# FACTORS EXPLAINING MOVEMENTS IN THE IMPLIED VOLATILITY SURFACE

**SCOTT MIXON**

This article explores the relationship of changes in the S&P 500 index implied volatility surface to economic state variables. Observable variables can explain some of the variation in implied volatility, with the majority of explanatory power from index returns. Although the contemporaneous return is most important for explaining changes in short dated volatility, the path of the index is important for explaining changes in long dated volatility. Other variables also display statistically significant relations to volatility changes. Shocks to the Nikkei 225, short-term interest rates, and the corporate/government bond yield spread are correlated with small, systematic changes in implied volatility. The results suggest a multifactor model for market volatility, with factors other than index returns adding negligible explanatory ability. © 2002 Wiley Periodicals, Inc. Jrl Fut Mark 22:915–937, 2002

I would like to thank Stephen Figlewski and an anonymous referee, as well as participants at the Florida State University financial mathematics seminar and the 2000 Southern Finance Association meeting, for helpful comments. This article reflects the opinions of the author and does not necessarily reflect the views of UBS Warburg LLC.
For correspondence, Scott Mixon, UBS Warburg LLC, 677 Washington Boulevard, Stamford, Connecticut 06901; e-mail: scott.mixon@ubsw.com

*Received November 2001; Accepted February 2002*

■ *Scott Mixon is in the Equity Derivatives Research area at UBS Warburg LLC in Stamford, Connecticut.*

The Journal of Futures Markets, Vol. 22, No. 10, 915–937 (2002)

Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/fut.10039

## INTRODUCTION

Although the benchmark Black-Scholes option pricing formula assumes a constant volatility for the underlying asset, this assumption is almost always violated in reality. Quoting option prices with a misspecified model gives rise to the calculation of an implied volatility matrix, or surface; this matrix contains Black-Scholes implied volatilities for various strike prices and maturities at a point in time. The structure of this matrix can be interpreted as particular distributional deviations from the log-normality of prices assumed by Black and Scholes. Moreover, the so-called term structure and strike structure of volatility changes through time as economic conditions change.

To guide financial theory, this article characterizes the dynamic structure of that surface through time, and relates the changes in implied volatility to several economic state variables. This is relevant because tightly parameterized option pricing models have difficulty accommodating the structure in the data (e.g., Bates, 1996, 2000; Bakshi, Cao, & Chen 1997). It has also been found that the Black-Scholes pricing biases change through time: the implied volatility term structure changes slope and sign of slope, and the volatility smile changes slope (see, e.g., Bates, 2000; Rubenstein, 1985). This presumably reflects the fact that the option pricing model is reduced form or partial equilibrium in nature, and cannot directly accommodate larger macroeconomic shocks. A particular specification of the stock price process is assumed, with all other events in the economy relegated to an error term. In reviewing the changing nature of the deviations from the Black-Scholes option pricing model, Hull (1997, p. 509) concludes that "[p]ossibly macroeconomic variables affect stock option prices in a way that is as yet not fully understood." One view is that "[t]he ultimate research agenda may therefore be to identify those omitted 'fundamentals' that are showing up as parameter shifts in current option pricing models" (Bates, 1996, p. 102).

This article explores the movements in the implied volatility surface without imposing a tightly parameterized model. In the first section, systematic movements in the surface are described by examining the principal components of first-differenced volatility from S&P 500 OTC options with a wide range of strikes and maturities of up to 5 years. In the second section, changes in at-the-money implied volatility for the entire at-the-money term structure are investigated directly. These changes are regressed on innovations in a number of macroeconomic state variables to relate implied volatility to the larger economy.

This article is closest in spirit to Schwert (1989) and Franks and Schwartz (1991), who relate U.S. stock market volatility and changes in nearby implied volatility for FTSE 100 options, respectively, to several state variables. It is also related to other papers examining the principal components of implied volatility data. Zhu and Avallaneda (1997) analyze at-the-money (ATM) FX options and characterize the time series properties of the principal components. Skiadopoulos, Hodges, and Clewlow (1999) investigate exchange traded S&P 500 options, relating the unobserved factors to index returns. A major difference in focus differentiates this research from previous work. The focus here is on placing implied volatility changes into a macroeconomic context, whereas previous research in this field has often been motivated by the need to find plausible models for pricing and hedging options.

The conclusions are as follows. First, three unobserved components can explain, on average, approximately 90% of the variation in implied volatilities. The first factor, which moves the entire term structure of volatility in the same direction, explains 80–90% of the variation in option volatilities with maturities of 1 year or less (which are heavily traded on exchanges). It is strongly related to contemporaneous market returns, and exhibits negative autocorrelation. It affects the skew of near term out of the money puts and calls, in a way that is consistent with a random intensity jump model featuring the correlation of stochastic volatility and jumps. The second factor is somewhat associated with market returns. Its explanatory ability is concentrated on options with greater than 1 year to maturity (which are mostly traded OTC); an interpretation is that it reflects longer term macroeconomic expectations and risks. Interestingly, the path of the index appears highly important in determining the level of the long-dated volatility. The third factor is interpreted as a pure jump fear component: it explains a small proportion ($\sim$5%) of the variance, and reflects changes in the slope of the implied volatility smile at the 1 month horizon. It is not strongly related to the postulated economic variables.

Because of the close relationship between volatility expectations and ATM volatility, the relationship between observable variables and the ATM term structure of implied volatility is also investigated directly. Contemporaneous index returns significantly affect the term structure, but with a monotonically declining effect as time to maturity increases. Lagged index returns do not significantly affect volatility at the short end of the term structure, but they are highly significant for determining volatility further out along the curve, indicating path dependence on the relation between implied volatility and returns. Foreign returns shift the term structure in a parallel fashion. Changes in the short rate affect only the nearest maturity options, but the spread between corporate bond yields and government bonds is associated with volatility changes across the entire term structure.

A few unobserved components suffice to explain the variation in implied volatility, but observed variables account for a much smaller proportion of the variation. The explanatory ability of observed variables declines as option maturity increases. Changes in observable variables explain movements in near-dated option volatility, but unobservable variables explain movements in longer dated volatility.

This article represents an improvement over the previous literature in several ways. First, the dataset represents an improvement over previously utilized data. By using OTC option marks with a fixed time to maturity, rather than exchange traded options with a telescoping time to maturity, as in Skiadopoulos, Hodges, and Clewlow (1999), the noise in the data is reduced. This richer and much larger data set allows for much sharper conclusions about the data generation process than previously reported. This research also relates movements in the entire implied volatility surface to various macroeconomic state variables, which has not been done before. This is the most important contribution: relating the stylized facts about the implied volatility surface to the state of the economy. The use of principal components is primarily as a convenient data reduction technique rather than as the focus of the article.

The article is structured as follows. The first section describes the data set used in the empirical work. The next section reviews the estimation procedures. The third section presents the estimation results. It includes evidence on the amount of variation explained by each factor for various options, and the sensitivity of volatility at various strikes and maturities to the factors. It also includes the results of regressions that relate movements in the volatility surface to economic state variables. The final section provides concluding thoughts.

