] \tag{A1}$$

$$E_0[\sigma_t^2] = \theta(1-e^{-kt}) + \sigma_0^2 e^{-kt} \tag{A2}$$

[6] The results in this paper are inconsistent with a constant crash probability, which would simply raise implied volatility above realized volatility by a constant value.

Eq. (A1) describes implied variance as a function of known variables. It is an expectation under the risk neutral distribution shown in Eq. (3). Eq. (A2) describes the evolution of instantaneous variance under the physical measure.

Then

$$\begin{aligned} g(\sigma_0^2,\lambda,\theta,k,T) &= \frac{1}{2}E_0\left[\sigma^2_{T/2,T}\right]-\left(\sigma^2_{0,T}-\frac{1}{2}\sigma^2_{0,T/2}\right)\\ &= \frac{1}{2}\left[\theta(1-e^{-kT/2})+\sigma_0^2e^{kT/2}\right]\left[\frac{1-e^{-(k+\lambda)T/2}}{(k+\lambda)T/2}\right]\\ &\quad+\frac{1}{2}\theta\left[\frac{k}{k+\lambda}\right]\left[1-\frac{1-e^{-(k+\lambda)T/2}}{(k+\lambda)T/2}\right]-\sigma_0^2\left[\frac{1-e^{-(k+\lambda)T}}{(k+\lambda)T}\right]\\ &\quad-\theta\left[\frac{k}{k+\lambda}\right]\left[1-\frac{1-e^{-(k+\lambda)T}}{(k+\lambda)T}\right]+\frac{1}{2}\sigma_0^2\left[\frac{1-e^{-(k+\lambda)T/2}}{(k+\lambda)T/2}\right]\\ &\quad+\frac{1}{2}\theta\left[\frac{k}{k+\lambda}\right]\left[\frac{1-e^{-(k+\lambda)T/2}}{(k+\lambda)T/2}\right]. \end{aligned} \tag{A3}$$

Direct substitution verifies that $g(\sigma_0^2, 0, \theta, k, T)=0$. Collecting terms on θ and $\sigma^2{}_0$ and noting the expression has the form $g(\cdot)=a+b\sigma^2{}_0$, the expression simplifies to

$$a=\theta\left[\frac{1}{(k+\lambda)T}\right]\left[(1-e^{-kT/2})(1-e^{-(k+\lambda)T/2})-\left[\frac{k}{k+\lambda}\right](1-e^{-(k+\lambda)T})^2\right] \tag{A4}$$

and

$$b=\left[\frac{1}{(k+\lambda)T}\right](e^{-(k+\lambda)T}-e^{-(k+\lambda)T/2})(1-e^{\lambda T/2}). \tag{A5}$$

First, consider the case where $k+\lambda>0$. The first term in Eq. (A5) is positive, the second term is negative, and the third term is positive, therefore $b<0$. Next, consider the case where $k+\lambda<0$. Then the first term is negative, the second term is positive, and the third term is positive, therefore $b<0$.

Proof of Proposition 2. Using the same process as used in the proof of Proposition 1, I can write $g'(\sigma^2{}_0, \lambda, \theta, k,T)$ as a function of structural parameters. I then derive conditions under which $b'<b$, where b is derived in Proposition 1.

$$g'(\sigma_0^2,\lambda,\theta,k,T)=\frac{1}{3}E_0[\sigma^2_{T/2,T}]+\frac{1}{3}E_0[\sigma^2_{T,3T/2}]-\left(\sigma^2_{0,3T/2}-\frac{1}{3}\sigma^2_{0,T/2}\right)=a'+b'\sigma_0^2, \tag{A6}$$

where $b'=\left(\left[\frac{1-e^{-(k+\lambda)T/2}}{(k+\lambda)T/2}\right]\left[\frac{1}{3}e^{-KT/2}+\frac{1}{3}e^{-KT}+\frac{1}{3}\right]-\left[\frac{1-e^{-(k+\lambda)3T/2}}{(k+\lambda)3T/2}\right]\right)$ and a' is defined analogously to the proof of Proposition 1. I need to show the conditions under which $b'<b$. This leads to the expression

$$2\left[\frac{1}{(k+\lambda)T}\right][e^{-KT}-e^{-(K+\lambda)T}][1-e^{-(K+\lambda)T/2}]<\left[\frac{1}{(k+\lambda)T}\right](e^{-(k+\lambda)T}-e^{-(k+\lambda)T/2})(1-e^{\lambda T/2}). \tag{A7}$$

This condition can be reduced to

$$kT<2\ln(2)+2\ln(e^{-\lambda T}-1)-2\ln(e^{-\lambda T/2}-1), \tag{10}$$

which can be interpreted as saying that the condition obtains as long as λ is relatively large compared to k. For example, suppose T is 13 weeks and k is 0.12 (the value from the example shown in Fig. 1), then the condition holds for all $\lambda<0$. For $T=26$ weeks and $k=0.12$, the condition is true for $\lambda<\lambda^*$, where the critical value $\lambda^*\approx-0.025$.

