A Nonlinear Factor Analysis of S&P 500 Index Option Returns

Christopher S. Jones
Simon School of Business
Carol Simon Hall 3-110P
University of Rochester
Rochester, NY 14627
(716) 275-3491
c_s_jones@simon.rochester.edu
http://www.simon.rochester.edu/fac/c_s_jones

January 2001

Abstract

This study confronts the growing evidence that multiple sources of priced risk appear necessary to explain the expected returns of equity index options. A general class of nonlinear latent factor models is estimated using a data set of over 32,000 daily return observations on S&P 500 index put options of varying maturity and moneyness. The results show that while priced factors other than the return on the underlying security contribute to these expected returns, factor-based models are insufficient to explain their magnitude. For a variety of option classes, but particularly short-term out-of-the-money puts, the magnitude of the mispricing remains large. Out-of-sample tests confirm the usefulness of the semiparametric latent factor approach in hedging a book of put options and in forming portfolios that exploit apparent option mispricing.

JEL classifications: C11, C14, G12
Keywords: Bayesian analysis, factor models, option pricing, stochastic volatility

1 Introduction

There is growing evidence that extraordinary returns may be obtained through the sale of options on the S&P 100 and S&P 500 Indexes. In a recent paper, Coval and Shumway (2000) examine the returns on delta-hedged option positions, focusing on the at-the-money straddle, and find average returns close to minus three percent per week. Since these positions are approximately invariant over short time horizons to movements in the underlying index, their CAPM betas should be near zero and their expected returns close to the risk-free rate. Coval and Shumway conclude that large deviations from the CAPM are the norm, which they suggest is evidence that some other systematic factor, such as stochastic volatility, might be priced by the market. Similar conclusions are made by Jackwerth (2000), who finds risk-adjusted profitability from selling puts, and by Bakshi and Kapadia (2000), who find that the volatility risk premium contributes significantly to higher prices for calls and puts. Buraschi and Jackwerth (2000) concur that more than one priced risk factor is necessary to explain option prices.

A less direct set of evidence is the now standard observation that the Black-Scholes implied volatilities of equity index options are upward-biased predictors of future realized volatility. Numerous studies, including Day and Lewis (1992), Fleming, Ostdiek, and Whaley (1995), Christensen and Prabhala (1998), and Blair, Poon, and Taylor (1999) have documented that implied volatilities of short-term at-the-money equity index options are on average several percent higher than the volatilities realized over the option's life. This bias is apparent even in the VIX index (an index of the implied volatility of hypothetical 22-day at-the-money S&P 100 option), whose average over the period from January 1986 to September 2000 was 20.4 percent. The realized index volatility over the same period was just 18.2 percent. After January 1990, realized volatility was 15.8 percent, but average implied volatility declined just to 19.2 percent. Under the Black-Scholes assumptions, at-the-money put and call options were overpriced by between 20 and 25 percent over this period.

Finally, a related conclusion can be made from the estimation of the risk-neutral parameters of stochastic volatility models used in option pricing. Both Bakshi, Cao, and Chen (1997) and Bates (2000) argue that the risk-neutral parameters required to fit the Heston (1993) model and more general stochastic volatility models to options prices are unrealistic. Large volatility risk premia are found by Benzoni (1999), Chernov and Ghysels (2000), Jones (2000), and Pan (2000). In many of these papers, the price of volatility risk is sufficiently high to induce explosive volatility dynamics under the risk-neutral measure. While this explosivity is not a theoretical problem, it implies volatility term structures that are too upward-sloping to be realistic.

While there is a high level of agreement that volatility risk is priced and contributes to higher option prices, the sufficiency of volatility risk premia to explain the high prices of options and the large negative returns they generate remains unproven. The purpose of this paper is therefore to ask whether the extraordinary returns that appear possible by short selling put options can be explained by *any* set of priced factors, or whether options appear mispriced under very general assumptions about the nature of systematic risk.

In order to consider as large a class of models as possible, I adopt a semiparametric approach to option pricing and hedging. Option hedge ratios, expected returns, and other unknown functions will be approximated using flexible but parsimonious sets of orthogonal polynomials. Using this approach we can approximate models that have been proposed already and consider many more models that have not. In addition, I make almost no assumptions about what the sources of systematic risk in options actually are, giving the approach an additional degree of flexibility beyond what is typically allowed in the literature. While the latent factors are undoubtedly related to observables such as the market return and its volatility, stochastic dividends, interest rates, and risk aversion might also be important sources of systematic risk. This approach allows for the possibility that these sources of risk are important.

The flexibility of the polynomial approximations and the latent nature of the factors makes estimation accuracy particularly important. To address this, I adopt a Bayesian approach with uninformative priors, which should provide an approximation to maximum likelihood and therefore result in high efficiency. Second, I use an extremely large data set of put option returns, totaling over 32,000 observations, which should allow a large number of parameters to be estimated with high accuracy. Finally, I follow Bauer and Tamayo (2000) by using observables, such as the market return, to provide some information about latent systematic factors without assuming that the two are the same.

The semiparametric nature of the analysis links it to a line of research initiated by Hutchinson, Lo, and Poggio (1994) and recently refined by Garcia and Gençay (2000) that seeks to identify the option pricing formula using nonparametric methods. Related approaches have been pursued by Aït-Sahalia and Lo (1998), Broadie, Detemple, Ghysels, and Torrès (2000) and Christoffersen and Hahn (1999), among others. The aim of this paper is fundamentally different, however, since its focus, like that of Coval and Shumway (2000), is on modeling expected option returns, rather than option prices, and decomposing these expected returns into components related to systematic risk factors and components attributed to mispricing.

The estimation results of the paper show that no set of systematic risks appears likely to explain the consistently negative returns of many option portfolios. While the various modifications of the general model perform differently, it is unambiguous that short-term out-of-the-money put options have expected returns that are too negative to be consistent with a factor-based explanation. Other options appear to be mispriced as well, but which options these are depends on the model specification.

Several out-of-sample hedging and portfolio allocation exercises are considered to reinforce these results and to reduce the concern that the in-sample results suffer from model overfitting. I find that for the purpose of hedging a book of options, the flexible semiparametric approach often outperforms hedging strategies based on matching the Black-Scholes "Greeks." The portfolio allocation exercise shows that substantial excess returns may be made with relatively low risk, yielding out-of-sample Sharpe ratios on zero-beta portfolios several times higher than those of the market index or a naive put writing strategy.

The outline of the paper is as follows. Section 2 introduces and motivates the model, while section 3 discusses the orthogonal polynomial approximation of the model. Estimation methods are covered in section 4, and estimation results appear in section 5. Section 6 contains some in-sample and out-of-sample hedging and portfolio allocation experiments. Section 7 concludes.

2 The model

This section describes the construction of a factor model for option returns in which the factor loadings, idiosyncratic variances, and related quantities are nonlinear functions of time to expiration, moneyness, and a state vector.

2.1 Motivation

Suppose that asset price uncertainty is driven by a K-dimensional diffusion process S_t that is assumed to have zero drift. We observe the prices of a variety of options on the same underlying security that differ on the basis of time to expiration and moneyness. I posit that the price processes of all these options are time-homogeneous, and that they obey the SDE

$$\frac{dP_{it}}{P_{it}} = \mu\left(\tau_{it}, \kappa_{it}, \zeta_t\right) dt + \beta\left(\tau_{it}, \kappa_{it}, \zeta_t\right) dS_t, \tag{1}$$

where

- τ_{it} is the time to expiration of option i at time t,
- κ_{it} is the moneyness (strike / forward price) of option i at time t,
- ζ_t is a $1 \times J$ vector of information variables measurable at time t,

and where the functions $\mu : \Re^+ \times \Re^+ \times \Re^J \to \Re^1$ and $\beta : \Re^+ \times \Re^+ \times \Re^J \to \Re^K$ are identical across all options.

Fundamental risk in the economy is considered to be the noise process S_t, which is allowed to exhibit time-varying covariances as long as these covariances depend only on ζ_t. By definition, the price of risk $\lambda(\zeta_t)$ is the $K \times 1$ function that satisfies

$$\mu(\tau_{it}, \kappa_{it}, \zeta_t) = r_t + \beta(\tau_{it}, \kappa_{it}, \zeta_t)\,\lambda(\zeta_t) \tag{2}$$

for all τ_{it}, κ_{it}, and ζ_t, where r_t is the instantaneous short rate.

We may therefore rewrite the price process as

$$\frac{dP_{it}}{P_{it}} = [r_t + \beta(\tau_{it}, \kappa_{it}, \zeta_t)\,\lambda(\zeta_t)]\,dt + \beta(\tau_{it}, \kappa_{it}, \zeta_t)\,dS_t, \tag{3}$$

2.2 The discretized process

For tractability, I will work with the following discretized version of the model:

$$\frac{P_{it+1} - P_{it}}{P_{it}} = r_t + \beta(\tau_{it}, \kappa_{it}, \zeta_t)\,\lambda(\zeta_t) + \beta(\tau_{it}, \kappa_{it}, \zeta_t)\,E_{t+1}, \tag{4}$$

E_{t+1} is a $K \times 1$ vector of mean-zero random shocks, again with possibly time-varying covariances that depend on ζ_t.

Alternatively, by defining $R_{it+1} = (P_{it+1} - P_{it})/P_{it} - r_t$ and $F_{t+1} = \lambda(\zeta_t) + E_{t+1}$ we have

$$R_{it+1} = \beta(\tau_{it}, \kappa_{it}, \zeta_t)\,F_{t+1}, \tag{5}$$

which is reminiscent of familiar factor model notation.[1]

It is important to note that β depends only on lagged information, so it correctly represents the *ex ante* covariance between returns and factors.

2.3 Residual noise

Given a large enough K, the above model may be capable of matching many of the most salient return characteristics. However, with a small value of K, say three, the correlation structure imposed in (5) is unrealistic, as it implies that zero-variance hedges may be formed using just a few assets.

To enhance model realism while maintaining a tractable form (which requires that K be small), I will assume that all returns are subject to idiosyncratic conditionally Gaussian shocks. These shocks are assumed independent across assets and across time. As they are non-systematic, they are assumed unpriced.

Although the shocks are independent, they are not identically distributed. Consistent with the previous setup, I allow the residual standard deviation to depend on τ, κ, and ζ_t. The revised return equation therefore becomes

$$R_{it+1} = \beta(\tau_{it}, \kappa_{it}, \zeta_t)\,F_{t+1} + \epsilon_{it+1}, \tag{6}$$

where $\epsilon_{it+1} \sim \mathrm{N}(0, \sigma^2(\tau_{it}, \kappa_{it}, \zeta_t))$.

[1] As Coval and Shumway (2000) note, logarithmic returns will be equal to negative infinity for the final observation of an option that expires out of the money, hence expected returns will equal negative infinity as well. I therefore work with simple returns exclusively.

2.4 Mispricing

So far, I have assumed that mean excess returns are proportional to factor loadings. To allow for the possibility that option expected return deviate from this assumption, I introduce a function $\alpha(\tau, \kappa, \zeta_t)$ to capture systematic differences in expected returns from those implied by the factor model. With a less-than-formal justification, I refer to these deviations as mispricing. We therefore have

$$R_{it+1} = \alpha(\tau_{it}, \kappa_{it}, \zeta_t) + \beta(\tau_{it}, \kappa_{it}, \zeta_t) F_{t+1} + \epsilon_{it+1}. \tag{7}$$

Under the null hypothesis, $\alpha(\tau_{it}, \kappa_{it}, \zeta_t) \equiv 0$.

2.5 Relating factors to observables

So far, virtually no structure has been imposed on F_t, the systematic factors of the model. As the sum of a mean-zero shock vector with time-varying covariances and a vector of unspecified risk premia, the model currently has too much flexibility to be estimated. I seek to operationalize the model while allowing the structure to remain as flexible as possible.

There are two broad econometric approaches to estimating factor models such as (7). The first approach involves the preselection of certain economic variables that are believed by the researcher to explain the returns under observation. Two well-known examples are the papers by Chen, Roll, and Ross (1986) and Fama and French (1993). The other common approach is to use the returns themselves to infer the realizations of the factors, as was done in different ways by Connor and Korajczyk (1986) and Geweke and Zhou (1996), among many other papers too numerous to mention.

Recently, however, Bauer and Tamayo (2000) have proposed an intermediate approach in which observable economic variables other than returns are used to help infer, but not completely determine, the realizations of the factors. Adopting a similar approach, I assume a nonlinear relation between the $1 \times K$ factor vector and a row vector of *contemporaneous* information variables ξ_t:

$$F_t = f(\xi_t) + \eta_t, \tag{8}$$

where η_t is a $1 \times K$ matrix of i.i.d. residuals that are assumed Gaussian for tractability. As in Geweke and Zhou (1996), model identification requires that the covariance matrix of η_t be prespecified. Since the choice is inconsequential, for convenience we assume that

$$\mathrm{Cov}(\eta_t', \eta_t) = \mathbf{I}.$$

For determining what variables to include in ξ_t, it is helpful to recall the definition of F_t,

$$F_t = \lambda(\zeta_{t-1}) + E_t,$$

as the sum of a factor risk premium and a zero-mean random shock. It is clear that at a minimum, ξ_t should include ζ_{t-1}, because of the latter's impact on the price of risk. But ξ_t should also include some variables that are not measurable at time $t-1$ that should be related to the factor noise, E_t.

