Option Prices and Implied Volatility Dynamics under Bayesian Learning*

Massimo Guidolin Allan Timmermann
University of California, San Diego

June 29, 2000

Abstract

This paper shows that many of the empirical biases of the Black and Scholes option pricing model can be explained by Bayesian learning effects. In the context of an equilibrium model where dividend news evolve on a binomial lattice with unknown but recursively updated probabilities we derive closed-form pricing formulas for European options. Learning is found to generate asymmetric skews in the implied volatility surface and systematic patterns in the term structure of option prices. Data on S&P 500 index option prices is used to back out the parameters of the underlying learning process and to predict the evolution in the cross-section of option prices. The proposed model leads to lower out-of-sample forecast errors and smaller hedging errors than a variety of alternative option pricing models, including Black-Scholes.

*We wish to thank Alexander David, Jose Campa, Bernard Dumas, Claudio Michelacci, Enrique Sentana and seminar participants at Bocconi University, CEMFI, Federal Reserve Bank of St. Louis, INSEAD, McGill, UCSD, University of Virginia and University of York for discussions and helpful comments.

1. **Introduction**

Although Black and Scholes' (1973) formula remains the most commonly used option pricing model in financial markets, a large literature has documented its strong empirical biases. Most commonly, such biases are associated with the appearance of systematic patterns (smiles or skews) in the implied volatility surface produced by inverting market prices and solving for the unknown volatility parameter (e.g. Rubinstein (1985, 1994) and Dumas, Fleming and Whaley (1998)). Implied volatility also appears to be systematically related to the term structure of option contracts (Das and Sundaram (1999)).

Several pricing models have been proposed to overcome these problems. These include stochastic volatility (Hull and White (1987), Wiggins (1987), Melino and Turnbull (1990), Heston (1993)) and ARCH models (Duan (1995)); models with jumps in the underlying price process (Merton (1976)); jump-diffusion models (Ball and Torous (1985), Amin (1993)); and models incorporating transaction costs (Leland (1985)). Bakshi, Cao and Chen (1997) summarize the empirical performance of these models. Most option pricing models fail to improve significantly on the empirical fit of the Black-Scholes (BS) model and, by modifying the stochastic process followed by the underlying asset price, do not provide a direct economic explanation for the systematic shortcomings of BS. Nevertheless, this literature has contributed significantly to our understanding of the requirements of an option pricing model that can fit observed data.

In this paper we relax the key assumption underlying the BS model of full information about the stochastic process that drives fundamentals. More specifically, we assume that fundamentals evolve on a binomial lattice with 'up' and 'down' probabilities that are unknown to investors who update their probability estimates using Bayes' rule. The underlying asset price process is determined by embedding the learning mechanism in an equilibrium model. In equilibrium, asset prices reflect all possible future probability distributions of the parameter estimates and there are no expected gains from implementing trading strategies based on the unfolding of estimation uncertainty.

While there are now many papers studying the asset pricing effects of learning,[1] the only other study that explicitly considers its derivative pricing implications ap-

[1] E.g., Bossaert (1995, 1999) and Timmermann (1993).

pears to be David and Veronesi (1999). Investors in their model face a filtering problem which leads them to update the probability of which of two states fundamentals are currently in. Since investors do not know the true state, shocks to fundamentals are effectively drawn from a non-Gaussian mixture distribution and hence the BS model is not obtained in the limit. In contrast, we assume that fundamentals follow a binomial process, whose limit is Gaussian. This allows us to isolate the effect of investors' recursive estimation of parameter values in a model that converges asymptotically to BS. Another difference is that we study the ability of the learning model to predict out-of-sample the evolution in the entire cross-section of option prices and to generate lower hedging errors than Black-Scholes. The importance of such exercises for evaluating different option pricing models has recently been emphasized by Dumas et al (1998).

Our approach is also related to a large literature that infers the market's probability beliefs from asset prices (Rubinstein (1994), Jackwerth and Rubinstein (1996), Aït-Sahalia and Lo (1998)) although our results explicitly incorporate the effect of investors' beliefs on equilibrium prices. The assumption that the market updates its beliefs through Bayes rule adds a new aspect to this exercise and provides an understanding of the time-series dynamics of implied volatility surfaces.