## DATA

Two sets of option data are examined in this study. The first set of option data consists of a cross section of S&P 500 OTC option implied volatilities for the last trading day each week from November 28, 1997 to June 22, 2001, for a total of 186 observations after taking log first differences. At each point in time, the trading desk of a major international broker-dealer provided Black-Scholes volatilities for the following expirations: 1, 3, 6, 12, 24, 36, 48, and 60 months. For each of these maturities, the implied volatility is observed for each of the following five options: out of the money puts with strike prices of 90 and 95% of spot, at-the-money calls, and out of the money calls with strike prices of 105 and 110% of spot. Hence, at each observation date, a $40 \times 1$ vector of S&P 500 implied volatilities is observed.[1] The second set of option data consists of a longer time series of ATM implied volatilities for options expiring in 1, 3, 6, 12, 24, 36, 48, and 60 months. This set of data begins May 9, 1994, and runs until June 22, 2001, for a total of 372 weekly observations before differencing.

Given the origin of the data, it is possible that firm-specific effects influence the analysis. However, there are reasons to believe these effects are not strongly affecting the results. First, the focus is on systematic changes in the volatility surface, rather than outliers or specific events. Second, the data covers options that are relatively near the money and are less likely to exhibit large differences from typical market values than, say, far out of the money puts might.

The data, especially the longer ATM series, covers several different market environments. It includes the quiet, low volatility period of 1994–1995, as well as the period of upward trending volatility of 1996 and early 1997. The violent market movements of late 1997 and 1998, with the associated swings in volatility are also included. Finally, the steadier but high levels of volatility of 1999–mid-2001 are also included in the analysis. Overall, the data represents a relatively long period in the option market (over 7 years) under a variety of economic conditions.

It is worth considering the nature of the OTC equity option market to provide context for the data. In the United States, the market is active and economically large, although it is not as large as the exchange traded option market. (Outside the United States, the OTC market is larger than the OTC market.) According to Bank for International Settlements statistics (Bank for International Settlements (BIS), 2001), the notional amount outstanding in June 2001 for U.S. equity options was $242 billion. At the same time, the notional amount outstanding for exchange traded index options was $1.13 trillion.

There is also a substantial market for OTC equity options that have longer maturities than the bulk of listed options have. Although statistics for the United States are not broken out, the BIS estimated that the global outstanding notional value for OTC equity options was $1.56 trillion. Of this, approximately 29% of the notional was for options with less than 1 year until expiration, 50% was for options with

[1]The data analyzed is the raw data (bid-side volatilities, for fixed percentage moneyness) provided by the trading desk. The moneyness of the option volatilities is not adjusted for maturity, as suggested by Dumas, Fleming, and Whaley (1998). Hence, the analysis is directly comparable with other volatility term structure research, for example, Das and Sundaram (1999), and with market practice.

1–5 years until expiration, and 21% was for options with greater than 5 years until expiration. Estimates for previous years show a similar tendency for the notional value of long-dated options to be of comparable size to the notional value for options with less than 1 year until expiration.

## ESTIMATION METHOD

Suppose the logarithmic changes in implied volatility at a given date can be approximated as coming from a linear model of the form

$$d\sigma = \mu + LF + \varepsilon \tag{1}$$

where $E[F] = E[\varepsilon] = 0$, $E[\varepsilon F'] = 0$, $E[FF'] = I$, and $E[\varepsilon\varepsilon'] = \Psi$, where $\Psi$ is a diagonal matrix. The vector $d\sigma$ is a $p \times 1$ vector consisting of logarithmic first differences in implied volatility for the maturities and five values of moneyness described above.[2] The parameter $p = 40$. There are $m$ common factors that are associated with changes in implied volatility, which are collected in the $m \times 1$ vector $F$. The associated $p \times m$ factor loading matrix $L$ represents the sensitivity of the volatility to each of the factors. The following covariance structure for the volatility changes is implied: $\Sigma = LL' + \Psi$.

It is well known that the estimated factor loadings are not unique; they are determined only up to an orthogonal rotation, and the solutions cannot be distinguished on the basis of the observed data. Any orthogonal rotation is valid, and it corresponds to changing the directions of the basis vectors for the space spanned by the factors. Changing the directions of the vectors alters the interpretation of the factors, but it does not alter the space spanned by the factors. In this article, the unrotated factors are examined, and this arbitrariness should be kept in mind.

The estimation of the factor loadings is determined by the principal components method. If the estimated factor loadings, $\hat{L}$, and specific variances, $\hat{\Psi}$, are treated as if they are the true values, and the specific factors, $\hat{\varepsilon}$, are treated as error vectors, the factor realizations, or scores, can be recovered. Following Bartlett (1937) and Knez, Litterman, and Scheinkman (1994), weighted least squares is used for this purpose. This method accounts for the fact that the diagonal elements of $\Psi$ are

[2]Estimation using simple differences provided qualitatively similar results as estimation using log first differences, but the use of log changes is more consistent with the positively skewed nature of implied volatility data. Also, when the diagnostics for structural breaks in the Estimation Results Section are performed in simple difference form, they provide strong evidence that the simple difference specification is not stable.

not likely to be equal. Specifically, for each date $t$, the sum of squared errors, weighted by the reciprocal of estimated variances, is

$$\sum_{i=1}^{p} \frac{\varepsilon_i}{\psi_i} = \varepsilon' \Psi^{-1} \varepsilon = (d\sigma - \mu - LF)' \, \Psi^{-1}(d\sigma - \mu - LF) \qquad (2)$$

The objective is to minimize equation 2 with respect to $F$ at each date. Replacing the population values of the relevant variables with their sample analogs, the solution is to choose $\hat{F}_t = (\hat{L}'\hat{\Psi}^{-1}\hat{L})^{-1}\hat{L}'\hat{\Psi}^{-1}(d\sigma_t - \hat{\mu})$. The estimated factor scores have virtually zero correlation with each other.

## ESTIMATION RESULTS

Table I presents the variance decomposition results for the three-factor model of S&P 500 volatility dynamics. As described above, the estimation results are from the principal components solution method. Maximum likelihood estimation of the factor model was also performed, with similar results for explanatory ability. However, the results are not as readily interpretable.[3] To avoid this problem, the principal components solution is preferred. Likelihood ratio tests from maximum likelihood

**TABLE I**

Proportion of Variation in S&P 500 Implied Volatility Explained by Three Factors

| *Time to Expiration (Months)* | *Total Variance Explained (%)* | *Proportion of Total Variance Explained by* | | |
|---|---|---|---|---|
| | | *Factor 1* | *Factor 2* | *Factor 3* |
| 1 | 97.8 | 84.4 | 7.0 | 6.4 |
| 3 | 94.8 | 93.1 | 1.4 | 0.2 |
| 6 | 96.3 | 90.6 | 5.6 | 0.1 |
| 12 | 95.1 | 81.1 | 14.0 | 0.0 |
| 24 | 93.0 | 68.3 | 24.1 | 0.6 |
| 36 | 90.4 | 50.9 | 37.8 | 1.7 |
| 48 | 84.6 | 43.6 | 40.0 | 1.0 |
| 60 | 76.7 | 34.8 | 40.4 | 1.5 |

Note. The table reports the proportion of variance in weekly log first differences in S&P 500 implied volatility explained by the first three principal components. For each option maturity, the explained variance is the average across options with strike prices equal to 90, 95, 100, 105, and 110% of spot. The data are 186 weekly observations of changes in implied volatility from December 5, 1997 to June 22, 2001.