Since ζ_t represents a state vector used for describing return means and covariances, it is likely to contain relatively persistent variables, such as a market volatility indicator, that could explain presumably persistent variation in these moments. In addition to lagged values of these variables, the vector ξ_t should include variables such as the market return or the *change* in the volatility indicator, variables that are likely to capture some of the unpredictable variation in F_t.

2.6 Return (non-)normality and model limitations

It is important to note that several of the previous assumptions have eliminated the types of non-normality considered in the model. First, the assumption that idiosyncratic errors (ϵ_{it}) in returns are conditionally Gaussian implies that all conditional non-normality in returns is due to systematic components. Since not much is known about the instrument-specific component of option returns, this assumption is questionable. Nevertheless, by allowing idiosyncratic variances to depend on time to expiration, moneyness, and other state variables, normality is not imposed at the unconditional level.

Second, the conditional distribution of F_t given ξ_t is Gaussian, since the residual η_t is a multivariate standard normal. Unconditionally, F_t is non-normal through its dependence on ξ_t, whose dynamics are not modeled. This is both because the variables in ξ_t are likely non-normal, but also because ξ_t impacts F_t in a nonlinear manner. In addition, even if F_t were Gaussian, unconditional non-normality in option returns would be generated from the dependence of β on time to expiration, moneyness, and one or more state variables.

Finally, in addition to it's role in producing non-normality, ξ_t is also the only possible source for time-varying correlations in the factors, as the covariance matrix of η_t is constant. Changes in the correlations of the returns of different options will also be affected by time variation in β, however.

3 Approximating the unknown functions

Since the true functional forms of $\alpha(\cdot)$, $\beta(\cdot)$, $\sigma(\cdot)$, and $f(\cdot)$ are unknown, flexibility in their specification is desirable. I use a flexible set of orthogonal polynomial basis functions that are capable of very accurate approximation while retaining econometric tractability.

3.1 Choice of approximation scheme

For approximation using basis function, there are at least three issues that must be addressed. The first is to decide on the a set of basis functions. One approach might be to use the set of all monomials and their products, leading to the familiar Taylor series approximation. However, since many of these basis functions are highly collinear, a more efficient and easily interpreted approximation scheme uses orthogonal polynomials.

All functions will therefore be approximated using Legendre polynomials, a set of polynomial basis functions that are orthogonal to one another on the interval $[-1, 1]$. Since they are polynomials, the Legendre polynomials span the same space as a Taylor series of the same order. However, after a univariate data series is rescaled to lie in the interval $[-1, 1]$, different terms of the Legendre series will be closer to uncorrelated than will terms of the Taylor series. If the data are uniformly distributed over this interval, the correlation will be zero. In practice, deviations from uniformity will lead to a large reduction, but not elimination, of correlation between terms in the expansion relative to the Taylor series.

The second issue involves choosing the variable in which the expansion is taken. The function $g(x) = |x|^{2.1}$, for example, need not be expanded in x, but might instead be expanded in some function of x, for example $h(x) = x^2$. This would undoubtedly be a vast improvement if only first-order expansions were considered.

Ideally we would choose $h(x) = g(x)$, so that a first-order expansion would be an exact fit. Since the function $g(x)$ is generally unknown, this choice is infeasible. To the extent that we are willing to "guess" the approximate form of $g(x)$, however, we may be able to improve the approximation with a fixed number of terms. The expansion variables I choose are based on an analysis of the Black-Scholes model. Although this model is no doubt an oversimplification of reality, it still may be a useful guide for selecting variables for the unknown, more complicated model. I provide some evidence below that indicates that this is the case.

Lastly, we must choose the order of the approximation, which is difficult when $g(x)$ is unknown. Again, to the extent that we are willing to guess a realistic if incorrect proxy for $g(x)$, we might learn the necessary number of terms to include. After much experimentation, some of which is reported below, I have settled on a third-order expansion. This expansion consists of a constant, the first, second, and third order Legendre polynomials in each variable, and cross-terms formed by taking products of the first and second order polynomials of different variables.

3.2 A suggestive analysis of approximation accuracy

For a suggestive analysis of how well these approximations might fare, I consider several experiments. In the first experiment I ask how the approximation method would fare for the Black-Scholes model. I generate 25,000 random maturities and moneyness levels by drawing τ_i ($i = 1, ..., 25,000$) from a uniform distribution between 10 and 250 days, and κ_i from a uniform distribution between .8 and 1.1. I assume that $\beta(\cdot)$ measures the sensitivity of option returns to a change in the price of the underlying asset. Assuming a constant volatility of one percent per day and a constant interest rate of 6/264, percent per day[2], I therefore compute the function $\beta(\tau_i, \kappa_i)$ as the Black-Scholes put delta divided by the Black-Scholes put price.

The approximation considered is a third-order Legendre series. Since the Legendre polynomials are orthogonal on $[-1, 1]$, variables are rescaled to lie in that interval. Let τ^* and κ^* denote the rescaled versions of τ and κ, constructed as

$$\tau_j^* = 2\left(\frac{\tau - \min_i\{\tau_i\}}{\max_i\{\tau_i\} - \min_i\{\tau_i\}}\right) - 1 \tag{9}$$

for τ^* and similarly for κ^*. Then the Legendre series approximation is given by

$$\begin{aligned}\beta(\tau_i, \kappa_i) \approx\ & b_0 + b_1 P_1(\tau_i^*) + b_2 P_1(\kappa_i^*) + b_3 P_2(\tau_i^*) + b_4 P_2(\kappa_i^*) + b_5 P_3(\tau_i^*) \\ & + b_6 P_3(\kappa_i^*) + b_7 P_1(\tau_i^*) P_1(\kappa_i^*) + b_8 P_2(\tau_i^*) P_1(\kappa_i^*) + b_9 P_1(\tau_i^*) P_2(\kappa_i^*),\end{aligned} \tag{10}$$

where $P_n(\cdot)$ denotes the n^{th} order Legendre polynomial. After adding an error term, we may estimate the vector b using OLS. This results in an unadjusted R-squared of .8917, indicating that the series expansion captures the general shape of the true function but is easily distinguished from it.

To improve the fit without adding additional terms, the second approximation uses some intuition from the Black-Scholes formula to change the variables in which the expansion is taken. The new approximation is

$$\begin{aligned}\beta(\tau_i, \kappa_i) \approx\ & b_0 + b_1 P_1(x_i^*) + b_2 P_1(y_i^*) + b_3 P_2(x_i^*) + b_4 P_2(y_i^*) + b_5 P_3(x_i^*) \\ & + b_6 P_3(y_i^*) + b_7 P_1(x_i^*) P_1(y_i^*) + b_8 P_2(x_i^*) P_1(y_i^*) + b_9 P_1(x_i^*) P_2(y_i^*),\end{aligned} \tag{11}$$

where x_i^* and y_i^* are rescaled versions of

$$x_i = \log(\tau_i) \quad \text{and} \quad y_i = \frac{\kappa_i - 1}{\sqrt{\tau_i}}. \tag{12}$$

[2] A daily rate of 6/264 percent represents an annualized rate of 6 percent in a year with 264 trading days.

The approximation therefore standardizes moneyness by the time to expiration and then lets the additional impact of τ enter through its logarithm. Using the new approximation as the basis of a regression equation results in an R-squared of .9999.

Figure 1 provides a graphical illustration of the two approximation methods. While both approximations capture the general shape of the true function, the absolute errors of the approximation (10) are as large as 300% of the true function. The approximation (11) can be seen to reproduce the true function almost perfectly.

Alternatively, we may redefine $\beta(\cdot)$ as the sensitivity to changes in volatility (B-S vega divided by price), the interest rate (B-S rho divided by price), or time (B-S theta divided by price). The results of approximating these different function are summarized in panel A of Table 1. The table shows that the approximation that uses renormalized variables is generally far superior. Only in measuring the sensitivity of returns to changes in the interest rate does the approximation (10) perform slightly better, and in this case both approximations are nearly perfect.

To gauge whether this renormalization is beneficial in more complex models, I also perform an experiment with Heston's (1993) square root stochastic variance model.[3] In addition to the values of τ_i and κ_i simulated above, I also draw 25,000 values of instantaneous volatility, v_i, from the uniform distribution between .5 and 2.5 percent per day.

Because the "Greeks" must be calculated numerically from prices that themselves are computed using numerical integration, some combinations of τ, κ, and v_i generated extremely large errors in the ratio of the partial derivative (delta, vega, theta, or rho) to the option price. These errors were confined to options that were extremely cheap, so I redrew all combinations of τ, κ, and v_i that resulted in puts prices below .001 cents per dollar of notional principal. Since the average put price in the sample was 2.6 cents per dollar notional and fewer than 1.5 percent of draws were repeated, the remaining puts still represent an extremely wide variety of moneyness, maturity, and volatility.

The $\beta(\cdot)$ function now depends on three variables, τ_i, κ_i, and v_i, and the first approximation considered consists of a third-order Legendre series in these three untransformed variables, rescaled to lie between -1 and 1. As before, we also consider a transformation that renormalizes the moneyness measure and converts remaining variables to logarithms. Now writing moneyness in terms of the number of standard deviations between the current stock price and the option's strike, the second approximation uses rescaled versions of the transformed variables

$$\log(\tau_i), \quad \frac{\kappa_i - 1}{\sqrt{\tau_i} v_i}, \quad \text{and} \quad \log(v_i). \tag{13}$$

The R-squares from estimating each Legendre series in a regression are reported in panel B of Table 1. Results again indicate the superiority of using the transformed variables, with the second approximation yielding R-squares that are again close to one. Although the R-squares are not as close to one as they were for the Black-Scholes model, part of the reduction is probably due to errors in the numerical integration required to compute option prices and Greeks under the Heston model.

3.3 The final approximation

In the primary model considered in the paper, $\zeta_t = \hat{v}_t$, where $\hat{v}_t$ is a market volatility proxy to be determined later. It is likely that the correct specification of ζ_t would include a number of other variables such as interest rates and dividend yields. In the interests of parsimony, and because no variable is as obviously important as market volatility, I restrict ζ_t to be univariate.

[3] The model (with parameter values on a daily frequency) is given by $dS_t = (.06/264)S_t dt + \sqrt{V_t} S_t dB_{1t}$, $dV_t = .01(.0001 - V_t)dt + .001\sqrt{V_t} dB_{2t}$, and $\text{Corr}(dB_{1t}, dB_{2t}) = -.5$.

Each of the K dimensions of $\beta(\tau_{it}, \kappa_{it}, \hat{v}_t)$ will be approximated by a separate third-order Legendre series in rescaled versions[4] of

$$\log(\tau_{it}), \quad \frac{\kappa_{it} - 1}{\sqrt{\tau_{it}}\hat{v}_t}, \quad \text{and} \quad \log(\hat{v}_t). \tag{14}$$

The logarithm of $\sigma(\tau_{it}, \kappa_{it}, \hat{v}_t)$ is similarly approximated. We can therefore write

$$\begin{aligned} \beta(\tau_{it}, \kappa_{it}, \hat{v}_t) &\approx X_{it}\mathbf{B} \\ \log \sigma(\tau_{it}, \kappa_{it}, \hat{v}_t) &\approx X_{it}s, \end{aligned} \tag{15}$$

where $\mathbf{B}$ is $L \times K$, s is $L \times 1$, and X_{it} is the $1 \times L$ vector of the Legendre series terms derived from $\log(\tau_{it})$, $(\kappa_{it} - 1)/(\sqrt{\tau_{it}}\,\hat{v}_t)$, and $\log(\hat{v}_t)$.

For the third-order expansion in three variables, $L = 19$, implying for a three-factor model that there are $(3+1) \times 19 = 76$ unknown parameters in $\mathbf{B}$ and s alone. Because the number of option return observations is large, this may not be an excessive number of parameters, and estimation errors will hopefully be moderate.

Since estimation error tends to be larger in means than in covariances, it is desirable to approximate the intercept function $\alpha(\cdot)$ more parsimoniously. To approximate this function I use a second-order Legendre series in the variables in (14), appropriately rescaled. Letting W_{it} denote the $1 \times H$ vector of terms in this expansion, we therefore have

$$\log \alpha(\tau_{it}, \kappa_{it}, \hat{v}_t) \approx W_{it}a, \tag{16}$$

where a is an $H \times 1$ vector of unknown parameters.