Simply replacing the assumption in the BS model of known 'up' and 'down' probabilities with Bayesian learning is found to generate biases similar to those observed in option prices. Consistent with recent empirical evidence (Aït-Sahalia and Lo (1998)), learning alters the shape of the state price density perceived by investors by adding to tail probabilities. Furthermore, learning effects can generate implied volatility smiles as well as a variety of non-constant term structures of implied volatility.[2] By inverting the resulting model, we can infer the dynamics of the parameters of the Bayesian learning scheme from data on S&P 500 index option prices. Independently of the time period over which theoretical option prices are matched with observed prices, we find that estimated parameters are remarkably stable over time and that our model provides an excellent in-sample and out-of-sample fit in addition to generating smaller hedging errors than the BS model and various alternatives proposed in the literature.

[2]Das and Sundaram (1999) show that the most popular alternatives to BS — jump-diffusion and stochastic volatility models — fail to simultaneously generate implied volatility smiles and term structures that match the complex features of the data.

The outline of the paper is as follows. Section 2 introduces the data set and briefly describes systematic biases in the BS option pricing model. Section 3 presents the binomial lattice model under full information and Bayesian learning is introduced in Section 4 which also derives explicit formulae for European option prices. Section 5 characterizes the equilibrium effect of learning on option prices and calibrates the option pricing model under learning so it can be compared to the data from Section 2. The parameters characterizing the maintained recursive learning process are inferred from option prices in Section 6 and used to predict option prices out-of-sample. Section 7 concludes.

2. **Biases in the Black-Scholes Model**

This section briefly outlines the systematic pricing biases in the BS option pricing model and thus serves as a benchmark for the empirical analysis. Our data set of option prices from the CBOE consists of S&P 500 index option prices[3] covering the period Jan. 4, 1993 to Dec. 31, 1993 and is identical to that used in Aït-Sahalia and Lo (1998) and has been filtered in the same way.[4] While the calendar period is relatively short compared to that used by Dumas et al (1998), the number of observations (13,090) is comparable due to the fact that we use daily rather than

[3] As stressed by Rubinstein (1994) the market for S&P 500 index options on the CBOE provides a case study where the conditions required by BS seem to be well approximated in terms of volumes, continuity of the trading process and hedging opportunities.

[4] We thank Yacine Aït-Sahalia for making his data set available to the public. In-the-money options are thinly traded, their prices are notoriously unreliable and are thus discarded from the data set. Out-of-the-money and near-at-the-money put prices, P, are translated into call prices, C, using put-call parity. All information contained in liquid put prices has thus been extracted and converted into call prices. The remaining put options are discarded from the data set without loss of information. We explicitly take into account that the index pays out daily dividends and follow Aït-Sahalia and Lo (1998) in using the continuous dividend yield, δ, implied by daily values of the index, S, and prices of futures contracts, F, of given maturity, τ. By the spot-futures parity

$$F_{t,\tau} = S_t e^{(r_{t,\tau} - \delta_{t,\tau})\tau} \Longrightarrow \delta_{t,\tau} = \frac{1}{\tau} \ln\left(\frac{S_t}{F_{t,\tau}}\right) + r_{t,\tau}.$$

The implied value of $\delta_{t,\tau}$ is a measure of the continuously compounded dividend yield that, as of time t, is expected between t and $t+\tau$, while $r_{t,\tau}$ is the riskfree rate over this period. The only difference with respect to the data employed by Aït-Sahalia and Lo (1998) is that we limit ourselves to options with less than 200 days to expiration.

weekly data in all the exercises.

2.1. *Implied Volatility Surfaces*

Initially we confirm the existence of a systematic skew in the relationship between BS implied volatility and moneyness. Figure 1 plots the implied volatility surface against moneyness for six maturities represented in our data sample.[5] For most days in the sample period, the curve relating BS implied volatilities to the strike price is skewed. This is of course inconsistent with the maintained assumption of a constant diffusion in the BS model.

2.2. *Term Structure of Implied Volatility*

The data also reveals a systematic term structure in implied volatilities. Using three alternative values of moneyness over the period Jan. 18, 1993 to Jan. 25, 1993, Figure 2 shows that the implied volatilities of at-the-money options and options with moneyness less than one increase with time to expiration. For in-the-money options the pattern is somewhat weaker: some days implied volatility is an increasing and convex function of time to expiration; other days implied volatility is a concave function of moneyness and occasionally the pattern is constant or even declining.

2.3. *State Price Densities*

Recently Rubinstein (1994) and Jackwerth and Rubinstein (1996) proposed to extract state price densities (SPD) from implied binomial trees. This is another powerful procedure for demonstrating biases in the BS model whose assumption is that the state price density is log-normal. Figure 3 shows the SPD inferred from option contracts with at least 50 calendar days to expiration and averaged across calendar days and maturities (a total of 334 estimated SPDs). To ensure that SPDs on different days are comparable, all plots use standardized logarithmic returns. Particularly important in economic terms is the tail behavior of the SPD since this may provide information about the jump risk expected by markets, c.f. Bates (1991). Therefore we plot in the bottom of Figure 3 the estimated tail behavior of

[5]The finding of skews in implied volatility is generic and holds across different maturities.

the average SPD.