[3]The identifying assumption that $L'\Psi^{-1}L$ is diagonal is convenient for maximum likelihood estimation, but it may lead to factors that are not easily interpretable. See the discussion in Johnson and Wichern (1982, p. 405).

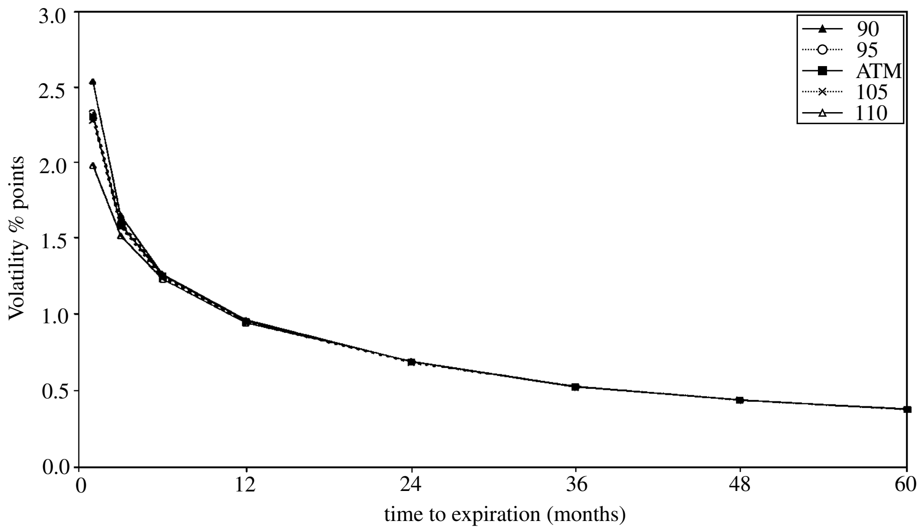
estimation indicate that models with up to 15 factors are statistically rejected. Nonetheless, the extra factors provide little explanatory power, and the three-factor model is adequate to explain the majority of systematic movements.

The first principal component explains most of the variance in options with maturities of up to 12 months. The explanatory power of this factor diminishes with option maturity, falling from over 80% at the 1 year or less maturity to approximately 30% for 3 to 5 year maturities. The variation in volatility for longer maturity options is explained mostly by a second factor. The explanatory ability of the second factor rises with the option maturity. The second factor accounts for very little of the variation in the 1 year or less horizon, but 30–50% of the 3 to 5 year maturity variation. Exchange traded options are very liquid for contracts with less than 1 year to maturity and longer term options traded by institutional investors are OTC, and this difference in clientele may help explain the nature of the results. In particular, the longer dated OTC market is less liquid and more opaque than the market for short-term options. This difference in risks no doubt influences the findings. Finally, the third factor explains a much smaller proportion of the total movement (approximately 5%), and this explanatory power is focused at the 1 month horizon.

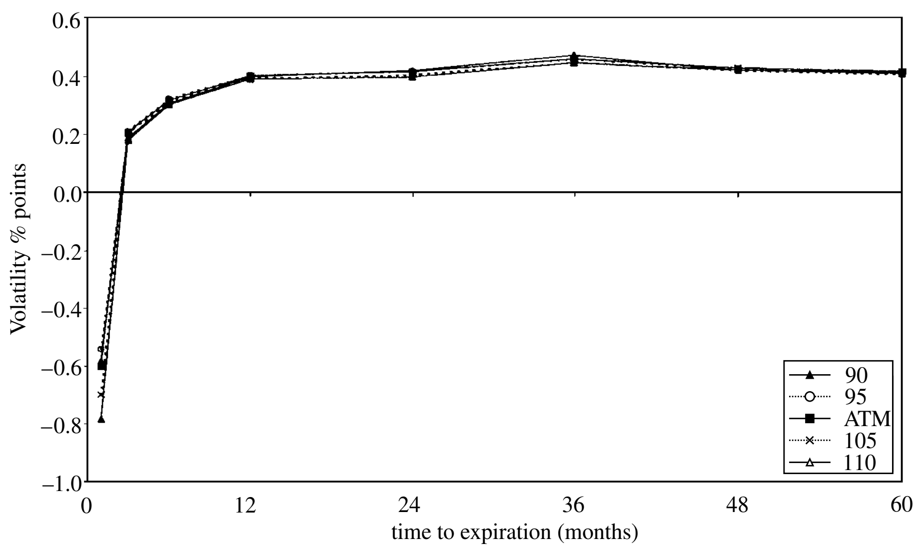
## Factor Loadings

Panels A, B, and C of Figures 1 display the estimated factor loadings for the S&P. For ease of interpretation, the loadings are transformed to represent changes in annualized volatility percentage points. The raw loadings (which measure percentage changes) are used to construct changes in the median term structure over the sample period. The character of the three factors is readily apparent. Factor 1 represents a shock that affects all maturity and moneyness options in the same direction. Because the option volatility can be thought of as approximately the expected average risk neutral volatility over the life of the option, the interpretation of this factor is entirely consistent with standard mean-reverting stochastic volatility models. A shock to volatility raises the instantaneous volatility, but the effect is expected to dampen over time. This leads to a shock that increases the level of implied volatility at all horizons, with the strongest increase at the shortest horizons.

Factor 1 also affects the out-of-the-money option volatility differently than at the money volatility. An increase in volatility is generally associated with an increase in the steepness of the implied volatility smile (and a corresponding increase in the skewness of the implied risk neutral density). Panel A of Figure 1 shows, for example, that a one standard deviation shock to the first S&P 500 factor drives the 110% call volatility up by 2 percentage points, while the 90% put volatility increases by nearly 3 percentage points. This does not appear to be an artifact of the data or



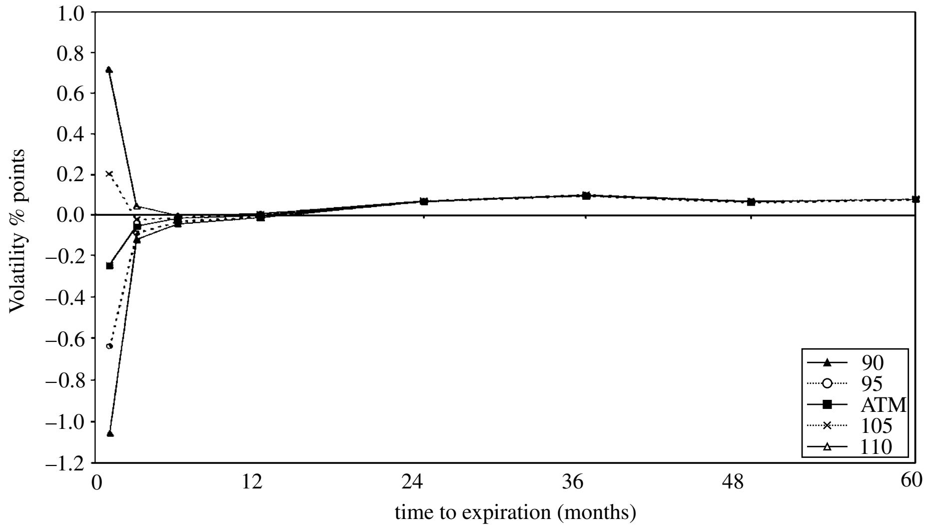

Panel A: Factor 1




Panel B: Factor 2

**FIGURE 1**

Loadings for the three-factor model of S&P 500 implied volatility. Panels A, B, and C show the factor loadings in the log change in implied volatility for the three factors. The graph of each factor for each option moneyness represents the change in the median volatility term structure that would be caused by a one standard deviation shock to that factor. The units of the y-axis are in volatility percentage points, while the x-axis represents months to expiration for the options. The data are scaled to represent changes in the median observed volatility surface. (Continued)




Panel C: Factor 3

**FIGURE 1**
(Continued)

statistical technique, as it is corroborated by correlations computed from exchange traded option data. The correlation of changes in 1 month at-the-money volatility and changes in the skew (measured by the 90% put volatility less the 110% call volatility) for the S&P is 0.56.