In analyzing the factor equation $F_t = f(\xi_t) + \eta_t$, the effective sample size is even smaller, since the number of time periods in the sample is much smaller than the number of observed option returns. I therefore put a special emphasis is placed on minimizing the number of parameters in the approximation of the function $f(\xi_t)$. As suggested earlier, ξ_t will include all variables in ζ_{t-1} as well as variables that should help explain the unpredictable portion of the factor realization F_t. Less theory is available to help guide the approximation scheme for $f(\cdot)$, but several general comments may be made. First, variables such as the return on the S&P 500 probably impact F_t in an approximately linear way. Less obvious is the impact of a change in the volatility proxy, which possibly has an asymmetric relation with some unobserved volatility factor. To decrease the number of parameters that must be estimated while allowing for some nonlinearities in the relation between ξ_t and F_t, we will approximate $f(\xi_t)$ as a second-order Legendre series in ξ_t with all cross-terms set to zero.

Letting Z_t denote the $1 \times M$ vector of all terms in the Legendre expansion of $f(\xi_t)$, we therefore have the approximate factor equation

$$F_t \approx Z_t\,\mathbf{G} + \eta_t, \tag{17}$$

where $\mathbf{G}$ is a $M \times K$ parameter matrix.

4 Estimation

The model is estimated in a Bayesian framework using a Gibbs sampling approach similar to Geweke and Zhou (1996). In this section I summarize the model, introduce the data set, and describe the algorithm used to compute posterior distribution of all quantities of interest.

[4] Correlation between different terms of the expansion will be reduced if the underlying variable is close to uniform on $[-1, 1]$. To increase the uniformity over this interval, we allow a small number of terms to fall outside of it. In no way does this change the functional span of the approximation. The new rescaling scheme for a variable x is $x^* = 2(x - x_{1\%})/(x_{99\%} - x_{1\%}) - 1$, where $x_{1\%}$ and $x_{99\%}$ are the first and 99^{th} percentiles of the sample distribution of x.

4.1 The econometric specification / model summary

To summarize the previous section, the model considered in the remainder of the paper consists of a returns equation,

$$R_{it} = W_{it}a + X_{it}\mathbf{B}F_t' + \epsilon_{it}, \tag{18}$$

where

$$\epsilon_{it} \sim \mathrm{N}\left(0, \exp\left(X_{it}s\right)\right), \tag{19}$$

and a factor equation,

$$F_t = Z_t\,\mathbf{G} + \eta_t, \tag{20}$$

where

$$\eta_t \sim \mathrm{MVN}(0, \mathbf{I})\,. \tag{21}$$

4.2 The data

4.2.1 Option returns

To homogenize input and reduce the quantity of output, I report results only for put options. The data set consists of most returns between 1992 and 1996 realized from holding a put option from the close of one trading day to the close of the next trading day. Because of data problems identified in other studies, I only consider options with at least ten trading days prior to expiration.

The original data set was obtained from the Chicago Board Options Exchange and included all intraday quotes over the five year sample. A much smaller daily data set was constructed for each option contract by keeping the quote that was closest to but no later than 3:00pm. The resulting end-of-day data set consisted of 64,793 observations of option prices. About 550 observations were eliminated because they fell on days for which some information variable (such as the VIX index or the interest rate) was not observed. Four observations were eliminated for having Black-Scholes implied volatilities above 100%, and 2,834 were dropped for violating arbitrage bounds or for having a price below three eighths of a dollar. To eliminate potential data errors, 68 observations were eliminated because they represented large price reversals, either a 200% return followed by a -50% return or a -50% return followed by a 200% return.

Simple returns are calculated using the average of the bid and the ask prices, following Coval and Shumway (2000). Because many options were not observed on subsequent days, the set of returns data that were produced from these price observations is smaller, with 44,750 observations. Finally, 1,226 observations on 217 days were thrown out because fewer than ten returns were observed on that day.

The resulting data set therefore consists of a total of 43,524 returns observed on 1,019 days, and average of 42.7 returns per day. The composition of this sample, sorted by moneyness and time to expiration, is reported in Table 2. It is apparent that the greatest number of observations are for slightly out-of-the-money (OTM) puts. As time to expiration increases, returns are observed over a wider range of strikes, with deep OTM puts nonexistent for very short maturities. For all maturities, deep in-the-money (ITM) puts are relatively rare, reflecting the well-known illiquidity of the markets for these securities.

Finally, for the purpose of performing some out-of-sample diagnostics, the last six months of the sample were excluded from the estimation. A total of 32,228 option returns were therefore used in the estimation, with 11,296 returns in the 123-day “hold-out” sample.[5]

[5] It was somewhat surprising that one quarter of the return observations came from the last ten percent of the sample in calendar time, but this seems simply to be a reflection of the rapid growth in the S&P 500 options market.

4.2.2 Information variables

In addition to the returns data, the predetermined variables ζ_t and ξ_t must be specified. As mentioned above, ζ_t will consist solely of a proxy for lagged market volatility. Although it is likely that other state variables influence the mean and covariance structure of put option returns, none does so as clearly as market volatility. Another obvious variable for inclusion in ζ_t would be a short-term interest rate. Between 1992 and 1996, however, short rate volatility was extremely low, making the impact of this variable even lower than what other studies have argued is already low. I therefore omit this variable from ζ_t.

The market volatility proxy chosen is the Chicago Board Options Exchange Market Volatility Index (VIX), lagged one day. This index represents an average of eight Black-Scholes implied volatilities from options on the S&P 100 index. The construction of the VIX index is described in detail in Whaley (1993).[6]

The main advantages of the VIX index are its apparent quality and its observability in real time. Even at high frequencies one does not see the major reversals that would be evidence of errors in the data used to construct the index. Blair, Poon, and Taylor (1999) have found it to be a reliable indicator of future stock market volatility, generally outperforming measures of volatility based on past returns but essentially equivalent to a returns-based measure constructed from index returns at five minute intervals.

Use of the VIX could be criticized for several reasons. First, it is a measure of the volatility of the S&P 100, not the S&P 500. Second, as a measure of implied volatility, it does not necessarily coincide with true volatility under the "physical" or true probability measure. Finally, it is a hypothetical measure that does not generally equal the implied volatility of any particular option traded that day.

While these are valid criticisms, if the VIX index and the "true" S&P 500 volatility have relation that is approximately deterministic and not extremely nonlinear, then the Legendre expansions in the VIX proxy should be very similar to the expansions that would be generated were the true market volatility used instead. It seems plausible that the S&P 100 and S&P 500 market volatilities should not differ too much, and that S&P 100 implied volatilities are closely related to true market volatility, even if the relation is not quite linear or unbiased. I proceed with these caveats in place.

Finally, the variables in ξ_t must also be specified. Since ξ_t should contain variables useful for explaining factor risk premia, I again include a lagged value of the market volatility proxy, VIX. For explaining the unexpected component of F_t, I include the contemporaneous return on the S&P 500, the change in the logarithm of the VIX index, and the change in the logarithm of total daily volume on the NYSE index.

While the market return and the change in log volatility are natural candidates for explaining variation in options returns, the change in log volume requires some justification. Hong and Stein (1999) argue that there should be a relation between skewness and stock turnover, since differences of opinion among traders leads to more asymmetric returns. Chen, Hong, and Stein (2000) find support for the hypothesis in an analysis of individual stocks. Since the extreme case of negative skewness is a stock market crash, I use the change in trading volume as a variable to capture possibly time-varying "crash fears."

[6] The essential features can be described briefly. All eight options are near-term and near-the-money and are very actively-traded. Each option's implied volatility is computed using a binomial tree that is adjusted for time-varying dividends, changing interest rates, and the possibility of early exercise. The S&P 100 index level used in the calculation is perfectly contemporaneous, and to mitigate the problem that some prices underlying the index may be stale the implied volatilities of each call and its corresponding put are averaged. Put/call averages implied volatilities of different moneyness (all close to at-the-money) are then interpolated to produce the at-the-money implied volatility for each maturity. This term structure of implied volatilities is then interpolated to yield a hypothetical 22-day implied volatility.

4.3 Details of the posterior calculations

The unknown quantities of the model include the intercept parameters a, the slope parameters **B**, the residual variance parameters s, the factor sensitivity parameters **G**, and the unknown factors themselves, **F**. Posterior distributions of all these quantities are computed using the Gibbs sampler, in which we draw from the conditional distribution of each term given the remaining terms. Under general conditions these draws, performed in sequence, converge to the unknown posterior distribution.

Throughout this subsection, it will be more convenient to replace the double subscript "$i\ t$" (for security i at time t) with a single subscript $i = 1, ..., N$, where $N = 32228$ is the size of the entire sample. Therefore τ_i, κ_i, and all other i-subscripted symbols are to be interpreted as variables corresponding to return i. Less conventionally, F_i will refer to the row of **F** (a $T \times K$ matrix) that is contemporaneous with the observation of R_i, so F_i is *not* the i^{th} row of **F**.

4.3.1 Prior beliefs

The priors used in the analysis are completely flat, or exactly proportional to a constant. Since most analysis is based on linear regression theory, the distributions used to construct the Bayesian posterior will typically have the same form as classical sampling distributions. While it is not without controversy that flat priors represent prior ignorance, the extremely large size of the data set may make such issues unimportant.

4.3.2 Drawing a and B

Similarly to other factor model setups, there is a rotational indeterminacy between **B** and F_i' in the equation

$$R_i = W_i a + X_i \mathbf{B} F_i' + \epsilon_i, \tag{22}$$

since the combination of **B** and $\mathbf{F}'$ is indistinguishable from $\mathbf{BQ}'$ and $\mathbf{QF}'$ for all orthogonal $K \times K$ matrices **Q**.[7] I solve this problem as in Geweke and Zhou (1996) by imposing that the square submatrix consisting of the first K rows of **B** be lower triangular with positive diagonal elements. For a K factor model, therefore, $K(K-1)/2$ elements of **B** will be constrained to equal zero, and K elements will be constrained to be positive.

Using proposition 10.4 of Hamilton (1994), we may rewrite the return equation as

$$R_i = W_i a + (F_i \otimes X_i)\,\mathrm{Vec}\,(\mathbf{B}) + \epsilon_i, \tag{23}$$

where $(F_i \otimes X_i)$ denotes the $1 \times LK$ Kronecker product of F_i and X_i and $\mathrm{Vec}\,(\mathbf{B})$ is the $LK \times 1$ vector formed by stacking the K columns of **B**.

Let b denote the vector formed by stacking a on top of all elements of $\mathrm{Vec}\,(\mathbf{B})$ that are not constrained to equal zero by the identification assumption.[8] Also let **Y** denote the $N \times (LK - K(K-1)/2)$ matrix formed by stacking $(F_i \otimes X_i)$ for $i = 1, ..., N$, and then eliminating the columns that correspond to the zero elements of $\mathrm{Vec}\,(\mathbf{B})$. Now writing the equation in matrix form, we have

$$R = [\mathbf{W}\ \mathbf{Y}]\, b + \epsilon. \tag{24}$$

[7] The matrix must be orthogonal to be consistent with the subsequent identifying restriction that $\mathrm{Cov}(\eta_t', \eta_t) = I$. Recall also that for orthogonal matrices $\mathbf{Q}' = \mathbf{Q}^{-1}$.

[8] For the one factor model, no variable is constrained to equal zero, but the first element of $\mathrm{Vec}\,(\mathbf{B})$ must be positive. For a two-factor model, the $L+1$ element is removed and the first and $L+2$ elements are constrained to be positive. The three-factor model also removes the $2L+1$ and $2L+2$ elements and also requires the $2L+3$ element be positive, and so on.

Given $\mathbf{\Sigma}$, the diagonal matrix of idiosyncratic error variances, we would have the standard Bayesian GLS result under flat priors,

$$b|\mathbf{F}, s \sim \text{MVN}\left(\hat{b}, \mathbf{V}_b\right), \tag{25}$$

where

$$\hat{b} = \left([\mathbf{W}\ \mathbf{Y}]' \mathbf{\Sigma}^{-1} [\mathbf{W}\ \mathbf{Y}]\right)^{-1} [\mathbf{W}\ \mathbf{Y}]' \mathbf{\Sigma}^{-1} R$$

and

$$\mathbf{V}_b = \left([\mathbf{W}\ \mathbf{Y}]' \mathbf{\Sigma}^{-1} [\mathbf{W}\ \mathbf{Y}]\right)^{-1},$$

were all elements of the vector b unconstrained. Since K elements of b must be positive to satisfy the identification assumption, we must augment the draw of b from (25) with an accept/reject step in which draws that violate that assumption are discarded.

The parameters of interest, a and $\mathbf{B}$, are given by the appropriate partitions of b.

4.3.3 Drawing s

We can rewrite the return equation as

$$R_i = W_i a + X_i \mathbf{B} F_i' + \exp\left(X_i s\right) e_i, \tag{26}$$

where e_i is an i.i.d. standard normal random variable.

This approximately implies

$$\log\left(\left[R_i - W_i a - X_i \mathbf{B} F_i'\right]^2 + .001\right) = 2X_i s + e_i^*, \tag{27}$$

where $e_i^* = \log(e_i^2)$ and where the approximation arises from the extra .001 term appearing in the left hand side.[9] Given a, $\mathbf{B}$, and $\mathbf{F}$, this is a standard linear regression, with the complication that e_i^* is not Gaussian or even mean zero, although it is i.i.d.