Compared to the lognormal benchmark, the SPD is clearly leptokurtic. Market participants assign high value to future 'extreme' outcomes that under a lognormal SPD would receive a much smaller state price. To demonstrate this point, Table I compares the no-arbitrage prices of a state-contingent asset paying off one dollar when the S&P 500 returns exceed (in absolute value) a certain number of standard deviations of returns under the estimated SPDs versus under a lognormal benchmark. Higher prices for tail-contingent securities reflect the fact that under the estimated SPDs the market assigns a much higher risk-neutral price to a dollar paid out in either crash or strongly bullish states of the world.

3. **Asset Prices under Full Information**

The empirical findings reported in Section 2 confirm the presence of systematic biases in the BS model and suggest that a more general option pricing model is required. This section characterizes option prices in a full information equilibrium model which has the BS model as a limiting case and sets up a benchmark from which to evaluate option pricing biases. Learning is introduced in Section 4.

3.1. *Fundamentals on a Binomial Lattice*

Our starting point is a version of the infinite horizon, representative agent endowment economy studied by Lucas (1978). There are three assets: A one-period default-free, zero-coupon bond in zero net supply trading at a price of P_t and earning interest of $r_t = (1/P_t - 1)$; a stock traded at a price of S_t whose net supply is normalized at 1; and a European call option written on the stock with $\tau \equiv T - t$ periods to expiration, strike price K and current price C_t.

The stock pays out an infinite stream of real dividends $\{D_{t+k}\}_{k=1}^{\infty}$. Dividends are perishable and are consumed in the period when they are received. Dividend growth rates, $g_{t+k} = \frac{D_{t+k}}{D_{t+k-1}} - 1$, follow a Bernoulli process that is subject to change m times in each unit interval. Within the interval $[t, t+\tau]$ dividends thus follow a $v = \tau m$-step binomial process. For each interval the dividend growth rate is g_h with probability π or g_l with probability $1 - \pi$:

$$g_{t+k+1} = \begin{cases} g_h & \textit{with prob. } \pi \\ g_l & \textit{with prob. } 1-\pi \end{cases} \qquad \forall k \geq 0, \quad \pi \in (0,1) \tag{1}$$

Without loss of generality we assume that $g_h > g_l > -1$ so that dividends are non-negative provided $D_t > 0$. This gives a standard recombining binomial tree similar to the one adopted by Cox, Ross and Rubinstein (1979) for the underlying asset price process. We follow the literature in normalizing the parameters by the incremental time unit: $1 + g_h = e^{\sigma\sqrt{\frac{dt}{v}}}$, $1 + g_l = (1 + g_h)^{-1}$, and $\pi = \frac{1}{2} + \frac{1}{2}\frac{\mu}{\sigma}\sqrt{\frac{dt}{v}}$. As $\frac{dt}{v} \to 0$, the distribution of dividends converges weakly to a geometric Brownian motion with constant drift and diffusion (μ, σ).

To price assets we assume a perfect capital market. There are unlimited short sales, perfect liquidity, no taxes, no transaction costs or borrowing and lending constraints and markets are open at all the nodes of the binomial lattice where news on dividends are generated.

The representative investor has power utility

$$u(Z_t) = \begin{cases} \frac{Z_t^{1-\gamma}-1}{1-\gamma} & \gamma < 1 \\ \ln Z_t & \gamma = 1 \end{cases} \tag{2}$$

where Z_t is real consumption at time t. We focus on the case where $\gamma \leq 1$; models with $\gamma > 1$ have the counter-intuitive property that stock prices decline when the probability of high growth of the fundamentals increases, c.f. Abel (1988). The representative agent chooses bond, stock, and call option holdings to maximize the discounted value of expected future utilities derived from consumption subject to a budget constraint:

$$\max_{\left\{Z_{t+k}, w^s_{t+k}, w^b_{t+k}\right\}_{k=0}^{\infty}} E_t\left[\sum_{k=0}^{\infty} \beta^k u(Z_{t+k})\right] \tag{3}$$
$$s.t.\ Z_{t+k} + w^s_{t+k} S_{t+k} + w^b_{t+k} P_{t+k} = w^s_{t+k-1}(S_{t+k-1} + D_{t+k-1}) + w^b_{t+k-1},$$

where $\beta = \frac{1}{1+\rho}$, ρ is the rate of impatience and w^s_{t+k} and w^b_{t+k} represent the number of stocks and bonds in the agent's portfolio as of period $t + k$.[6]