Factor 2 represents a "spread" factor that acts to separate the short and long end of the volatility term structure. A shock to this factor lowers the 1 month option volatility but increases longer dated volatility. It uniformly affects options of all moneynesses. The finding of a significant second factor is consistent with a volatility process, which has short-lived shocks and a more persistent component. Table I indicates that factor 2 explains much of the movement in longer dated volatility, and very little of the short-dated volatility moves, which indicates that factor 2 may be proxying for a low-frequency component of stock market volatility. Such a two-factor model has been estimated in, for example, Engle and Lee (1993) and Gallant, Hsu, and Tauchen (1999) for spot markets and Bates (2000) for the option market. This second important factor is also consistent with the conclusion of Bakshi, Cao, and Chen (2000) that long-dated options have information not readily available from short dated options.

Factor 3 describes a factor that explains a relatively small fraction of the overall surface movement. Panel C of Figure 1 displays the factor loadings. The loadings imply a factor with a strong effect at the 1 month horizon. It strongly affects out-of-the-money puts and calls differently, acting to sharply increase the steepness of the implied volatility smile. A positive shock to this factor seems related to an increased jump fear component in the return process. The effects are largely for the 1 month volatility.

Skiadopoulos, Hodges, and Clewlow (1999) examine the correlation matrix of the S&P 500 volatility surface using exchange traded options from 1992–1995 and find two interpretable factors. They find that the first factor is a shift factor, but the second factor has a "z-shape." The second factor acts to change the slope of the implied volatility smile. Their analysis is hampered by the nature of their data, which buckets options into groups having 10–90 days to maturity, 90–180 days to maturity, and 180–270 days to maturity. The data available for this research allows more precise statements regarding the nature of the latent factors. In particular, the availability of long-dated OTC options provides information on the second factor that is not available from options with a year or less to expiration.

## The Relation to State Variables

This section relates movements in implied volatility to observable variables. In the first subsection, the analysis is carried out using the factor scores from above. In a second subsection, a much longer time series of at-the-money volatility is used to explore the relations. The variables chosen are the contemporaneous and lagged log return to the S&P 500, the contemporaneous return to the Nikkei 225, the 3-month constant maturity Treasury bill rate, slope of the yield curve, proxied by the 10-year constant maturity Treasury yield less the 3-month rate, and the spread of Moody's AAA index yield over the 30-year constant maturity Treasury bond yield. Each of the interest rate variables are first differenced to represent innovations. Similar variables have been used by Chen, Roll, and Ross (1986), Schwert (1989), Franks and Schwartz (1991), and many others. Most of the variables reflect the state of the U.S. economy, while the Nikkei return is best interpreted as a proxy for conditions in the rest of the world.

By making the assumption that changes in implied volatility are a result of changes in state variables plus noise terms (i.e., no feedback from volatility to the variables), sharp statements based on linear regressions are available. Table II reports the results from regressing the factor scores for each of the three factors for the S&P on the observable variables. Some of the coefficients are interpretable and consistent with intuition. The regression of factor 1 produces a significant negative coefficient on the domestic index return. A negative (positive) return increases (decreases) volatility for all maturities. The corporate bond spread also has a significant positive coefficient in the S&P regression, although the

**TABLE II**
Regression of Factor Scores on State Variables

| | *F1* | *F2* | *F3* |
|---|---|---|---|
| Constant | −0.003 | −0.010 | −0.003 |
| | (−0.07) | (−0.18) | (−0.07) |
| S&P 500 | −0.221 | −0.010 | 0.038 |
| | (−10.99) | (−0.34) | (1.34) |
| S&P 500(−1) | 0.020 | −0.100 | −0.025 |
| | (1.06) | (−3.74) | (−0.95) |
| Nikkei | −0.027 | −0.031 | 0.010 |
| | (−1.55) | (−1.26) | (0.43) |
| T-bill | −1.036 | −0.000 | 0.672 |
| | (−2.29) | (−0.00) | (1.13) |
| Slope | −0.291 | 0.774 | 0.247 |
| | (−0.72) | (1.35) | (0.47) |
| AAA-30Y | 3.684 | 2.192 | 2.654 |
| | (3.13) | (1.31) | (1.84) |
| MA(1) | −0.182 | −0.167 | −0.424 |
| | (−2.41) | (−2.19) | (−6.05) |
| $\overline{R}^2$ | 54.4% | 9.3% | 12.5% |

Note. Factor scores are extracted from weekly first log differences in index option volatility for S&P 500 options with 1, 3, 6, 12, 24, 36, 48, and 60 months to expiration. Independent variables are listed in the first column. Estimated coefficients and *t*-statistics (in parentheses) are reported where applicable. *S&P 500* returns are contemporaneous log price changes in the S&P 500 multiplied by 100; *S&P 500(−1)* returns are the log price changes in the index lagged 1 week. *Nikkei 225* returns are contemporaneous log price changes in the Nikkei 225 multiplied by 100. The variable *T-bill* is the change in the 3-month constant maturity (cm) T-bill yield. *Slope* is the change in the 10 year cm Treasury bond yield less the 3-month cm T-bill yield. *AAA-30Y* is the change in Moody's AAA index yield less the 30-year cm Treasury yield. *MA(1)* refers to the moving average correction of order 1.

significant coefficient on short rates is negative. Conditioned on the index return, a lower interest rate increases volatility. This mirrors the finding of Franks and Schwartz (1991) and is consistent with a leverage effect on volatility. For a given index value, a lower interest rate increases the value of corporate debt, and hence, the debt/equity ratio and the riskiness of the cash flows. Interpretation of the factor 2 regression is more difficult. The most interesting fact is that the contemporaneous index return is not significant, but the lagged index return is highly significant. The factor 3 regression provides little help in interpreting the factor, other than a strong tendency towards mean reversion.

The results are somewhat mixed for interpreting the regressions. Some coefficients have the intuitively expected sign. Given the nature of principal component analysis, which combines the information in many variables into a few factors, this should not be surprising. An interesting finding concerns the $R^2$ values for the regressions. Factor 1, which mostly explains the variation in options of less than 1 year, is explained by observable variables much better than is factor 2, which tends to explain the variation in options of more than 1 year. Over 90% of the variation in factor 2 is left unexplained.