An efficient and very accurate approximate solution for the posterior of s was proposed by Kim, Shepard, and Chib (1998, hereafter KSC). Their procedure consists of approximating the distribution of $e_i^* = \log(e_i^2)$ by a mixture of seven different normal distributions. The normal means, variances, and mixture probabilities calculated by KSC are reproduced in Table 3.

The procedure involves the introduction of a latent state variable π_i that takes on an integer value between one and seven indexing the normal distribution from which e_i^* is drawn. The KSC procedure therefore requires the addition of an additional "block" in the Gibbs sampler, the draw of the π_i, although this block is by itself of little interest.

Following KSC, we draw π_i for $i = 1, ..., N$ from the probability mass function

$$P\left(\pi_i = j | a, \mathbf{B}, \mathbf{F}, s\right) \propto q(j) \phi\left(\log\left(\left[R_i - W_i a - X_i \mathbf{B} F_i'\right]^2 + .001\right); 2X_i s + c(j), d(j)\right),$$

where $\phi(x; \mu, v)$ is the density function of a $N(\mu, v)$ random variable evaluated at x.

Conditional on π_i, we may view e_i^* as a normal random variable with mean $c(\pi_i)$ and variance $d(\pi_i)$. After subtracting this mean from each side of (27), we may proceed to draw s using Bayesian GLS. Letting C denote the column matrix formed from the N values of $c(\pi_i)$ and $\mathbf{D}$ the diagonal matrix formed from the corresponding values of $d(\pi_i)$, we have

$$s | a, \mathbf{B}, \mathbf{F} \sim N\left(\hat{s}, V_s\right), \tag{28}$$

[9] As discussed in Kim, Shepard, and Chib (1998), adding the offset .001 before taking logs makes the transformation more robust since it reduces the "inlier" problem associated with taking logarithms of very small numbers.

where

$$V_s = \left((2\mathbf{X})'\mathbf{D}^{-1}(2\mathbf{X})\right)^{-1}$$

and

$$\hat{s} = V_s\,(2\mathbf{X})'\,\mathbf{D}^{-1}\left(\log\left[(R - [\mathbf{W}\,\mathbf{Y}]b)^2 + .001\right] - C\right),$$

and where b and $\mathbf{Y}$ are defined as in (24).

4.3.4 Drawing F_t

For this draw I follow the derivations of Geweke and Zhou (1996), which necessitates the introduction of some new notation.

- N_t: the number of return observations at time t
- R^t: the $N_t \times 1$ vector of these returns,
- $\mathbf{\Sigma}^t$: the $N_t \times N_t$ diagonal matrix of all time t idiosyncratic error variances ($\exp(X_i s)$)
- ϵ^t: the $N_t \times 1$ vector of unobserved time t errors
- $\mathbf{X}^t$: the N_t rows of $\mathbf{X}$ corresponding to all time t observations

From the equation

$$F_t = Z_t\,\mathbf{G} + \eta_t$$

and the identifying restriction that $\mathrm{Cov}(\eta_t', \eta_t) = I$, we have, conditional on Z_t, X^t, $\mathbf{G}$, b, and s,

$$F_t' \sim \mathrm{MVN}((Z_t\,\mathbf{G})', \mathbf{I})\,. \tag{29}$$

From the equation

$$R^t = \beta^t F_t' + \epsilon^t,$$

where $\beta^t = X^t\mathbf{B}$, we have

$$R^t \sim \mathrm{MVN}\left(\beta^t\mathbf{G}'Z_t', \beta^t\beta^{t\,\prime} + \mathbf{\Sigma}^t\right). \tag{30}$$

Since the covariances of F_t and R^t are just equal to β^t, we have

$$\begin{bmatrix} F_t' \\ R^t \end{bmatrix} \sim \mathrm{MVN}\left(\begin{bmatrix} \mathbf{G}'Z_t' \\ \beta^t\mathbf{G}'Z_t' \end{bmatrix}, \begin{bmatrix} \mathbf{I} & \beta^{t\,\prime} \\ \beta^t & \beta^t\beta^{t\,\prime} + \mathbf{\Sigma}^t \end{bmatrix}\right) \tag{31}$$

Thus we have the conditional distribution

$$F_t'|R^t \sim \mathrm{MVN}\left(\mathbf{G}'Z_t' + \beta^{t\,\prime}\left(\beta^t\beta^{t\,\prime} + \mathbf{\Sigma}^t\right)^{-1}\left(R^t - \beta^t\mathbf{G}'Z_t'\right), \mathbf{I} - \beta^{t\,\prime}\left(\beta^t\beta^{t\,\prime} + \mathbf{\Sigma}^t\right)^{-1}\beta^t\right) \tag{32}$$

As Geweke and Zhou (1996) suggest, I use Woodbury's identity,

$$(\beta\beta' + \mathbf{\Sigma})^{-1} = \mathbf{\Sigma}^{-1} - \mathbf{\Sigma}^{-1}\beta\left(\mathbf{I} + \beta'\mathbf{\Sigma}^{-1}\beta\right)^{-1}\beta'\mathbf{\Sigma}^{-1} \tag{33}$$

to speed up the inversion of the $N_t \times N_t$ matrix $\beta^t\beta^{t\,\prime} + \mathbf{\Sigma}^t$, since the identity implies that only the the diagonal matrix $\mathbf{\Sigma}^t$ and the $K \times K$ matrix $\mathbf{I} + \beta^{t\,\prime}(\mathbf{\Sigma}^t)^{-1}\beta^t$ need be inverted.

4.3.5 Drawing G

The equation

$$\mathbf{F} = \mathbf{Z}\,\mathbf{G} + \eta \tag{34}$$

is a multivariate regression, since **F** is a $T \times K$ matrix. However, since $\text{Cov}(\eta_t) = I$, the K equations are independent, so we can use standard univariate methods to draw the K columns of **G** one at a time.

5 Estimation results

Altogether, eight different specifications are investigated. The first three are one, two, and three-factor versions of the approach described in section 4. In addition, I consider a number of models in which the factors are prespecified as some combination of the market return, the change in the log of VIX, the change in the log of NYSE trading volume, and the market return squared. When factors are prespecified, the econometric algorithm remains identical to before except that **F** is fixed and **G** need no longer be estimated.[10]

For each specification, a Gibbs sampler was run for 110,000 iterations, with the first 10,000 iterations discarded to negate the effects of initial conditions. The remaining 100,000 draws of the parameter vector are assumed to come from the posterior distribution and form the basis for all inferences drawn in the remainder of the paper.

5.1 Parameter estimates

Since the coefficients the Legendre polynomials are difficult to interpret, I do not report them in the paper. Instead, Table 4 gives a summary of the number of parameters estimated in each specification and the overall accuracy with which they are estimated. With one latent factor ($K = 1$), for example, there are 57 parameters estimated in the model. Out of the 57, 48 can be signed with at least 95 percent probability, meaning that for each of the 48 at least 95 percent of the marginal posterior distribution was on one side of zero. Finally, the median absolute "t-ratio", or posterior mean divided by posterior standard deviation, is equal to 4.44, indicating that most posterior means are far from zero.

In general, parameters of the $\alpha(\cdot)$ function are less accurately estimated than are parameters of the $\beta(\cdot)$ function, which are in turn less accurately estimated than are parameters of the $\sigma(\cdot)$ function. For the latent factor models, the parameters in the $f(\cdot)$ function that links observables to the latent factors tend to be estimated accurately as well.

The specifications in which factors are prespecified imply much sharper inferences about the $\alpha(\cdot)$ function. As in all empirical work, the greater accuracy resulting from stronger assumptions must be weighed against the possibility of model misspecification. Nevertheless, it is interesting that seven out of the ten parameters of $\alpha(\cdot)$ were signed with 95 percent probability when the factors were taken to be the market return and the change in log VIX, since $\alpha(\cdot) \neq 0$ is inconsistent with exact factor pricing. Since these two prespecified factors should be highly correlated with the two factors in a standard stochastic volatility model, this result suggests that such conventional models are probably incapable of pricing the cross section of index options. A more direct analysis of this claim is made below.

[10] Furthermore, when factors are prespecified the identification restrictions are unnecessary, making the **B** matrix unrestricted.

5.2 Interpreting the factors

In general, interpretation of latent factors must be performed with caution. In the present case, interpretation is facilitated by the assumed relation (8) between observables, such as the market return, and the latent factors. Assuming that the estimate of the **G** matrix of (17) is accurate enough, we can calculate the approximate correlation between the factors and observable variables such as the market return, the change in the logarithm of the VIX index, and the change in the logarithm of total NYSE volume. We therefore take **G** equal to its posterior mean and calculate correlations using the equation

$$\mathbf{F} = \mathbf{Z}\,\mathbf{G} + \eta \tag{35}$$

and the identification restriction that $\text{Cov}(\eta_t) = \mathbf{I}$.

The results, tabulated in Table 5, are not too surprising. In the single factor model ($K = 1$), the factor is positively correlated with market returns and is negatively correlated with changes in the VIX index, consistent with the fact that these two observables are strongly negatively correlated. Observables explain about 93 percent of the variation in the unobserved factor realizations.

For the two-factor case ($K = 2$), the link between volatility and underlying returns shocks is decoupled somewhat. Both factors remain highly correlated with the market return, but only one is highly correlated with changes in the log of the VIX index. As with the single factor, most factor variation is explained by observables, but a nontrivial part is not, with R-squares for both factors around .9.

Adding a third factor ($K = 3$) appears to introduce factor variation that is somewhat orthogonal to observables, as only 39 percent of the variation in the second factor is explained. One possible interpretation is that the second factor captures some aspect of option return dynamics that is outside the scope of what is usually considered relevant information for option pricing. Another interpretation is that adding a third factor amounts to model overfitting. We examine this second possibility indirectly through the out-of-sample hedging and portfolio allocation results.

5.3 Mispricing results

Of primary interest is whether the function $\alpha\,(\tau, \kappa, v_t)$ is reliably nonzero for any values of τ, κ, or v_t. Since Table 4 showed that many of the parameters a that enter this function could be signed with high posterior probability, it is natural to expect that the function $\alpha(\cdot)$ is nonzero as well. In this section I investigate the shape of $\alpha(\cdot)$, which reveals both the magnitude of the mispricing for different bonds and under different volatility states. Because the results are functional, a graphic approach to reporting results seems most efficient. To save space, I focus on put options of three maturities: one month, three months, and six months. The top rows of Figures 2, 3, and 4 plot the posterior mean of $\alpha(\cdot)$ as a function of moneyness and the current value of the VIX index for the models with one to three latent factors. Aso plotted in each graph is a coarse grid depicting the $\alpha(\cdot) = 0$ plane. The bottom rows gives the posterior probability that $\alpha(\cdot) > 0$ for each level of moneyness and VIX.

In every case, at least some put options have negative alphas with very high posterior probability, and no put option considered ever appears with high confidence to offer a positive α. Although the patterns of mispricing differ sizably across specifications, a few general characteristics are common. First, depending on the maturity and the specification, either short-term puts or out-of-the-money puts appear most mispriced, with the largest negative alphas. In the most severe cases, mispricing is largest for options that are both short term and OTM. Long-term options tend to have an α closer to zero, although it should be noted that short positions in these assets will earn this abnormal return for a longer period of time.

One observation common across all three maturities is that the addition of a second factor substantially reduces the magnitude of the mispricing. This is particularly true on days when the VIX index is low, for which the one factor models fares poorly. Nevertheless, there is reliable evidence that $\alpha(\cdot) < 0$ even for the two factor model, indicating that the introduction of a second priced risk factor, such as stochastic volatility, can reduce apparent mispricing but not eliminate it.

It is somewhat interesting that with two latent factors, the degree of mispricing is fairly constant over different moneyness and VIX levels, and, to a lesser extent, maturities. Since stochastic volatility models are the most popularly implemented form of two factor models, these results have implications for the shortcomings of stochastic volatility. Specifically, the "missing factor" in a stochastic volatility model must have similar implications for the expected returns of a wide variety of securities. This provides a basis on which to analyze extensions such as jumps, stochastic interest rates, stochastic dividends, and multi-factor volatility processes.

5.4 Comparison with Heston (1993)

Since the Heston (1993) model represents one of the most common examples of a two-factor option pricing model, a comparison of a few of its implications with those of the semiparametric analysis seems natural. Heston formulates a model in terms of a stock price and an instantaneous variance process as

$$\begin{aligned} dS_t &= \mu S_t dt + \sqrt{V_t} S_t dB_{1t} \\ dV_t &= \kappa(\theta - V_t)dt + \sigma\sqrt{V_t} dB_{2t}, \end{aligned}$$

where $\mathrm{Corr}(dB_{1t}, dB_{2t}) = \rho$. For pricing under the risk neutral distribution, μ is replaced by the instantaneous riskless rate r, κ is replaced by its risk-neutral counterpart $\kappa^* = \kappa - \lambda$, and θ is replaced by $\theta^* = \theta\kappa/\kappa^*$.