Standard dynamic programming methods yield the following Euler equations for stock and bond prices:

$$\begin{aligned} S_t &= E_t\left[Q_{t+1}(S_{t+1} + D_{t+1})\right] \\ P_t &= E_t\left[Q_{t+1}\right] \end{aligned} \tag{4}$$

[6] Since the call is a redundant asset which does not expand the set of attainable consumption patterns, option holdings do not enter into the program and the equilibrium stock and bond prices can be determined independently of the option price.

where $Q_{t+1} = \beta \frac{u'(Z_{t+1})}{u'(Z_t)} = \beta \left(\frac{Z_{t+1}}{Z_t}\right)^{-\gamma}$ is the pricing kernel defined as the product of the discount factor and the intertemporal marginal rate of substitution in consumption.

Guidolin and Timmermann (1999) price the underlying stock and risk-free bond in this setting subject to a transversality condition.[7] For convenience, we state asset prices using the transformed parameters $g_l^* = (1+g_l)^{1-\gamma}-1$ and $g_h^* = (1+g_h)^{1-\gamma}-1$.

Proposition 1 (Guidolin and Timmermann (1999)) *The full information rational expectations stock price is given by*

$$S_t = \frac{1 + g_l^* + \pi(g_h^* - g_l^*)}{\rho - g_l^* - \pi(g_h^* - g_l^*)} D_t,$$

while the full information bond price is

$$P_t = \frac{(1+g_l)^{-\gamma} + \pi\left[(1+g_h)^{-\gamma} - (1+g_l)^{-\gamma}\right]}{1+\rho}.$$

A property of the solution is that the stock price is homogeneous of first order in dividends and that the ex-dividend stock price follows the same binomial lattice $\{g_h, g_l, \pi, m\}$ as dividends.[8] This means that dividends and stock prices follow a stationary Markov chain.

3.2. *Option Prices under Full Information*

Pricing European calls and establishing the link to Black-Scholes is straightforward under full information. This follows from noting that in our model (i) arbitrage opportunities are ruled out; (ii) markets are complete; (iii) ex-dividend stock prices inherit the binomial lattice structure $\{g_h, g_l, \pi, m\}$ from dividends; (iv) although the underlying asset pays out cash dividends, the option is European and early exercise

[7] The transversality condition is:

$$\lim_{T\to\infty} E_t\left[\left(\prod_{k=1}^{T} Q_{t+k}\right) S_{t+T}\right] = 0.$$

[8] When combined with the result in Pliska (1997) that a multiperiod security markets model is complete if and only if all possible sequences of single period models obtained by decomposing the binomial lattice are formed by complete models it follows that the asset market in our model is complete and the call option is a redundant asset.

is not possible. Therefore we can draw on the result of Cox et al. (1979) that the price of a European option converges to the BS value provided the parameters are suitably adjusted as the number of increments to the lattice goes to infinity. Let r be the risk-free rate, δ the dividend yield, and μ and σ respectively the (annual) mean and volatility of the dividend growth rate, while dt is an interval of fixed calendar time (typically $\tau/252$, supposing there are 252 trading days). The following proposition restates this result and shows the mapping between the deep parameters of our model and the BS inputs.

Proposition 2 *Suppose that the parameters have been scaled as follows:*

$$g_h^{(m)} = e^{\sigma\sqrt{dt/v}} - 1, 1 + g_l^{(m)} = \left(1 + g_l^{(m)}\right)^{-1}, \pi^{(m)} = \frac{1}{2} + \frac{1}{2}\frac{\mu}{\sigma}\sqrt{\frac{dt}{v}}$$

$$\rho^{(m)} = (1+r)^{\frac{dt}{v}} \left\{(1 + g_l^{(m)})^{-\gamma} + \pi^{(m)} \left[(1 + g_h^{(m)})^{-\gamma} - (1 + g_l^{(m)})^{-\gamma}\right]\right\} - 1 > 0$$

$$\delta = \frac{(1+r)\left\{(1 + g_l^{(m)})^{-\gamma} + \pi^{(m)} \left[(1 + g_h^{(m)})^{-\gamma} - (1 + g_l^{(m)})^{-\gamma}\right]\right\}^{\frac{v}{dt}}}{\left\{(1 + g_l)^{1-\gamma} + \pi^{(m)} \left[(1 + g_h^{(m)})^{1-\gamma} - (1 + g_l^{(m)})^{1-\gamma}\right]\right\}^{\frac{v}{dt}}} - 1.$$