*Practical Applications*

It is worthwhile to point out some of the practical applications of the principal component analysis and state variable regressions from above.

1. *Evaluation of hedging strategies for embedded options in insurance contracts or other option strategies:* Variable annuities in the United States and segregated funds in Canada are mutual fund investments with a specific maturity (say, 10 years) that are offered by insurance companies, and they often include an impressive variety of embedded options. For example, a policyholder may have a guaranteed death benefit (GMDB) and a guaranteed minimum accumulation benefit (GMAB) that equals the original investment (an ATM put). He may have embedded shout options that allow him to reset the guarantee level of the account value upon termination of the product to the account value at the reset date. This feature resets the life of the policy to 10 years from the date of the reset, which might be allowed at any time and up to twice per year for 10 years. He may be able to switch among asset classes at will. He also has the option to quit paying fees and to let the policy lapse.

The amount of policies outstanding is economically large, and the valuation of the embedded options is extremely difficult (Falloon, 1999, and Windcliff, Forsyth, & Vetzal, 2000, and references therein). Hedge strategies for various aspects of the policies can be evaluated by Monte Carlo simulation. For example, one way to hedge the reset option is to purchase a 10-year ATM put, and roll up the strike as necessary. For example, when the account value has risen and the policyholder resets, the formerly ATM put is now out of the money and the guarantee is for a higher index level. The company can sell the put and purchase a new one struck ATM.

The simulation requires the specification of the law of motion for the index and the entire implied volatility surface to evaluate the prices of the options. It is easy to simulate four random variables with the correlation properties of the three principal components and the index; it is straightforward to simulate plausible movements in the volatility surface in conjunction with the factor loadings. Without using the simplification of principal components, the analyst would have to specify the laws of motion for many points on the surface that are consistent (under any set of random draws) with real-world movements. Similar comments apply to realistic Monte Carlo evaluation of other hedging strategies.

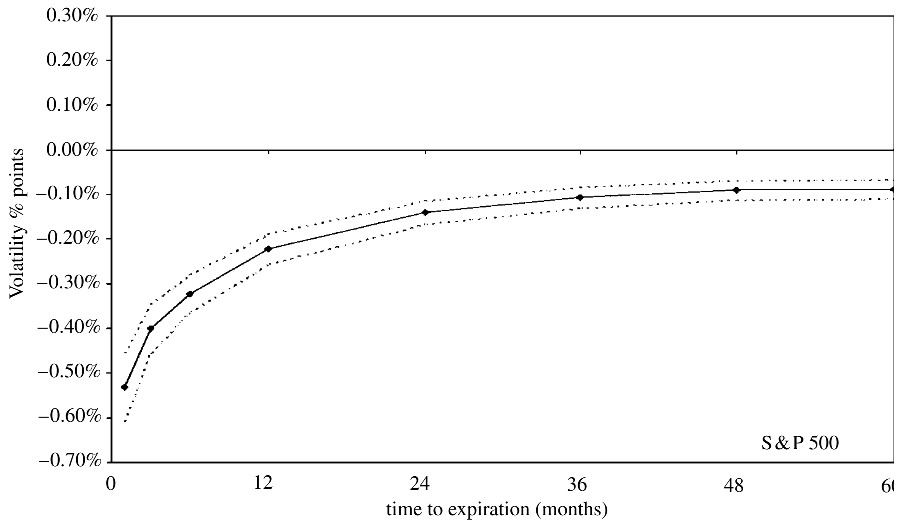
2. *Constructing hedge ratios consistent with empirical regularities:* the results from the regressions of factor scores on explanatory variables can be used to compute improved hedge ratios. The computations are simple due to the orthogonality of the principal components. For example, let $\delta_{BS}(\cdot)$ and $\nu_{BS}(\cdot)$ be the Black-Scholes delta and vega, respectively. Then

$$\begin{aligned}
\delta(\cdot) &= \frac{\partial C_t(S_t, \sigma[F_1(S_t), F_2(S_t), F_3(S_t)])}{\partial S_t} \\
&= \delta_{BS}(\cdot) + \nu_{BS}(\cdot)\left[\sum_{i=1}^{3} \frac{\partial \sigma[\cdot]}{\partial F_i}\frac{\partial F_i}{\partial S_t}\right] \\
&= \delta_{BS}(\cdot) + \nu_{BS}(\cdot)\left[\sum_{i=1}^{3} \frac{\partial \ln(\sigma[\cdot])}{\partial F_i}\frac{\partial F_i}{\partial \ln(S_t)}\frac{\sigma[F_i(S_t)]}{S_t}\right] \\
&= \delta_{BS}(\cdot) + \nu_{BS}(\cdot)\left[\sum_{i=1}^{3} L_{MTi}\,\beta_{\Delta Si}\frac{\sigma}{S_t}\right]
\end{aligned}$$

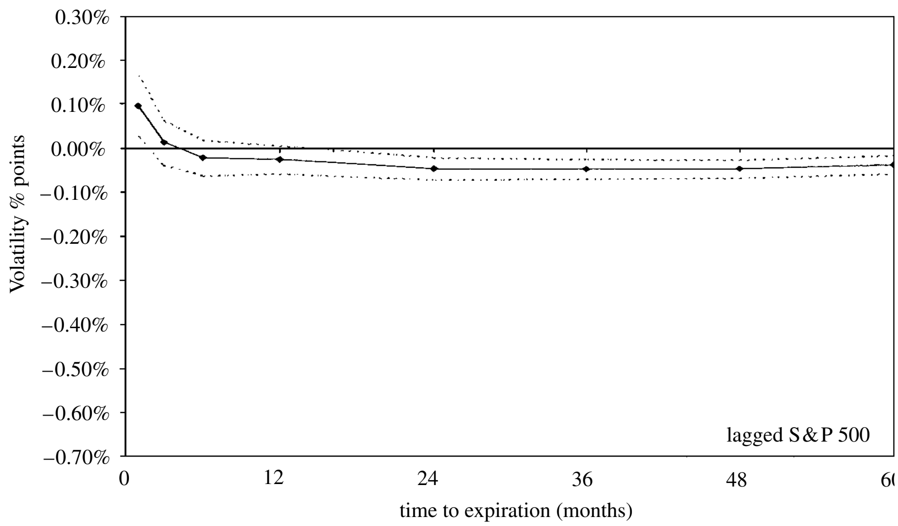
where $\beta_{\Delta Si}$ is the regression coefficient of stock returns on factor score $i$, and $L_{MTi}$ is the factor loading of factor $i$ on an option volatility with moneyness $M$ and time to expiration $T$. This expression can easily be evaluated at the proper values of $S$ and $\sigma$. It is valid for options with an estimated factor loading, but other values can be interpolated. Given the regression results above, it may be fruitful to truncate the sum after the first term. This method is computationally simple, yet it easily captures empirically realistic relationships. For example, the delta from the above equation is closer to zero than the Black-Scholes delta, as it accounts for the offsetting movements in spot and volatility.