Using this model as the data generating process for a hypothetical stock index, I simulate 1,000 trading days of underlying prices and instantaneous variances.[11] Put options on this index are introduced to mimic, in a somewhat stylized manner, the way that the CBOE lists new options. Every 22 trading days, as the shortest maturity options are just expiring, ten 88-day index put options are introduced at evenly-spaced strike prices between 80% and 110% of the current level of the index. On every day until expiration, each option is priced exactly according to Heston's formula. As in the actual data sample, very short-term and very cheap options were excluded, and returns were calculated for each remaining option from observations on consecutive days. All days with fewer than 10 observed returns were eliminated. Finally, on each day a hypothetical 22-day option was priced, and it's Black-Scholes implied volatility was taken to be the value of a simulated VIX index.

To clarify the analysis, factors were assumed observable, and were taken to be the market return and the change in the logarithm of the VIX index. While it might be more natural to look at the change in V_t rather than a measure based on implied volatilities, focusing on VIX is consistent with the previous analysis.

The Gibbs sampling algorithm was implemented to generate posteriors of all model parameters. Since $\alpha(\cdot) = 0$ and $\sigma(\cdot) = 0$ in the simulations, the posterior means of the estimates of these functions are very close to zero. Slightly positive values of $\sigma(\cdot)$ result from small approximation errors as the orthogonal polynomials fail to fit exactly the factor sensitivities of the Heston model.

The top two panels of Figure 5 plot the $\beta_{RM}(\cdot)$ and $\beta_{VIX}(\cdot)$ functions, or the sensitivities of put returns to market returns and to changes in the log of VIX, for a 3-month (66-day) option of varying

[11] I assume annual parameter values of $\mu = .12$ and $r = .06$ for the stock price process. The parameter values for the instantaneous variance process, taken from Pan's (2000) table 1 (SV model), are $\kappa = 7.1$, $\theta = .0137$, $\sigma = .32$, $\rho = -.53$, and $\lambda = 7.6$.

moneyness and given different levels of the VIX index.[12] The large negative betas in the top left panel reflect the significant leverage represented by option positions, while the positive betas for the VIX factor show the familiar positive relation between volatility and options prices.

Contrasting the Heston model with the two factor latent variable model estimated in the paper would be difficult because of the rotational indeterminacy of the latent factor betas. I therefore use the somewhat less flexible specification in which $\alpha(\cdot)$, $\beta(\cdot)$, and $\sigma(\cdot)$ are modeled semiparametrically but the factors are assumed observable and given by the market return and the change in log VIX. The beta functions are computed for each draw from the posterior distribution, and the means of these draws are plotted in the middle two panels of Figure 5.

The posterior mean of the semiparametric market return beta is broadly similar in shape to Heston's, as both models imply more negative betas for OTM puts and on days with low volatility. Under the flexible specification, however, the sensitivity to market returns is generally greater, particularly in markets with low volatility. The lower left panel of Figure 5 contains the surface of the posterior t-ratio, or the absolute value of the posterior mean divided by the posterior standard deviation. In all cases this t-ratio is extremely large, indicating a high degree of estimation accuracy.

While sensitivity to market returns is greater under the flexible specification, the right middle panel reveals that sensitivity to changes in the VIX index is lower than the Heston model predicts. This is particularly true for deep OTM puts, whose VIX beta can be as little as a third of the values implied by the Heston model. In addition, for some deep ITM options the posterior mean beta drops below zero (represented by the coarse grid), perversely indicating a negatively relation between volatility and option prices. The low t-ratios in the bottom right panel indicate, however, that the support for this negativity is relatively weak.

The general similarity of the Heston model to the semiparametric specification implies that the Heston model betas therefore represent reasonable if imperfect approximations of the factor sensitivities of put returns. Since the Heston model is known to underprice OTM puts (see Pan, 2000, for instance), it is unsurprising that the more flexible two factor models, as shown in section 5.3, are also incapable of pricing them.

6 Hedging and portfolio allocation

6.1 Hedging

I first consider a hedging exercise that is designed to mimic the problem faced by an option writer who has sold puts of a variety of strikes and maturities and would like hedge these options by purchasing a small number of more liquid puts.

Options are divided into three categories of moneyness and time to expiration, forming total of nine classes of puts. The different moneyness categories are deep out-of-the-money (ratio of strike to spot between 0.84 and 0.92), moderately out-of-the-money (strike/spot between 0.92 and 1), and moderately in-the-money (strike/spot between 1 and 1.08). Deep in-the-money puts are much less frequently traded and are therefore ignored. The three maturity categories are short (10 to 44 business days to expiration), medium (45 to 100 days), and long (101 to 254 days). On each day, one option is chosen from each of the nine classes. When there are multiple options in a class, the option with moneyness closest to the class moneyness midpoint is chosen.

[12] It is actually the posterior means of $\beta_{RM}(\cdot)$ and $\beta_{VIX}(\cdot)$ that are graphed, but the lack of an explicit error term in the data generating process leads to extremely small posterior standard deviations. For $\beta_{RM}(\cdot)$, these standard deviations are no more than one percent of the posterior mean in absolute value, while for $\beta_{VIX}(\cdot)$ they are no more than four percent.

This "book" of options is hedged with three other puts, which are chosen from relatively liquid ranges of maturity and moneyness. The first is a short-term at-the-money put, which, among all puts with time to expiration between 10 and 44 days, is the one whose strike/spot ratio is closest to one. The second is a medium term option, chosen similarly. Last is a short-term put with moneyness ratio closest to 0.88. Puts that are part of the "book" are disqualified from being chosen as one of the three hedge assets.

The hedging strategy attempts to set the total portfolio betas to zero without concern for residual variance. Since a well-diversified options book may be significantly larger in realistic situations, the importance of idiosyncratic variation in option returns is likely to be overstated in this exercise. With three dissimilar hedge assets, hedge positions that set overall portfolio betas to zero may always be found. To reduce the likelihood that extreme positions are taken in the hedge portfolio, I require that no hedge position is larger than 500% (in absolute value) of the value of the book. When this constraint is binding, the hedge portfolio is chosen to minimize the sum of the squares of the K hedged portfolio betas.

When there are fewer than three factors, three hedge assets are still used, so that reductions in hedged volatility that result from using a greater number of factors can not be attributed solely to the use of a more diversified set of hedge assets. To find a unique hedge portfolio, some proportions among the three hedge assets are fixed. When one or two factors are assumed, the position in the short-term ATM option is required to equal the sum of the positions in the short-term OTM option and the medium-term ATM option. When just one factor is assumed, the short-term OTM option and the medium-term ATM option are further required to be held in identical amounts. (This implies that 25% of the hedge assets are in the short-term OTM option, 50% are in the short-term ATM option, and 25% are in the medium-term ATM option.)

For a comparison, I also consider hedging strategies based on the Black-Scholes "Greeks," where the first beta would be defined by the ratio of the Black-Scholes delta to the price of the option. Other betas are calculated as vega divided by price (measuring sensitivity to changing volatility) and gamma divided by price. All Greeks are calculated using the options' own implied volatilities.

Tables 6 and 7 display the standard deviation of the unhedged returns of the option book as well as the hedged returns using a variety of hedge strategies. The sizes of the samples used to calculate the volatilities are somewhat reduced since only days with three satisfactory hedge assets are considered. In addition, days with fewer than five of the nine put classes observed are discarded. The in-sample results in Table 6 are calculating using just 199 out of the 1113 days in the sample. Out-of-sample results in Table 7 rely on 107 out of the 123 trading days from July to December 1996. The large difference in the proportion of trading days with sufficient data reflects the dramatic growth of the S&P 500 options market over the period from 1992 to 1996.

In addition to reporting return standard errors, following the methods of Diebold and Mariano (1995) I report pairwise t-tests of the hypothesis that the mean squared hedging error of one strategy is equal to that of another strategy. Suppose $e_{1,t}$ and $e_{2,t}$ represent two time series of hedging errors and construct a time series of performance differentials as

$$de_t^2 = e_{1,t}^2 - e_{2,t}^2.$$

Using heteroskedasticity-consistent standard errors, I calculate robust t-statistics for the mean of de_t^2 for each pair of strategies. A negative t-statistic indicates the superiority of strategy 1, while a positive statistic supports strategy 2.

In-sample results in Table 6 show that the unhedged return standard deviation of roughly 16 percent per day could be reduced to less than three percent per day under a variety of strategies. The top performer turned out to be a simple Black-Scholes delta-matching strategy, but other strategies performed comparably. The single latent factor model (row B), for example, had a return standard deviation of just 2.78 percent per day. The t-statistic of 4.7 it received for its performance over the unhedged portfolio (row A) confirms, not surprisingly, the benefits of hedging. Interestingly, no strategy dominates this one, as evidenced by the fact that none of the t-statistics in column B are greater than +2. Many strategies are inferior, however. All three-factor models perform poorly, while strategies based on Black-Scholes Greeks other than delta are also inferior.

It is therefore a surprise when, in Table 7, the model with three latent factors is found to be the best performer out of sample. Results for different models, however, are similar, with the only poor performance registered by the Black-Scholes strategy that matches vega in addition to delta. Overall, the out-of-sample results indicate that most approaches can be used to dramatically reduce risk, but that the most flexible methods introduced in this paper are particularly useful.

Finally, the results should reduce the concern that the highly parameterized models estimated represent a serious overfitting of the data, since even the most general model performs extremely well out of sample.

6.2 Optimal option portfolios

Given previous pricing results, it is worthwhile considering how an investor might optimally form a portfolio of different put options to exploit their apparent mispricing, in particular the abnormal returns offered by selling puts. This section considers a type of mean/variance optimization that might be performed by an investor attempting to capture these returns by trading solely in put options of different strikes and maturities. While it is natural to consider strategies that involve trading in the underlying as well, these strategies are currently beyond the scope of the paper.

The econometric approach used in the paper does not require a full characterization of the risk/return structure of the factors driving put returns. The dynamics of the market return, for example, are not modeled. Since the market return is a component of the factors, the statistical properties of the factors are therefore unknown. It is therefore impossible to deduce the mean and variance of option returns, except conditional on the factor realizations. Since portfolio choices are made without conditioning on these realizations, a full mean/variance analysis is not possible.

The alternative portfolio choice problem that I consider corresponds to an investor who would like to form a portfolio that is completely insensitive to factor risk. This zero beta portfolio is subject only to options' idiosyncratic risks, and will earn a positive expected return only to the extent that options are mispriced. These strategy us therefore only appropriate in situations where the believed mispricing is nontrivial and idiosyncratic risk is believed highly diversifiable. In addition, since the zero beta requirement is a constraint, a true mean/variance optimizer should face investment opportunities superior to the ones identified here.

Finally, the portfolio chosen requires zero investment, and so contains long and short positions of offsetting value. This choice was made because most put options appear to represent poor investments for the buyer, so forcing the investor to hold a portfolio with a net positive value would probably be suboptimal.

Given $N \times 1$ vectors of α, β, and σ, the investor would minimize portfolio variance for a set level of expected return μ_0 by solving[13]

$$\min \quad w'(\beta\beta' + D)w \quad \text{s.t.} \quad w'\alpha = \mu_0, \; w'\iota = 0, \; w'\beta = [0, ..., 0],$$

where $D = \text{diag}(\sigma_1^2, ..., \sigma_N^2)$, $A = [\alpha, \iota, \beta]$, and ι is an $N \times 1$ vector of ones. It is easy to see that if w^* is the solution of this optimization problem for a level of expected return μ_0, then $2w^*$ will be

[13] Since the portfolio betas are required to equal zero, we may assume without loss of generality that the factor means are zero and the factor covariance matrix is the identity matrix.

the solution for expected return $2\mu_0$, and so on. In other words, w^* is homogeneous in the expected return μ_0. One implication of this homogeneity is that the Sharpe ratio is independent of μ_0.

The optimal weights turn out to be proportional to the first column of $D^{-1}A'\left(A'DA\right)^{-1}$, where the constant of proportionality depends on μ_0. In the applications below, the constant μ_0 is chosen to set the standard deviation of daily portfolio returns equal to five percent.

For comparison, I also report the performance of strategies that are based in part on the Black-Scholes Greeks. Since there are no $\alpha(\cdot)$ and $\sigma(\cdot)$ functions associated with these strategies, I use the functions estimated for the latent factor models. When matching deltas alone, I use the functions estimated for a single latent factor. When matching delta plus vega or gamma, I use the functions for two latent factors. When all three Greeks are matched I use the three latent factor estimates of $\alpha(\cdot)$ and $\sigma(\cdot)$.

Both in-sample and out-of-sample results are reported in Table 8. For comparison with the optimal portfolios, an equal-weighted portfolio of short puts is also considered. Under every strategy, average excess returns were positive. In sample, return volatilities are fairly close to their target values of five percent per day. Some but not all of the average returns in sample are extremely high, with very large heteroskedasticity-robust t-statistics. Invariably in sample, the best performers are the hybrid strategies that use the $\alpha(\cdot)$ and $\sigma(\cdot)$ functions estimated for the latent factor models but that set Black-Scholes Greeks to zero rather than the latent factor model's own betas. The Sharpe ratios associated with these strategies are many times higher than Sharpe ratios obtainable by investing in the market index or by selling an equal-weighted basket of puts.