Then as $m \to \infty$, the price of the European call converges to its BS value:

$$\begin{aligned} C_t^{BS} &= S_t \Phi(d_1) e^{-\delta dt} - e^{-rdt} K \Phi(d_2) \\ d_1 &= \frac{\ln\left(\frac{S_t}{K}\right) + (r - \delta)\, dt + \frac{1}{2} v \left[\ln(1 + g_h^{(m)})\right]^2}{\sqrt{v}\ln(1 + g_h^{(m)})} \\ d_2 &= d_1 - \sqrt{v}\ln(1 + g_h) \end{aligned}$$

where $v = \tau m$ and $\Phi(\cdot)$ is the c.d.f. of the standard normal distribution.

Proof. See Appendix A. ■

As $m \to \infty$, $\frac{dt}{v}$ goes to zero, and the binomial lattice converges weakly to a geometric Brownian motion with parameters (μ, σ), the distributional assumption required by BS in continuous time.[9] This proposition shows that the results of Cox

[9] This point can also be shown by noting that as $m, v \to \infty$ the discrete state price density converges to a transformation of the lognormal:

$$\begin{aligned} \widetilde{f}(S_{t+T}) &= e^{-r\tau} \frac{1}{\sqrt{2\pi}\sigma\sqrt{\tau}} \frac{1}{S_{t+T}} \\ &\quad \times \exp\left\{ - \frac{\left[\ln(S_t) - \left(r - \delta + \frac{1}{2}\left[\ln(1+g_h)\right]^2\right) v\right]^2}{2v\left[\ln(1 + g_h)\right]^2} \right\}. \end{aligned}$$

et al. fully extend to our framework where dividends rather than stock prices are assumed to follow a binomial lattice. Notice, however, that while Cox et al. take the process for the underlying price as exogenous, we derive the underlying stock price in an equilibrium model in which preferences matter. This result can be also related to Stapleton and Subrahmanyam (1984) who value options in a general equilibrium model when markets are incomplete and the stock price evolves on a binomial lattice. In contrast to Stapleton and Subrahmanyam, our model assumes that markets are complete, but the exogenous binomial lattice process applies to dividends. Obviously in both cases preferences affect the equilibrium price of stocks as well as options. Since the BS option price is obtained in the limit under full information, this model is ideally suited to discuss the origin of BS pricing biases.

4. **Option Prices on a Learning Path**

It is common in the option pricing literature to assume a given process for the underlying asset price and then price the option as a redundant asset whose payoffs can be replicated in a dynamic hedging strategy invested in the risky asset and a riskfree bond. The standard setup assumes that the asset price process is stationary and hence that there are no learning effects. Once learning is introduced, an equilibrium model for the underlying asset price is required.

Suppose that the proportion of times dividends move up on the binomial lattice, π, is unknown to investors who recursively estimate this through the simple maximum likelihood estimator:

$$\widehat{\pi}_{t+k,j} = \frac{\sum_{j=1}^{m(t+k)+j} I_{\{m(t+k)+j=g_h\}}}{m(t+k)+j} = \frac{n_{m(t+k)+j}}{N_{m(t+k)+j}} \quad j = 0, 1, ..., m-1 \tag{5}$$

where $I_{\{m(t+k)+j=g_h\}}$ is an indicator function which is one when at the $m(t+k)+j$'th step of the lattice dividend growth is high, and is zero otherwise. n_i denotes the number of high growth states recorded up to node i, while N_i is the total number of nodes since time t. The indices $m(t+k)+j$ $(j = 0, 1, ..., m-1)$ take into account that learning happens on the binomial tree and not over calendar time. Investors are assumed to start out with prior beliefs $\{n_0, N_0\}$ and update these through Bayes rule.

Despite the presence of learning effects, the same features that simplified the solution of asset prices under full information are still in place: (i) consumption and dividends must coincide in general equilibrium; (ii) if markets are complete, investors form portfolio choices based only on the stock index and the bond; (iii) being redundant assets, options can be priced by no-arbitrage, using the unique risk neutral probability measure. This can be used to prove the following result:

Proposition 3 *The stock price under Bayesian learning (BL) is*

$$S_t^{BL} = D_t \underset{v\to\infty}{\ell im} \left\{ \sum_{i=1}^{v} \beta^i \sum_{j=0}^{i} (1+g_h^*)^j (1+g_l^*)^{i-j} \Pr{}_t^{BL} \left(D_{t+i}^j | n_t, N_t \right) \right\}$$

where the posterior distribution $\Pr{}_t^{BL} \left(D_{t+i}^j = (1+g_h)^j (1+g_l)^{i-j} D_t | n_t, N_t \right)$ *is given by*

$$\Pr{}_t^{BL} \left\{ D_{t+i}^j | n_t, N_t \right\} = \binom{i}{j} \frac{\prod_{k=0}^{j-1} (n_t+k) \prod_{k=0}^{i-j-1} (N_t - n_t + k)}{\prod_{k=0}^{i-1} (N_t + k)}.$$