*Macroeconomic Factors and the Volatility Term Structure*

The above analysis suggests that movements in the level of the term structure dominate movements in the surface. Figure 2 explores the direct relation between observable variables and changes in implied volatility. A regression of the log change in ATM-implied volatility on the observed variables was estimated separately for each S&P 500 option maturity available from 1 to 60 months. The regressions are meant as a linear projection onto the relevant explanatory variables rather than as a literal model for the evolution of implied volatility. The data are a longer series available from May 9, 1994 to June 22, 2001. To aid in interpretation, each independent variable is scaled by its mean absolute value. As before, the charts represent the change in implied volatility for the median term structure over the sample. The resulting coefficients thus reflect the



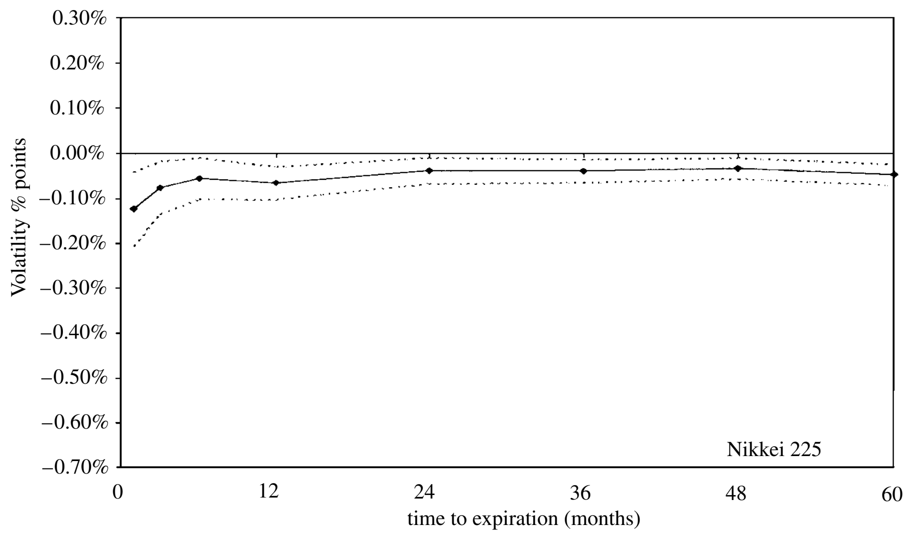

Panel A



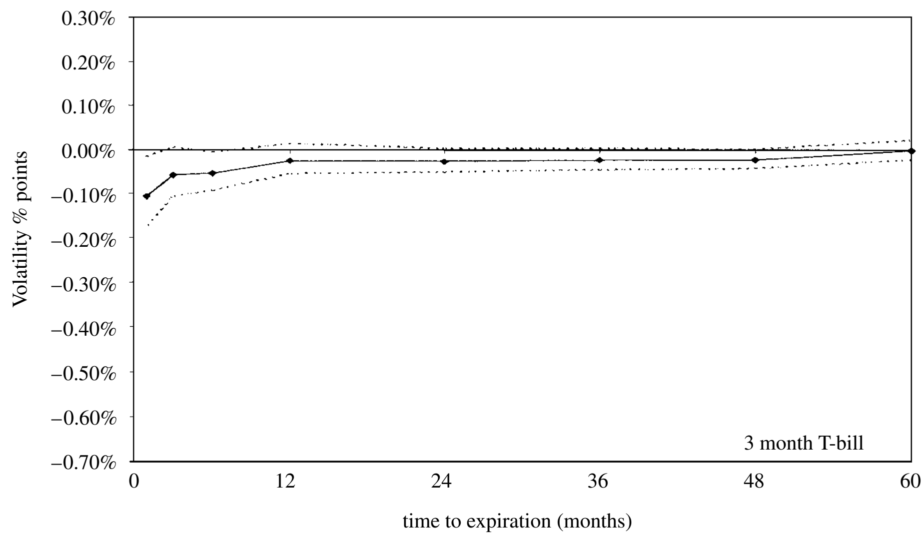

Panel B

**FIGURE 2**

Response coefficients for S&P 500 volatility term structure. Panels A through E display estimated response coefficients and 1.68 standard error bands for the median S&P 500 implied volatility term structure. The log first difference of each option maturity at-the-money volatility was regressed on the set of six independent variables. Each panel graphs the set of coefficients for the independent variable listed, plotted versus the maturity of implied volatility. Panels A and B show the coefficients on contemporaneous and lagged log returns of the S&P 500, respectively. Panel C shows the coefficients on log returns for the Nikkei 225. Panel D shows the coefficients on changes in the constant maturity 3-month Treasury bill rate. Panel E shows the coefficient on changes in the slope of the yield curve, measured by the constant maturity 10-year Treasury yield less the constant maturity 3-month T-bill rate. Panel F shows the coefficients on changes in the spread between Moody's AAA index yield and the constant maturity 30-year Treasury bond. The data are weekly observations from May 9, 1994 to June 22, 2001 for 371 observations. (Continued)