Out of sample, Sharpe ratios remain high, with the best performer being the three-factor Black-Scholes hybrid strategy. Very good performance was also returned by models not based on Black-Scholes. The two factor latent variable model, for instance, led to a strategy with a positive excess return of almost 2.5 percent *per day*. Other strategies were less impressive but still statistically significant.

One out of sample failure of the model is the higher level of realized return volatility, which had a target value of five percent. Nevertheless, for many strategies the average returns were high enough to result in extremely favorable risk-adjusted performance.

An important caveat to these results is that they are calculated, similarly to Coval and Shumway (2000), without taking transactions costs into consideration, which Phillips and Smith (1980) note can be substantial. It is therefore important that large Sharpe ratios are not taken as evidence of market inefficiency. Instead, these Sharpe ratios simply express the large deviations from exact factor pricing in a familiar and economically meaningful way.

7 Conclusion

The extremely large class of factor models considered in this paper appears to offer no rational explanation for the abnormal excess returns that appear possible in the market for S&P 500 index put options. While allowing for more than one priced factor did reduce the degree of mispricing for many options, two priced factors were insufficient to explain the abnormally negative returns on a wide range of put options. Somewhat surprisingly, adding a third factor appeared to make the mispricing worse.

In general, the different specifications led to varied implications about what options are most severely mispriced, but one conclusion that was common across specifications is that short-term out-of-the-money puts are overvalued.

Besides revealing large deviations from factor pricing, the usefulness of the semiparametric latent factor approach in hedging was demonstrated both in and out of sample. It was shown that most of the volatility in an options book could be eliminated by trading in just a few additional puts. The models also performed well in finding portfolios to exploit the apparent mispricing, resulting in Sharpe ratios that were often many times larger than those obtainable with more conventional investments, also both in and out of sample.

The results on hedging and portfolio optimization also gave reassurance that overfitting was not a major problem in the models estimated and provided good evidence that the fitted models represent reliable descriptions of the data generating process, even out of sample. In fact, the most flexible three factor models often had the best post-sample performance, indicating that departures from traditional two factor stochastic volatility models (with one returns factor and one volatility factor) may be fruitful.

Overall, the results suggest that a rational, risk-based explanation of expected options returns might not exist. If rational asset pricing is to be preserved, it might rely less on the representative agent paradigm and more on the presumably much different characteristics and possibly beliefs of option writers and option buyers.

References

[1] Aït-Sahalia, Y., and A. W. Lo, 1998, "Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices," *Journal of Finance* 53, 499-458.

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

[3] Bakshi, G., C. Cao, and Z. Chen, 2000, "Do Call Prices and the Underlying Stock Always Move in the Same Direction?" *Review of Financial Studies* 13, 549-584.

[4] Bakshi, G., and N. Kapadia, 2000, "Delta Hedged Gains and the Pricing of Volatility Risk," working paper.

[5] Bates, D. S., 2000, "Post-'87 Crash Fears in the S&P 500 Futures Option Market," *Journal of Econometrics* 94, 181-238.

[6] Bauer, G., and A. Tamayo, 2000, "The Cross-Section of Expected Returns with Imperfectly Observed Factors," working paper.

[7] Benzoni, L., 1999, "Pricing Options Under Stochastic Volatility: An Econometric Analysis," working paper.

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

[9] Blair, B., S. Poon, and S. J. Taylor, 1999, "Forecasting S&P 100 Volatility: The Incremental Information Content of Implied Volatilities and High Frequency Index Returns," working paper.

[10] Broadie, M., J. Detemple, E. Ghysels, and O. Torrès, 2000, "American Options with Stochastic Dividends and Volatility: A Nonparametric Investigation," *Journal of Econometrics* 94, 53-92.

[11] Buraschi, A., and J. C. Jackwerth, 2000, "The Price of a Smile: Hedging and Spanning in Option Markets," *Review of Financial Studies*, forthcoming.

[12] Chen, N., R. Roll, and S. A. Ross, 1986, "Economic Forces and the Stock Market," *Journal of Business* 59, 383-403.

[13] Chen, J., H. Hong, and J. Stein, 2000„ "Forecasting Crashes: Trading Volume, Past Returns, and Conditional Skewness in Stock Prices ," *Journal of Financial Economics*, forthcoming.

[14] Chernov, M., and E. Ghysels, 2000, "A Study towards a Unified Approach to the Joint Estimation of Objective and Risk Neutral Measures for the Purpose of Options Valuation," *Journal of Financial Economics*, 56, 407-458.

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

[16] Christoffersen, P., and J. Hahn, 1999, "Nonparametric Testing of ARCH for Option Pricing," *Proceedings of the Sixth International Conference on Computational Finance*, Y. S. Abu-Mostafa, B. LeBaron, A. W. Lo, and A. S. Weigend (eds.), MIT Press, Cambridge.

[17] Connor, G., and R. A. Korajczyk, 1986, "Performance Measurement with the Arbitrage Pricing Theory," *Journal of Financial Economics* 15, 373-394.

[18] Coval, J., and T. Shumway, 2000, "Expected Options Returns," *Journal of Finance*, forthcoming.

[19] Day, T., and C. Lewis, 1992, "Stock Market Volatility and the Information Content of Stock Index Options," *Journal of Econometrics* 52, 267-287.

[20] Diebold, F. X., and R. S. Mariano, 1995, "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics* 13, 253-263.

[21] Fama, E. F., and K. R. French, 1993, "Common Risk Factors in the Returns of Bonds and Stocks," *Journal of Financial Economics* 33, 3-53.

[22] Fleming, J., B. Ostdiek, and R. E. Whaley, 1995, "Predicting Stock Market Volatility: A New Measure," *Journal of Futures Markets* 15, 265-302.

[23] Garcia, R., and R. Gençay, 2000, "Pricing and Hedging Derivative Securities with Neural Networks and a Homogeneity Hint," *Journal of Econometrics* 94, 93-115.

[24] Geweke, J., and G. Zhou, 1996, "Measuring the Pricing Error of the Arbitrage Pricing Theory," *Review of Financial Studies* 9, 557-587.

[25] Hamilton, J. D., 1994, *Time Series Analysis*, Princeton University Press, Princeton.

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

[27] Hong, H., and J. Stein, 2000, "Differences of Opinion, Rational Arbitrage, and Market Crashes," working paper.

[28] Hutchinson, J. M., A. W. Lo, and T. Poggio, 1994, "A Nonparametric Approach to Pricing and Hedging Derivative Securities Via Learning Networks," *Journal of Finance* 49, 851-890.

[29] Jackwerth, J. C., 2000, "Recovering Risk Aversion from Option Prices and Realized Returns," *Review of Financial Studies* 13, 433-451.

[30] Jones, C. S., 2000, "The Dynamics of Stochastic Volatility: Evidence from Options and Underlying Markets," working paper.

[31] Judd, K. L., 1998, *Numerical Methods in Economics*, MIT Press, Cambridge.

[32] Kim, Shepard, and Chib, 1998, "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models," *Review of Economic Studies* 65 361-393.

[33] Pan, J., 2000, "'Integrated' Time-Series Analysis of Spot and Options Prices," working paper.

[34] Phillips, S. M., and C. W. Smith, Jr., 1980, "Trading Costs for Listed Options: The Implications for Market Efficiency," *Journal of Financial Economics* 8, 179-201.

[35] Whaley, R. E., 1993, "Derivatives on Market Volatility: Hedging Tools Long Overdue," *Journal of Derivatives* 1, 71-84.

Table 1: Approximation Goodness of Fits

$\beta(\cdot)$	Approximation R^2 with Untransformed Variables	Approximation R^2 with Transformed Variables
	Panel A: Black-Scholes model	
delta / price	.8917	.9999
vega / price	.9000	.9998
rho / price	.9998	.9996
theta / price	.6022	.9917
	Panel B: Heston model	
delta / price	.9057	.9941
vega / price	.8913	.9871
rho / price	.9971	.9978
theta / price	.7067	.9850

Table 2: Sample Size by Time to Expiration and Moneyness

	$10 \leq \tau \leq 22$	$22 < \tau \leq 44$	$44 < \tau \leq 66$	$66 < \tau \leq 128$	$128 < \tau$
$\kappa \leq .80$	0	9	49	308	870
$.80 < \kappa \leq .90$	99	826	994	1832	1876
$.90 < \kappa \leq 1.00$	3623	6072	4124	4522	3766
$1.00 < \kappa \leq 1.10$	2263	3720	2580	2413	1953
$1.10 < \kappa$	25	34	25	101	207

Table 3: Mixing Approximation of log (e_i^2)

State j	Probability $q(j)$	Mean $c(j)$	Variance $d(j)$
1	0.00730	-11.40039	5.79596
2	0.10556	-5.24321	2.61369
3	0.00002	-9.83726	5.17950
4	0.04395	1.50746	0.16735
5	0.34001	-0.65098	0.64009
6	0.24566	0.52478	0.34023
7	0.25750	-2.35859	1.26261

Table 4
Parameter Posterior Summary

	Latent Factors			*Prespecified Factors*				
	$K = 1$	$K = 2$	$K = 3$	R_M	$R_M + \Delta\ln(\text{VIX})$	$R_M + \Delta\ln(\text{Vol})$	$R_M + \Delta\ln(\text{VIX}) + \Delta\ln(\text{Vol})$	$R_M + \Delta\ln(\text{VIX}) + R_M^2$
# of total parameters	57	84	110	48	67	67	86	86
# signed with 95% prob	48	63	84	39	46	58	61	67
Median absolute t-ratio	4.43	4.04	2.94	6.21	5.07	7.98	4.20	5.30
# of α parameters	10	10	10	10	10	10	10	10
# signed with 95% prob	6	4	4	9	7	9	8	8
Median absolute t-ratio	2.92	1.34	0.97	5.17	6.33	11.77	6.32	7.09
# of β parameters	19	37	54	19	38	38	57	57
# signed with 95% prob	19	28	43	18	27	35	40	45
Median absolute t-ratio	9.29	5.63	3.29	13.00	5.35	9.98	4.11	5.51
# of σ parameters	19	19	19	19	19	19	19	19
# signed with 95% prob	16	15	15	12	12	14	13	14
Median absolute t-ratio	4.35	4.12	3.48	2.86	3.93	3.96	3.02	3.57
# of f parameters	9	18	27					
# signed with 95% prob	7	16	22					
Median absolute t-ratio	6.21	5.02	3.04					

Table 5
Relations Between Latent Factors and Observables

Estimated factor		Correlation with R_M	$\Delta \ln(\text{VIX})$	$\Delta \ln(\text{Vol})$	Regression R^2
$K = 1$	first factor	0.93	-0.71	-0.02	0.93
$K = 2$	first factor	0.81	-0.78	-0.10	0.91
	second factor	-0.89	0.38	-0.14	0.89
$K = 3$	first factor	0.83	-0.78	-0.09	0.91
	second factor	-0.05	-0.35	-0.21	0.39
	third factor	0.92	-0.44	0.11	0.90

Table 6
Hedging In Sample

		Standard Deviation	Robust T-Statistic for Hedging Improvement over Alternative Strategy											
			A	B	C	D	E	F	G	H	I	J	K	L
A	Unhedged Book Return	0.1627												
				Hedged Returns										
B	Flexible / One Latent Factor	0.0278	4.70											
C	Flexible / R_M	0.0284	4.69	-1.78										
D	B-S Delta	0.0269	4.70	1.42	2.17									
E	Flexible / Two Latent Factors	0.0276	4.72	0.20	0.61	-0.40								
F	Flexible / $R_M + \Delta \ln(\text{VIX})$	0.0273	4.71	0.65	1.30	-0.26	0.41							
G	Flexible / $R_M + \Delta \ln(\text{Vol})$	0.0271	4.72	0.74	1.28	-0.16	0.64	0.38						
H	B-S Delta + Vega	0.0681	3.94	-5.81	-5.74	-5.87	-5.65	-5.70	-5.72					
I	B-S Delta + Gamma	0.0361	4.61	-2.34	-2.18	-2.54	-2.53	-2.55	-2.60	4.70				
J	Flexible / Three Latent Factors	0.0398	4.54	-2.70	-2.58	-2.87	-2.74	-2.78	-2.79	4.19	-0.76			
K	Flexible / $R_M + \Delta \ln(\text{VIX}) + \Delta \ln(\text{Vol})$	0.0633	4.01	-2.60	-2.57	-2.63	-2.61	-2.61	-2.63	0.49	-2.15	-2.38		
L	Flexible / $R_M + \Delta \ln(\text{VIX}) + R_M^2$	0.0453	4.45	-2.67	-2.60	-2.79	-2.71	-2.71	-2.75	3.14	-1.40	-0.84	1.47	
M	B-S Delta + Vega + Gamma	0.0421	4.50	-5.02	-4.76	-5.38	-4.98	-4.89	-4.93	4.74	-1.71	-0.84	1.90	0.53

A positive t-statistic in row H, column C, for example, indicates the superiority of strategy H over strategy C.