The bond price under Bayesian learning is

$$P_t^{BL}(\widehat{\pi}_t) = \widehat{E}_t \left[\beta (1+g_{t+1})^{-\gamma} \right] = \frac{(1+g_l)^{-\gamma} + \widehat{\pi}_t \left[(1+g_h)^{-\gamma} - (1+g_l)^{-\gamma} \right]}{1+\rho}.$$

For a proof, we refer to Guidolin and Timmermann (1999). Proposition 3 has several implications. First, the price-dividend ratio is no longer constant and depends (through n_t and N_t) on the current estimate $\widehat{\pi}_t$. Dividend changes acquire a self-enforcing nature: Positive dividend shocks lead to an increase in the stock price not only through the standard proportional effect, but also through the revision of the dividend multiplier.

Although dividend yields are now time-varying, in principle binomial methods could still be used to obtain the no-arbitrage price of European options on flexible trees (see, e.g., Chriss (1997)). However, risk-free rates in our equilibrium model are not only time-varying, but also a function of the state variable $\widehat{\pi}_{t+k}$. Furthermore, the interest rate process cannot be characterized as a recombining lattice. The value today of one dollar in the future depends not only on the *number* of high and low growth states occurring between today and the future, but also on their *sequence*. In other words, the appropriate discount factors become *path-dependent*. Since risk-free rates show up in the general risk-neutral valuation formula proved

by Harrison and Kreps (1979), the induced lattice for the call option also becomes non-recombining.[10]

This path dependence means that no-arbitrage methods are more complicated in an equilibrium model with learning. Nevertheless, European call options can still be priced by employing a change of measure based on the perceived probabilities:

Proposition 4 *On a Bayesian learning path, the no-arbitrage price of a European call with time-to-expiration τ and strike price K is*

$$C_t^{BL}(K,T,S_t^{BL}) = \sum_{j=0}^{v} \max\left\{0, S_{t+v}^{BL,j} - K\right\} \widetilde{P}_t^{BL}\left\{S_{t+v}^{j} | n_t, N_t\right\}$$

where $v = \tau m$, $S_{t+v}^{BL,j} = (1+g_h)^j (1+g_l)^{v-j} S_t^{BL} = (1+g_h)^j (1+g_l)^{v-j} \Psi_{t+v}^{BL}(n_t + j, N_t + v) D_t$, $D_{t+v}^{j} = (1+g_h)^j (1+g_l)^{v-j} D_t$ *($j = 0, 1, ..., v$), and*

$$\begin{aligned} \widetilde{P}_t^{BL}\left\{S_{t+v}^{j} | n_t, N_t\right\} &= \widetilde{P}\left\{D_{t+v}^{j} | n_t, N_t\right\} = \beta^v \left(\frac{D_{t+v}^{j}}{D_t}\right)^{-\gamma} \times \\ &\quad \times \binom{v}{j} \frac{\prod_{k=0}^{j-1}(n_t + k) \prod_{k=0}^{v-j-1}(N_t - n_t + k)}{\prod_{k=0}^{v-1}(N_t + k)} \end{aligned}$$

Proof. See Appendix A. ■

Under learning stock prices and beliefs no longer follow a stationary Markov chain, since both the possible rates of change of the stock index and the (perceived) probabilities of these changes follow heterogenous Markov chains. The time-varying transition matrix that determines how the risk-neutral distribution is updated is given by

$$\left[\widetilde{P}^{BL}\left\{X_{t+k+1} = n_t + j | X_{t+k} = n_t + i\right\}\right] = \widetilde{M}_{t+k}(i+1, j+1) \tag{6}$$

$$= \beta \begin{bmatrix} R_{t+k}^{d}(1+g_l)^{-\gamma}\frac{N_{t+k}-n_t}{N_{t+k}} & R_{t+k}^{u}(1+g_h)^{-\gamma}\frac{n_t}{N_{t+k}} & 0 & ... & 0 \\ 0 & R_{t+k}^{d}(1+g_l)^{-\gamma}\frac{N_{t+k}-n_t-1}{N_{t+k}} & R_{t+k}^{u}(1+g_h)^{-\gamma}\frac{n_t+1}{N_{t+k}} & ... & 0 \\ .... & ... & ... & & \vdots \\ 0 & 0 & 0 & ... & 1 \end{bmatrix}$$