Panel C



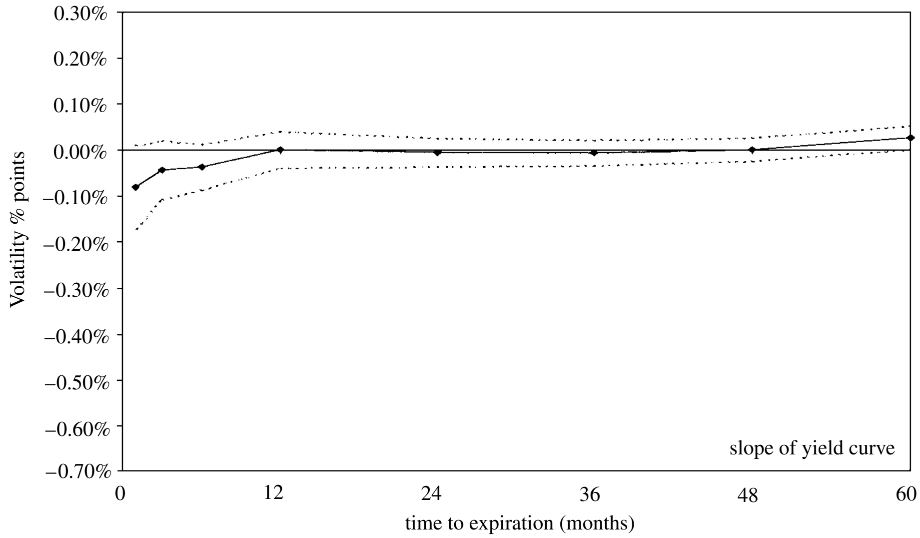

Panel D

**FIGURE 2**
(Continued)

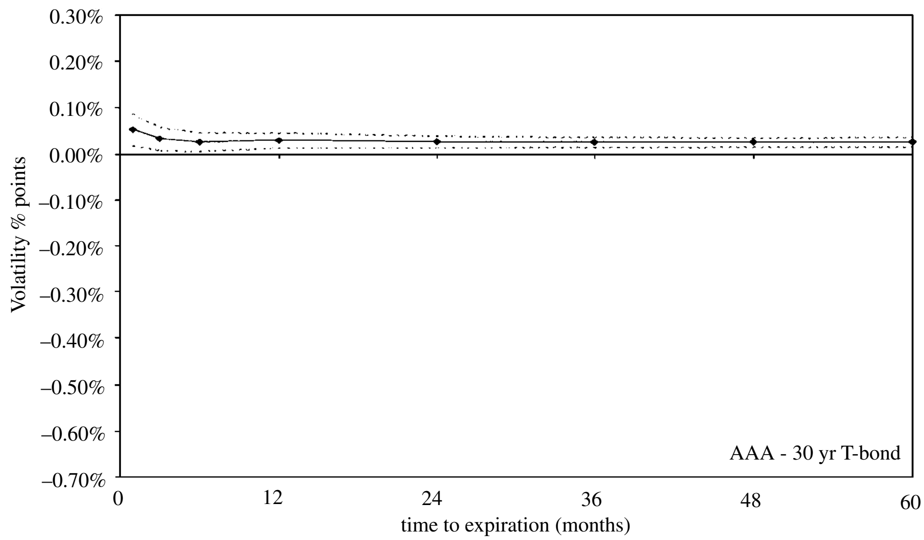
change in implied volatility for a positive innovation of typical size. The regression coefficients and 1.68 standard error bands for a given variable are graphed in Figure 3.[4]

Volatility is increased (decreased) at all maturities for a negative (positive) S&P return; the effect is decreasing with time to maturity of

[4]Tests for asymmetric responses of implied volatility to innovations in the variables provided no evidence against symmetry. Moving average error terms are included in several of the regressions.




Panel E




Panel F

**FIGURE 2**
(Continued)

the option. The lagged S&P return significantly affects volatility for all maturities except for the 1- and 3-month options. Beyond the 3-month maturity, the effect of the contemporaneous and lagged returns are approximately the same as each other and the same at all maturities. The interpretation is that the path of the index is very important in determining the level of the term structure beyond the shortest maturities. For example, a decrease in the index followed by a increase has little

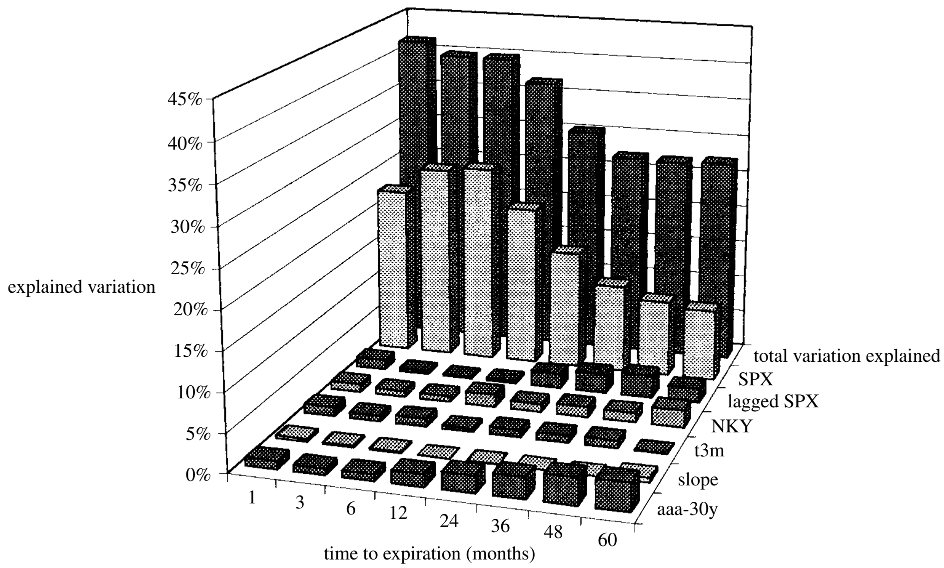



**FIGURE 3**
Variation in implied volatility explained by observable economic factors. The columns display the marginal explanatory power of each variable in a regression explaining the log change in at-the-money implied volatility for each maturity. The data are weekly observations from May 9, 1994 to June 22, 2001 for 371 observations.

effect on the long end of the term structure, but two weekly movements in the same direction have a strong effect.

Interestingly, Nikkei returns significantly affect the volatility at all maturities. The effect has the same sign as the effect of the S&P return, but the magnitude is much lower and represents a parallel shift at all maturities. An increase in the short rate significantly decreases volatility only for the 1-month maturity, with no significant effect otherwise. There is no evidence that a change in the interest rate term structure affects the volatility. Increases in the corporate bond spread effect a significant parallel shift in volatility, but the effect is typically small (less than 10 basis points). Finally, the regression $R^2$ values for the implied volatility regressions more clearly tell the same story as the factor score regressions. Explained variation declines almost monotonically from the 1-month maturity ($R^2 = 40\%$) to the 60-month maturity ($R^2 = 26\%$). Observable variables explain far less of the variation in long-dated volatility than in short-dated volatility.

Figure 3 displays the marginal increase in $R^2$ due to each independent variable, for each maturity available. For each maturity ATM option, the explanatory regression was estimated both with and without the studied variable. The observed increase in the $R^2$, given the effects of the other variables, is depicted as a column in the chart. The striking thing about the chart is that the total explained variation (including a moving average correction if necessary) is never higher than 45%. The contemporaneous index return offers the largest explanatory power, which is at a maximum around 25%, although this generally declines with the time to expiration. Some of this decline is matched by an increase in the marginal explanatory power of the lagged index return at longer maturities. The only other variable with non-negligible explanatory power is the corporate bond yield spread, which explains around 5% of the variation in the longer term volatilities.

Given the relatively long period and the changing market environments over which the regressions are estimated, it is useful to gauge how stable the estimated relationships are. A battery of tests was computed to test the null hypothesis of no structural break in the regressions. First, Chow tests were computed, with the null hypothesis that the break occurred in October 1997, about halfway through the sample. This point splits the sample into the early period of relatively low but rising volatility and generally positive returns and the later period of higher volatile volatility and a nontrending market. On the other hand, the breakpoint, if it exists, may have occurred at another time. The sup-$F$ test of Andrews (1993) was also computed. This test computes $F$-type tests at each possible breakpoint in the sample and compares the maximum test statistic to critical values that account for the bias caused by searching over many points.

The results of the tests are in Table III. The Chow tests find evidence of instability only in the regressions involving options that expire

**TABLE III**

Regression Instability Tests

| *Time to Expiration (Months)* | *Chow* | *Andrews sup-F* |
|---|---|---|
| 1 | 1.473 | 1.989 |
| 3 | 1.334 | 3.076 |
| 6 | 1.114 | 2.842 |
| 12 | 1.214 | 4.202 |
| 24 | 2.880* | 6.178 |
| 36 | 1.943 | 6.117 |
| 48 | 1.444 | 5.978 |
| 60 | 0.816 | 4.606 |

Note. Each column provides the test statistic for the null hypothesis of no parameter instability. The Chow test assumes that the breakpoint occurred in October 1997. The maximum value for the sup $F$-test is shown. The symbol "*" indicates significance at the 1% level. The $F$-test 5% critical value is 2.035; the 1% critical value is 2.69. The 5% critical value for the Andrews sup-$F$ test is 21.84; the 1% critical value is 26.23.

in 2 or 3 years. The $F$-statistic for the 2- year regression is larger than the 1% critical value and the $F$-statistic for the 3-year regression is close to the 5% critical value ($p$-value = 6.2%). Although the tests are not independent of each other, the results do not strongly suggest a structural break at that point. Judging by asymptotic critical values, the Andrews test does not reject the hypothesis of stability at any maturity. As with the Chow test, the statistics are not independent across regressions, and the results are suggestive rather than conclusive. Nonetheless, the statistics provide little obvious evidence that the regressions grossly misrepresent the true data generation process.