Table 7
Hedging Out of Sample

		Standard Deviation	Robust T-Statistic for Hedging Improvement over Alternative Strategy											
			A	B	C	D	E	F	G	H	I	J	K	L
A	Unhedged Book Return	0.1292												
			Hedged Returns											
B	Flexible / One Latent Factor	0.0316	7.16											
C	Flexible / R_M	0.0325	7.14	-2.66										
D	B-S Delta	0.0305	7.16	1.26	2.17									
E	Flexible / Two Latent Factors	0.0299	7.22	2.44	3.43	1.05								
F	Flexible / $R_M + \Delta\ln(\text{VIX})$	0.0303	7.23	1.74	2.94	0.48	-0.77							
G	Flexible / $R_M + \Delta\ln(\text{Vol})$	0.0294	7.26	2.77	3.86	1.14	0.54	1.62						
H	B-S Delta + Vega	0.0517	6.12	-4.63	-4.53	-4.91	-4.65	-4.62	-4.68					
I	B-S Delta + Gamma	0.0350	7.06	-1.04	-0.73	-1.42	-1.83	-1.63	-1.87	3.54				
J	Flexible / Three Latent Factors	0.0284	7.31	1.29	1.60	0.88	0.59	0.73	0.50	4.30	1.87			
K	Flexible / $R_M + \Delta\ln(\text{VIX}) + \Delta\ln(\text{Vol})$	0.0286	7.28	1.78	2.20	1.21	0.79	0.97	0.64	4.49	2.10	-0.20		
L	Flexible / $R_M + \Delta\ln(\text{VIX}) + R_M^2$	0.0320	7.19	-0.07	0.16	-0.31	-0.55	-0.45	-0.65	3.49	0.64	-1.47	-1.08	
M	B-S Delta + Vega + Gamma	0.0333	7.16	-0.84	-0.37	-1.41	-1.72	-1.48	-1.97	4.21	0.43	-3.77	-3.58	-0.46

A positive t-statistic in row H, column C, for example, indicates the superiority of strategy H over strategy C.

Table 8
Excess Returns of Zero Beta Portfolios with Maximum Sharpe Ratios

Strategy	*In Sample: Jan. 1992 - June 1996* Mean Return	Standard Deviation	Sharpe Ratio	Robust T-Stat	*Out of Sample: July 1996 - Dec. 1996* Mean Return	Standard Deviation	Sharpe Ratio	Robust T-Stat
Flexible / One Latent Factor	0.0123	0.0709	0.1738	5.20	0.0078	0.0759	0.1029	1.15
Flexible / R_M	0.0081	0.0521	0.1554	4.65	0.0194	0.0800	0.2429	2.70
B-S Delta	0.0295	0.0739	0.3993	11.96	0.0130	0.0911	0.1424	1.59
Flexible / Two Latent Factors	0.0042	0.0560	0.0745	2.23	0.0249	0.0937	0.2654	2.96
Flexible / $R_M + \Delta\ln(\text{VIX})$	0.0045	0.0418	0.1087	3.25	0.0197	0.0810	0.2428	2.70
Flexible / $R_M + \Delta\ln(\text{Vol})$	0.0033	0.0396	0.0835	2.50	0.0135	0.0621	0.2179	2.43
B-S Delta + Vega	0.0286	0.0529	0.5403	16.18	0.0213	0.0793	0.2652	2.98
B-S Delta + Gamma	0.0294	0.0639	0.4598	13.77	0.0249	0.0938	0.2652	2.95
Flexible / Three Latent Factors	0.0035	0.0476	0.0737	2.21	0.0104	0.0816	0.1276	1.42
Flexible / $R_M + \Delta\ln(\text{VIX}) + \Delta\ln(\text{Vol})$	0.0036	0.0386	0.0935	2.80	0.0174	0.0624	0.2786	3.10
Flexible / $R_M + \Delta\ln(\text{VIX}) + R_M^2$	0.0015	0.0353	0.0417	1.25	0.0059	0.0923	0.0639	0.71
B-S Delta + Vega + Gamma	0.0284	0.0611	0.4645	13.91	0.0368	0.1062	0.3465	3.86
Equal Weighted Short Put Return	0.0104	0.1798	0.0576	1.73	0.0102	0.1782	0.0573	0.64

Table 9
Sharpe Ratios of the S&P 500 Index

In Sample: Jan. 1992 - June 1996	0.0680
Out of Sample: July 1996 - Dec. 1996	0.1020
Extended Sample: Jan. 1963 - Dec. 1999	0.0379