[10] This is another important difference between our model and that of David and Veronesi (1999) which assumes multiple states for dividends. In a Lucas-type general equilibrim model this will result in a state-dependent consumption and risk-free rate process. However, David and Veronesi assume that the interest rate is fixed, suggesting that their results are partial equilibrium.

where $R^d_{t+k} \equiv 1+r^d_{t+k}$ is the gross interest rate factor that applies to the low growth state at $t+k$ etc. Computation of the risk neutral probabilities now requires keeping track of the risk-free rate as we move along the dividend tree, reflecting the path-dependence in this variable.

5. BS anomalies and learning effects

To better understand the sense in which Bayesian learning can explain the biases in the BS model we establish conditions under which learning systematically affects option prices. Estimation uncertainty reduces the underlying asset price when investors are risk averse, but also increases the asset price through the positive covariance between future asset payoffs and beliefs ($\widehat{\pi}_t$). When risk aversion is not 'too high' the second effect dominates. A similar result can be established by comparing option prices under learning to BS option prices:[11]

Proposition 5 *Suppose dividends follow a binomial process subject to the parameter restrictions $g^*_h \geq \rho > g^*_l$.. If investors have optimistic beliefs ($\widehat{\pi}_t \geq \pi \geq \frac{1}{2}$) then*

$$C^{BL}_t(K) \geq C^{BS}_t(K).$$

One can show that the difference between the call option price under Bayesian learning and under full information is positive for a zero strike price, increases over some interval of strike prices and then decreases towards zero. Figure 4 provides an example.

The technical condition $g^*_h \geq \rho > g^*_l$ is sufficient for the price-dividend ratio to be a monotonically increasing and convex function of beliefs, $\widehat{\pi}_{t+k}$. Optimistic beliefs ($\widehat{\pi}_{t+k} > \pi$) are sufficient for the difference between option prices $C^{BL}_t(K) - C^{BS}_t(K)$ to be monotonically increasing for low strike prices.

More generally there will be intervals for the strike price over which $C^{BL}_t(K) \geq C^{BS}_t(K)$ and others over which this inequality is reversed. Furthermore, although optimistic beliefs are sufficient, they are not a necessary condition for learning to systematically affect option prices. Effectively the cross-section of option prices allows us to infer the market's perception of fundamentals once the model is tested

[11] The proof, which is lengthy and therefore omitted from here, is available from the authors on request.

on option data. For instance, market prices for European calls systematically above BS predictions and a 'smiling' implied volatility shape would suggest strong learning effects and that investors are somewhat optimistic.

Ultimately an analysis must consider whether the Bayesian option pricing model can replicate the data in Section 2. To investigate this, we calibrate the parameters of the dividend process and fix the risk-free rate at a plausible level. Dividends are assumed to be paid out daily ($m = 1$). For a wide market index such as the S&P 500 this provides a fairly good approximation. The annual dividend growth rate (μ) is set to 3%, to match the average dividend growth rate around our sample period (June 1992 - June 1995). Volatility (σ) is set at 5%, slightly lower than the point estimate of the annual dividend growth rate which is likely to be inflated by the presence of periods where no dividend changes show up even though markets receive news more gradually.[12] Finally the annualized risk-free rate (r) is set to 4%, while the dividend yield is fixed at 3%, matching its estimate of 2.82% during the period 1992-1995. The annualized subjective discount rate is 0.02.

We consider European call options with 50 days to maturity ($\tau = v = 50$). From Proposition 2 it follows that no-arbitrage option prices are determined on a binomial lattice model with parameters

$$g_h = 0.00315,\ g_l = -0.00314,\ \pi = 0.5189,\ \rho = 7.96 \cdot 10^{-5},\ \gamma = 0.999.$$

Drawing on Proposition 5, we initially assume marginally biased initial beliefs and mild learning effects: $n_t = 42$ and $N_t = 80$, i.e. $\hat{\pi}_t = 0.525 > 0.519 = \pi$.[13] BS prices are based on the full information model.

5.1. *Implied Volatility Surfaces*

First consider implied volatility surfaces. BS implied volatility is a highly nonlinear function of option prices so a formal treatment is difficult and we provide the following heuristic explanation in the context of Figure 4. At-the-money BS option

[12] Campbell (1996) makes the same observation and estimates an annualized volatility of real US dividend growth of 7 percent.