References

Backus, David, Foresi, Silverio, Li, Kai, Wu, Liuren, 1997. Accounting for biases in Black–Scholes. Manuscript. New York University.

Bakshi, Gurdip, Kapadia, Nikunj, 2002. Delta hedged gains and the negative market volatility risk premium. Review of Financial Studies 16 (2), 527–566.

Bank for International Settlements, 2001. Quarterly Review (December).

Bates, David, 2000. Post-'87 crash fears in the S&P 500 futures option market. Journal of Econometrics 94, 181–238.

Bekaert, Geert, Hodrick, Robert, Marshall, David, 1997. On biases in tests of the expectations hypothesis of the term structure of interest rates. Journal of Financial Economics 44, 309–348.

Bekaert, Geert, Hodrick, Robert, Marshall, David, 2001. Peso problem explanations for term structure anomalies. Journal of Monetary Economics 48, 241–270.

Black, Fischer, 1989. How to use the holes in Black–Scholes. Journal of Applied Corporate Finance 1 (4), 67–73.

Buraschi, Andrea, Jackwerth, Jens, 2001. The price of a smile: hedging and spanning in option markets. Review of Financial Studies 14 (2), 495–527.

Buser, Stephen, Karolyi, Andrew, Sanders, Anthony, 1996. Adjusted forward rates as predictors of future spot rates. Journal of Fixed Income 6, 29–42 (December).

Byoun, Soku, Kwok, Chuck C.Y., Park, Hun Y., 2003. Expectations hypothesis of the term structure of implied volatility: evidence from foreign currency and stock index options. Journal of Financial Econometrics 1, 126–151.

Campa, Jose Manuel, Chang, P.H. Kevin, 1995. Testing the expectations hypothesis on the term structure of volatilities in foreign exchange options. Journal of Finance 50, 529–547.

Campbell, John Y., Shiller, Robert, 1991. Yield spreads and interest rate movements: a birds eye view. Review of Economic Studies 58, 495–514.

Coval, Joshua, Shumway, Tyler, 2001. Expected option returns. Journal of Finance 56 (3), 983–1009.

Dai, Qiang, Singleton, Kenneth, 2002. Expectation puzzles, time-varying risk premia, and affine models of the term structure. Journal of Financial Economics 63 (3), 415–441.

Diz, Fernando, Finucane, Thomas, 1993. Do the options markets really overreact? Journal of Futures Markets 13 (3), 299–312.

Fleming, Jeff, 1998. The quality of market volatility forecasts implied by S&P 100 index option prices. Journal of Empirical Finance 5, 317–345.

Fleming, Jeff, 1999. The economic significance of the forecast bias of S&P 100 index option implied volatility. Advances in Futures and Options Research 10, 219–251.

Harvey, Campbell, Whaley, Robert, 1991. S & P 100 option volatility. Journal of Finance 46, 1551–1561.

Heston, Steve, 1993. A closed form solution for options with stochastic volatility with applications to bond and currency options. The Review of Financial Studies 6, 327–343.

Heynen, Ronald, Kemna, Angelien, Vorst, Ton, 1994. Analysis of the term structure of implied volatilities. Journal of Financial and Quantitative Analysis 29, 31–56.

Hull, John, White, Alan, 1987. The pricing of options on assets with stochastic volatilities. Journal of Finance 42, 281–300.

Lamoureaux, Christopher, Lastrapes, William, 1993. Forecasting stock–return variance: toward an understanding of stochastic implied volatilities. Review of Financial Studies 6, 293–326.

Lo, Andrew, Wang, Jiang, 1995. Implementing option pricing models when asset returns are predictable. Journal of Finance 50 (1), 87–130.

Mixon, Scott, 2002. Factors explaining movements in the implied volatility surface. Journal of Futures Markets 22 (10), 915–937.

Newey, Whitney, West, Kenneth, 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.

Pan, Jun, 2002. The jump risk premia implicit in options: evidence from an integrated time-series study. Journal of Financial Economics 63 (1), 3–50.

Parkinson, Michael, 1980. The extreme value method for estimating the variance of the rate of return. Journal of Business 53, 61–66.

Poterba, James, Summers, Lawrence, 1986. The persistence of volatility and stock market fluctuations. American Economic Review 76, 1142–1151.

Poteshman, Allen, 2001. Underreaction, overreaction, and increasing misreaction to information in the options market. Journal of Finance 56, 851–876.

Scholes, Myron, 2000. Crisis and risk management. American Economic Review 90 (2), 17–21.

Stein, Jeremy, 1989. Do the options markets overreact? Journal of Finance 44, 1011–1023.