## CONCLUSION

This article has investigated the systematic factors explaining movements in the implied volatility surface for S&P 500 index options. A large number of option volatilities, ranging from 1 month up to 5 years to expiration and from 10% out of the money puts to 10% out-of-the-money calls were examined. The principal components analysis indicates that three unobservable components are sufficient to explain the majority of volatility movements. The first factor explains the variation of volatility for options with less than 1 year to maturity. It shifts the entire term structure in the same direction, although the nearest dated options are shifted by the greatest magnitude. The second factor explains movements in volatility for options with greater than 1 year to maturity. It is a spread factor that changes the slope of the implied volatility term structure. A third factor especially explains variation in 1-month option volatility, increasing the volatility of out-of-the-money puts over that of out-of-the-money calls.

Most importantly, the relation of the implied volatility surface to economic state variables is also examined. Domestic stock returns are highly significant in explaining changes in volatility, with a diminished effect at longer option maturities. Longer maturity options are significantly affected by both contemporaneous and lagged returns, which indicates that the path of the index determines the slope of the volatility term structure at longer maturities. For short maturity options, a decline in the index is associated with significant increases in volatility. For longer dated options, consecutive changes in the index of the same sign are associated with changes in implied volatility of the opposite sign (e.g., two week-on-week declines in the index raise long-dated volatility). Two weeks with differently signed returns produce a muted effect on long-term volatility. Foreign returns cause a parallel shift in the volatility term structure: the marginal effect of a decline in the Nikkei 225 is to raise index volatility slightly. Increases (decreases) in short rates decrease (increase) volatility. An increase in the spread between high quality corporate bond yields and government bond yields is also associated with a parallel shift upward in the volatility term structure.

The results are broadly consistent with Bayesian learning models of asset pricing. David and Veronesi (2000) assume that the aggregate dividend process for the economy switches between two unobservable states (recession and expansion), and conclude that the equilibrium implied volatility surface for options on the market fluctuates as investor beliefs are updated. A same-sign sequence of dividend innovations and returns causes investors' posterior beliefs to evolve. The results in this article are consistent with investors who update their beliefs on the economy using a variety of data on foreign and domestic stock returns as well as interest rates.

The results indicate that by combining principal components estimation and simple linear regressions, very simple methods for computing improved hedge ratios can be computed. Perhaps more importantly, the results are useful to analysts who wish to evaluate hedging strategies by using Monte Carlo simulations. This can be useful to analyze extremely complicated option portfolios that defy analytical solutions, for evaluating the risk to strategies that capture perceived mispricing or apparently attractive option trades, or for stress testing option portfolios. It is straightforward to simulate realistic changes to the implied volatility surface, complete with appropriately correlated shocks to the index. Without using data reduction techniques, the problem is far more difficult if not intractable.

The stylized facts produced by the investigation lead naturally to a two-factor stochastic volatility model with jumps for risk-neutral index volatility. These models require a quickly mean reverting component that reverts towards a more slowly moving component that defines the stochastic intermediate term mean. Furthermore, the increase in the option smile from volatility increases also points to a jump model where the jump intensity (or risk aversion) is correlated with the instantaneous volatility. The analysis provides guidance for estimation strategies that try to exploit the information in asset returns and option prices. At least three options are required to fully capture the dynamics of risk-neutral volatility. An out-of-the-money short-term option (e.g., a put), and an at-the-money or out-of-the-money short-term option (e.g., a call) are required to capture shifts in the near-term volatility smile. A single long-term option may be sufficient to capture shifts in the long-dated volatility and the term structure slope. At the same time, there would appear to be little advantage by including other variables in the estimation. For example, one could imagine that including short-term interest rates in estimation might provide valuable information above that provided by index returns. The analysis here suggests that changes in the index subsume almost all of the relevant information available from obvious macroeconomic state variables.

## BIBLIOGRAPHY

Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown changepoint. Econometrica, 61, 821–856.

Bakshi, G., Cao, C., & Chen, Z. (1997). Empirical performance of alternative option pricing models. Journal of Finance, 52, 2003–2050.

Bakshi, G., Cao, C., & Chen, Z. (2000). Pricing and hedging long-term options. Journal of Econometrics, 94, 277–318.

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

Bartlett, M. (1937). The statistical conception of mental factors. British Journal of Psychology, 28, 97–104.

Bates, D. (1996). Jumps and stochastic volatility: Exchange rate processes implicit in Deutsche Mark options. Review of Financial Studies, 9, 69–107.

Bates, D. (2000). Post-'87 crash fears in S&P 500 futures options. Journal of Econometrics, 94, 181–238.

Chen, N.- F., Roll, R., & St. Ross, S. (1986). Economic forces and the stock market. Journal of Business, 59, 383–403.

Das, S., & Sundaram, R. (1999). Of smiles and smirks: A term structure perspective. Journal of Financial and Quantitative Analysis, 34, 211–239.

David, A., & Veronesi, P. (2000). Option prices with uncertain fundamentals: Theory and evidence on the dynamics of implied volatilities. Manuscript, University of Chicago.

Engle, R., & Lee, G. (1993). A permanent and transitory component model of stock return volatility. Unpublished manuscript, University of California, San Diego.

Falloon, W. (1999). Canada's option nightmare. Risk, 12(8), 6.

Franks, J., & Schwartz, E. (1991). The stochastic behaviour of market variance implied in the prices of index options. The Economic Journal, 101, 1460–1475.

Gallant, A. R., Hsu, C.-T., & Tauchen, G. (1999). Using daily range data to calibrate volatility diffusions and extract the forward integrated variance. Review of Economics and Statistics, 81, 617–631.

Hull, J. (1997). Options, futures, and other derivatives. Upper Saddle River, NJ: Prentice Hall.

Johnson, R., & Wichern, D. (1982). Applied multivariate analysis. Englewood Cliffs, NJ: Prentice Hall.

Knez, P., Litterman, R., & Scheinkman, J. (1994). Explorations into factors explaining money market returns. Journal of Finance, 49, 1861–1882.

Litterman, R., Scheinkman, J. (1991). Common factors affecting bond returns. Journal of Fixed Income, 1(1), 54–61.

Rubenstein, 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, 455–480.

Schwert, G. W. (1989). Why does stock market volatility change over time? Journal of Finance, 44, 1115–1153.

Skiadopoulos, G., Hodges, S., & Clewlow, L. (1999). The dynamics of the S&P 500 implied volatility surface. Review of Derivatives Research, 3, 263–282.

Windcliff, H., Forsyth, P. A., & Vetzal, K. R. (2000). Valuation of segregated funds: Shout options with maturity extensions. Unpublished manuscript, University of Waterloo.

Zhu, Y., & Avellaneda, M. (1997). An E-ARCH model for the term structure of implied volatility of FX options. Applied Mathematical Finance, 4, 81–100.