Figure 1: Approximation Accuracy of the Legendre Polynomials

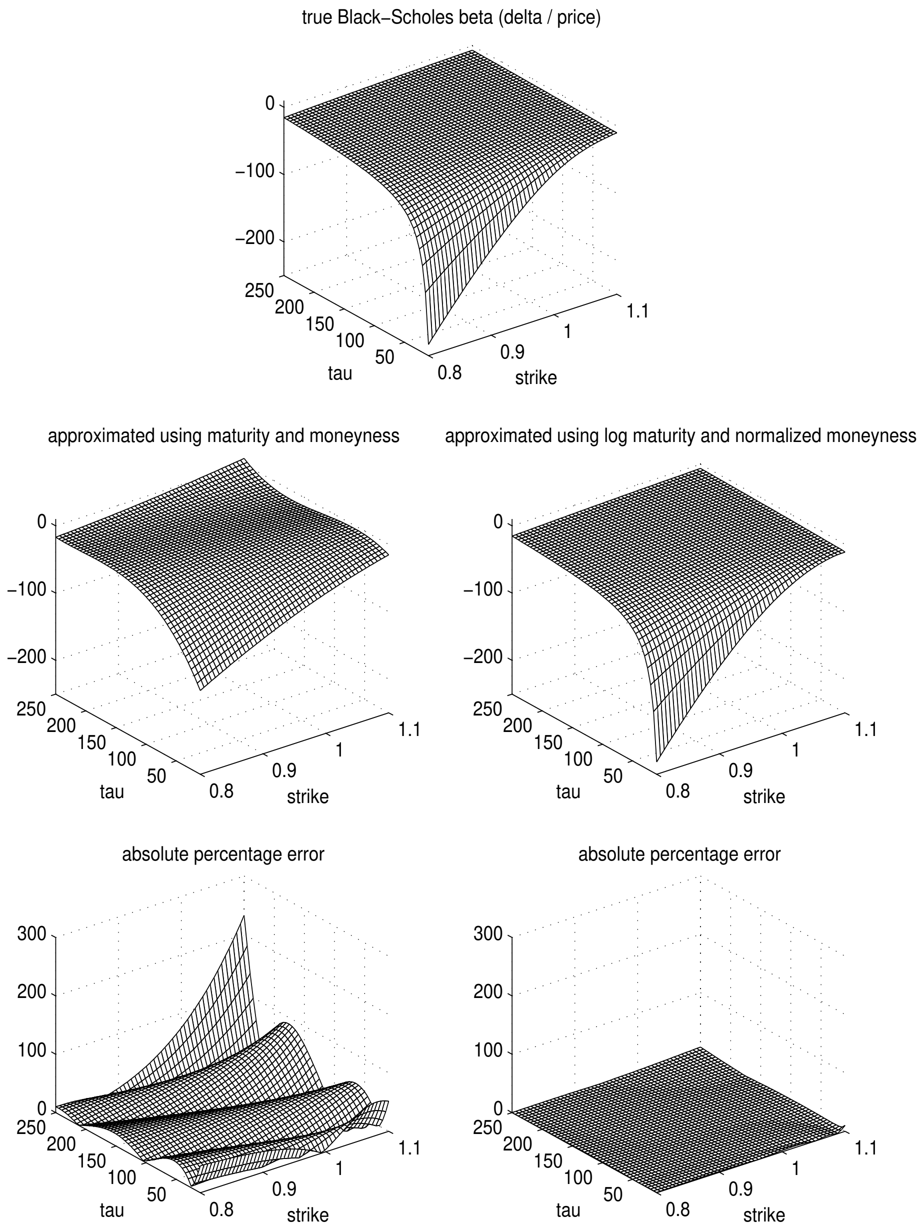


Figure 2: Mispricing (α) of 1-Month Options

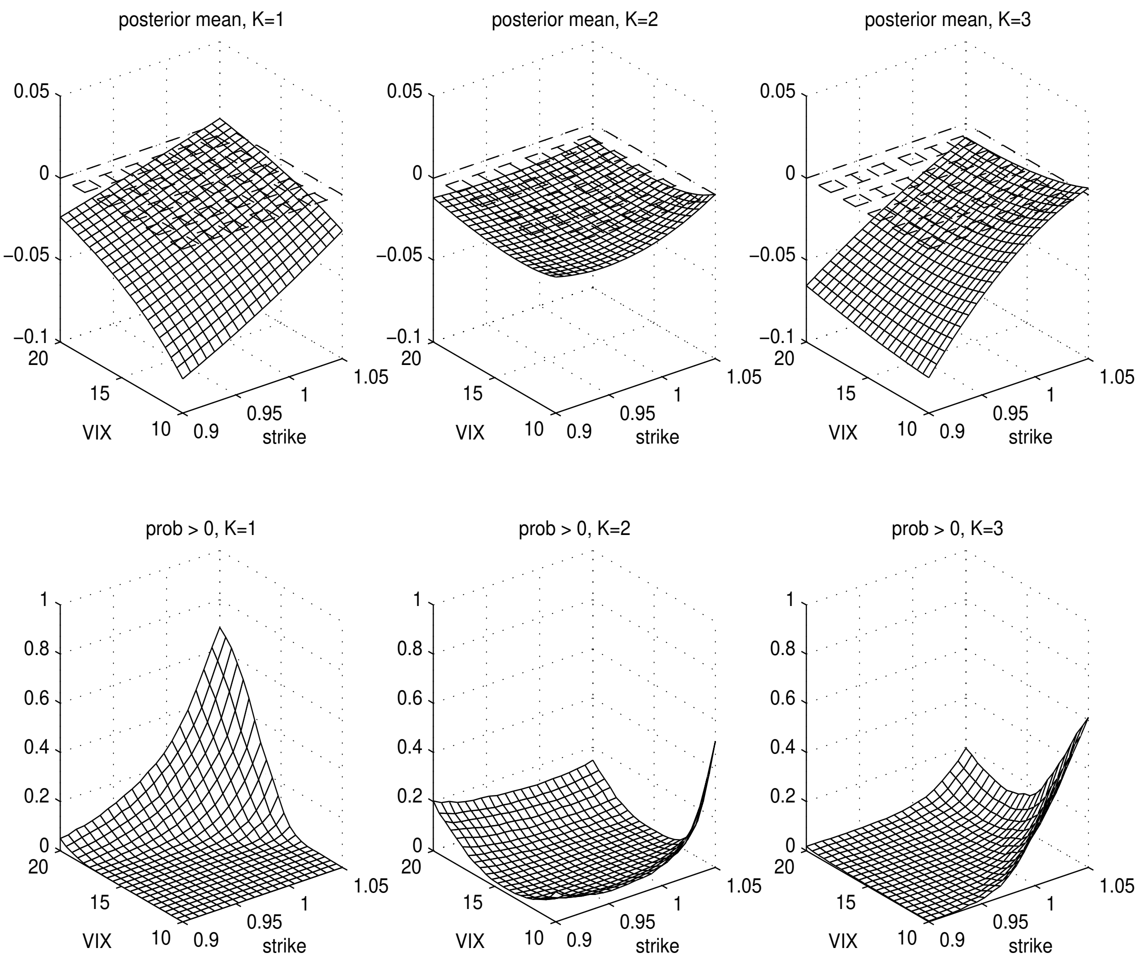


Figure 3: Mispricing (α) of 3-Month Options

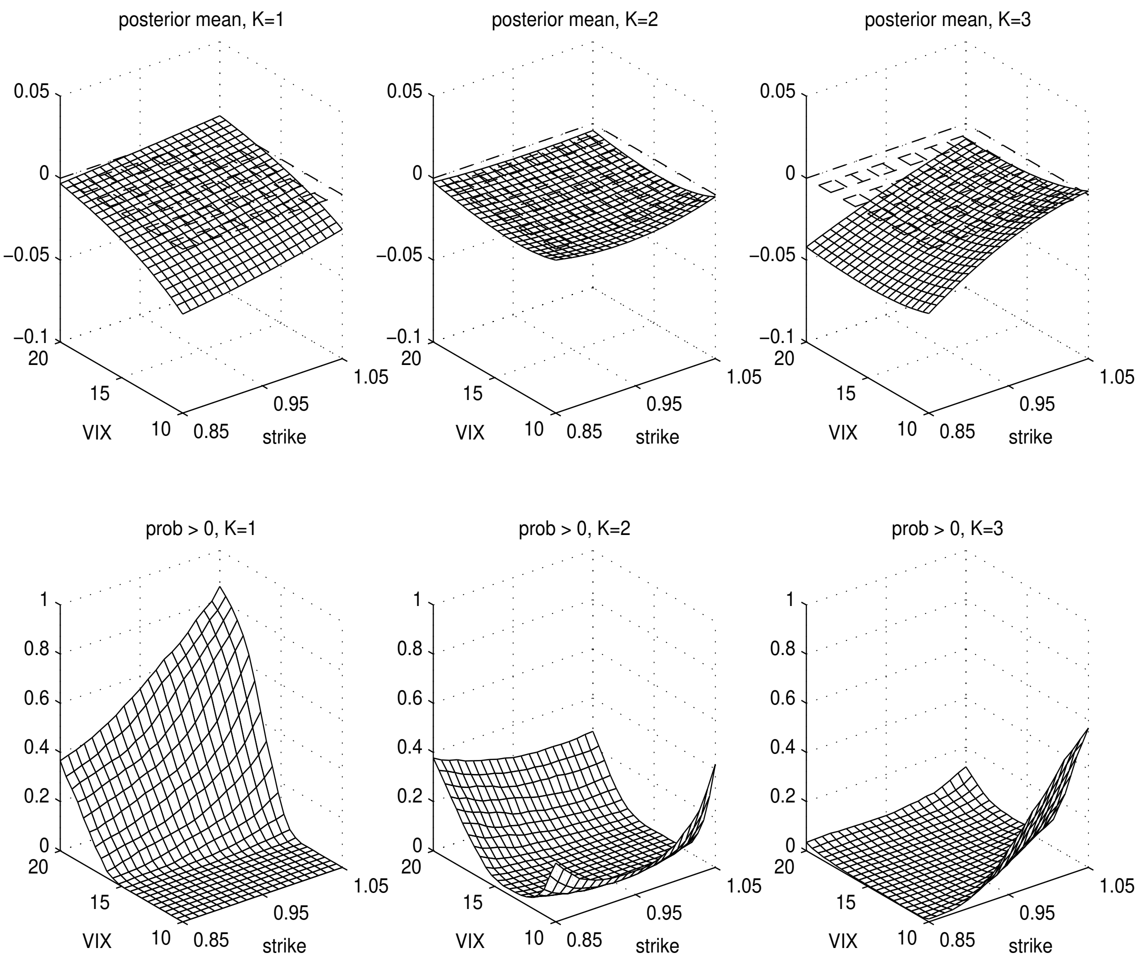


Figure 4: Mispricing (α) of 6-Month Options

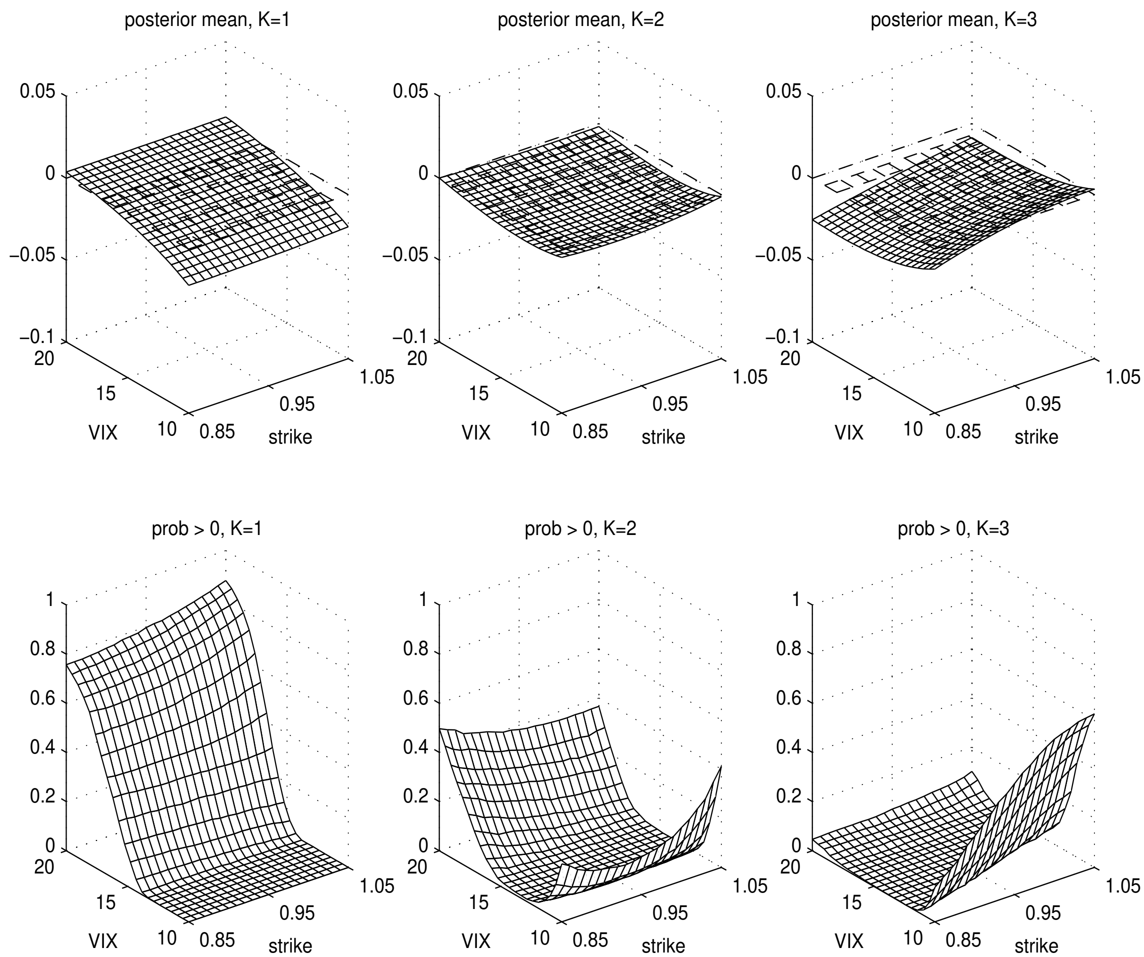
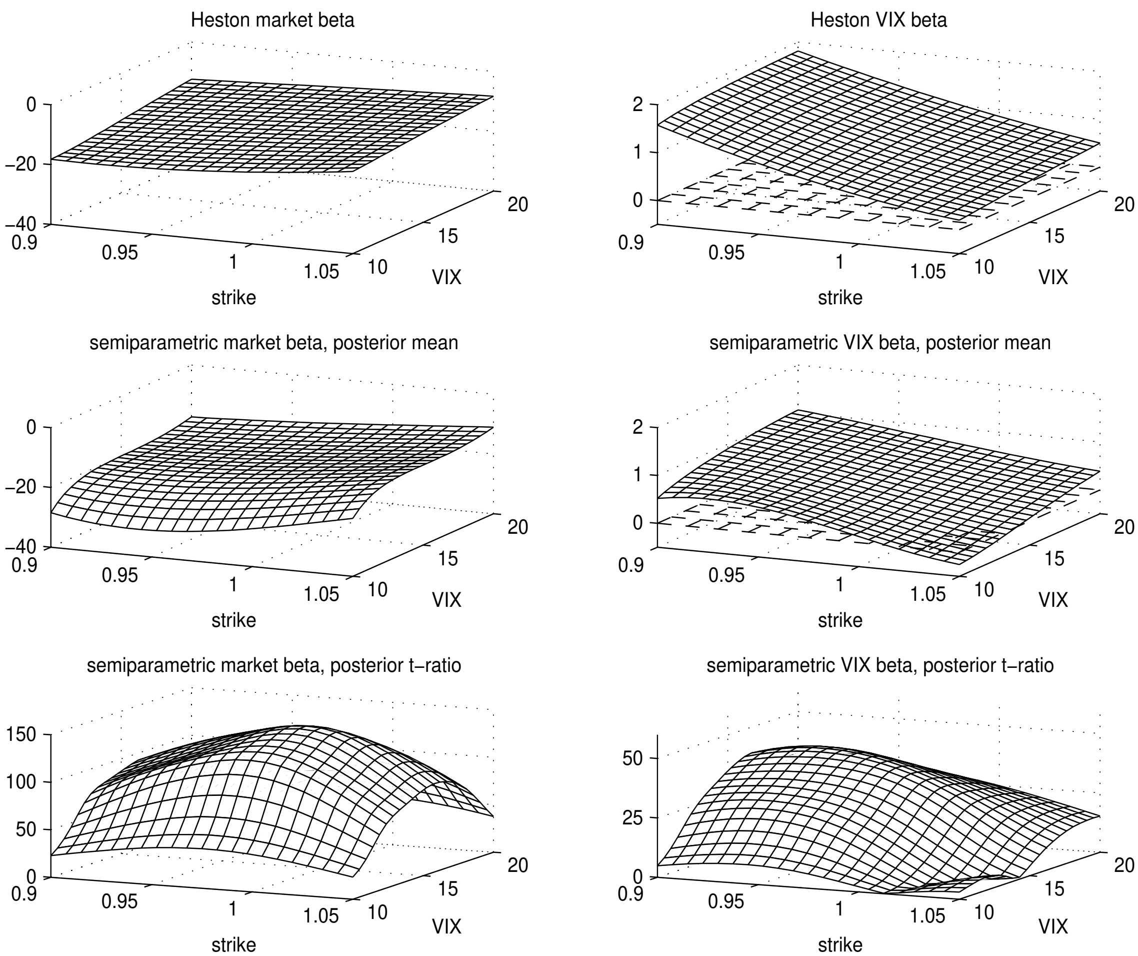


Figure 5: Factor Betas of 3-Month Options



A Notation summary

- a – the $H \times 1$ vector of parameters used in approximating $\alpha(\cdot)$
- b – the stacking of a and the nonzero elements of Vec $(\mathbf{B})$
- $\mathbf{B}$ – the $M \times K$ matrix of parameters used in approximating $\beta(\cdot)$
- $c(j)$ – the mean of the j^{th} normal in the mixture distribution of e_i^*
- C – the vector of means of e_i^* given the vector of π_i
- $d(j)$ – the variance of the j^{th} normal in the mixture distribution of e_i^*
- $\mathbf{D}$ – the diagonal matrix of variances of e_i^* given the vector of π_i
- e_i – the standard normal shock to return i
- e_i^* – the transformed residual $\log(e_i^2)$
- E_t – the time t vector of systematic random shocks
- $\mathbf{F}$ – the $T \times K$ matrix of factor realizations
- F_t – the $1 \times K$ vector of time t factor realizations
- F_i – the $1 \times K$ vector of factors pertaining to return R_i
- $\mathbf{G}$ – a $M \times K$ matrix of parameters used in approximating $f(\cdot)$
- H – the number of terms in the Legendre expansion of $\alpha(\cdot)$
- J – the size of ζ_t
- K – the number of factors
- L – the number of terms in the Legendre expansions of $\sigma(\cdot)$ and each dimension of $\beta(\cdot)$
- M – the number of terms in the Legendre expansion of $f(\cdot)$
- N – the total number of observed returns
- N_t – the number of returns observed at time t
- $P_n(\cdot)$ – the n^{th} order Legendre polynomial function
- $q(j)$ – the probability of the j^{th} normal in the mixture distribution of e_i^*
- R – the $N \times 1$ vector of all returns
- R^t – the $N_t \times 1$ vector of time t returns
- R_i – the i^{th} return (out of N)
- S_t – the K-dimensional diffusion process generating all systematic risk
- T – the number of days in the sample
- $\hat{v}_t$ – the time t volatility proxy (the VIX Index)
- v – a generic symbol for volatility
- V – a generic symbol for variance
- $\mathbf{W}$ – a $N \times H$ matrix of Legendre series terms for approximating $\alpha(\cdot)$
- $\mathbf{X}$ – a $N \times L$ matrix of Legendre series terms for approximating $\sigma(\cdot)$ and $\beta(\cdot)$
- $\mathbf{X}^t$ – the N_t rows of $\mathbf{X}$ corresponding to all time t returns
- X_i – the $1 \times L$ vector of expansion variables pertaining to return observation i
- $\mathbf{Z}$ – a $T \times M$ matrix of Legendre series terms for approximating $f(\cdot)$
- $\alpha(\cdot)$ – the return intercept function
- $\beta(\cdot)$ – the factor loading function
- ϵ – the $N \times 1$ vector of return equation residuals
- ϵ^t – the $N_t \times 1$ vector of time t return equation residuals
- ϵ_i – the residual of return R_i
- η – a $T \times K$ matrix of factor equation residuals
- η_t – a $1 \times K$ vector of time t factor equation residuals
- κ_i – the lagged moneyness (PV(strike)/price) corresponding to return observation i
- $\sigma(\cdot)$ – the residual standard deviation function
- $\mathbf{\Sigma}$ – the $N \times N$ diagonal matrix of Cov(ϵ)
- $\mathbf{\Sigma}^t$ – the $N_t \times N_t$ diagonal matrix of Cov(ϵ^t)
- τ_i – the lagged time to maturity corresponding to return observation i
- ξ_t – the vector of information variables that enter $f(\cdot)$,
- ζ_t – the $1 \times J$ vector of information variables that enter $\alpha(\cdot)$, $\beta(\cdot)$, and $\sigma(\cdot)$