[13] Under pessimistic beliefs the BL model is less able to produce option prices that fit the stylized facts of Section 2. For example, the implied volatility surface tends to increase as a function of the strike price. Although the SPD still has fatter tails than the log-normal distribution, most probability mass is shifted to the left of the return distribution when beliefs are pessimistic.

values are known to be sensitive — their *vega* is high — to the volatility input while deep out-of-the-money and in-the-money BS option values are not so sensitive to changes in σ. For low strike prices C_t^{BL} starts out above the BS value and their difference initially increases as the strike price rises. However, when deep-in-the money options are ruled out, the fact that vega also increases makes the volatility implied by Bayesian learning prices a decreasing function of the strike. For mildly in-the-money and at-the-money options, the difference between C_t^{BL} and the BS value starts declining, so that σ^{BL} declines as well. For deep out-of-the money options σ^{BL} also declines towards σ^{BS}. Therefore, if traded strikes do not extend to levels for which the vega is extremely low, a declining implied volatility skew emerges.[14] In summary, the option pricing model under learning can replicate the patterns observed in implied volatilities.

Figures 5 and 6 plot implied volatilities as a function of moneyness for the April 1993 maturity. These are representative of what we have found in other subperiods of our sample. Comparing the left and right windows, the resemblance between the implied volatility under learning and observed market values is striking. The lattice model under Bayesian learning appears to price options far more accurately than the BS model.

To further underline this point, Figure 7 compares observed option prices on February 22, 1993 and BL prices assuming $n_t = 43$, $N_t = 80$ ($\hat{\pi}_t = 0.538$). The fit is even more striking than in Figure 6 and is indicative of the ability of the model to fit implied volatility skews. Assuming that the markets really were on a Bayesian learning path on that day, the estimate $\widehat{\pi}_t = 0.538$ with a precision of $N_t = 80$ seems to accurately characterize investors' beliefs.

5.2. *State Price Densities*

Systematic differences between BS and BL European option prices must reflect differences in the underlying equivalent martingale measures employed by the market under the two alternative models or, equivalently, in the SPDs (cf. Harrison

[14]However, for very high strikes, a vega approaching zero can create a volatility smile. For very low strikes, the fact that $C_t^{BL} - C_t^{BS}$ is increasing, can also lead to a mildly increasing volatility curve. David and Veronesi (1999) report the occasional presence of 'frowns' of the implied volatility in the S&P 500 index option markets. In our model such frowns are generated when investors are strongly pessimistic.

and Kreps (1979)). Learning affects state price densities in two distinct ways: (i) the SPD perceived by investors changes; (ii) the support of the set of time $t+T$ equilibrium stock prices is widened. If the initial beliefs on π are unbiased, extreme events drawn from either end of the tail are perceived as more likely on a Bayesian learning path than under a lognormal distribution. Likewise, a wider support for the SPD also means that more extreme events now become possible.

When $\widehat{\pi}_t$ decreases towards π from above, BL continues to inflate the tails of the SPD perceived by the investors relative to the lognormal benchmark thus creating an implied volatility skew. This matches the stylized facts of the S&P 500 option data that densities are located more to the right than the lognormal benchmark but also attach positive probability mass to some crash events. Figure 8 shows that, consistent with the data in Section 2, Bayesian learning produces SPDs that are skewed to the right and have fatter tails than a lognormal.

5.3. *Term Structure of Implied Volatilities*

Next we vary the time to maturity (τ) from 10 to 150 days in steps of 10 days to study the implied volatility term structure resulting from Bayesian learning. Figure 9 shows the outcome of this exercise for three sets of strike prices: $K = 420$ (moneyness 1.04), $K = 435$ (moneyness 1), and $K = 455$ (moneyness of 0.96). Bayesian learning generates an upward sloping term structure for at-the-money and out-of-the money call options, while the term structure at first decreases and then increases for in-the-money call options. These patterns are broadly consistent with what was found in the data, cf. figure 2. The increase of about 2 percentage points in implied volatility between close-to-expiration options and long-term options is also plausible.[15]

5.4. *Vanishing learning effects*

In our setup investors use a consistent estimator of π, and as the precision goes to infinity ($N_t \to \infty$), $\hat{\pi}_t \to \pi$ and learning effects diminish. To study the consequence of this, we double the precision of investors' beliefs with respect to the experiments performed in section 5, keeping the mean constant, i.e. $n_t = 182$ and $N_t = 350$. The representative agent now brings experience of over 16 months of trading and

[15] See Campa and Chang (1995).

dividend realizations. Figure 10 shows the resulting option prices.

When learning effects are weak and investors have a more accurate estimate of π, BL option prices converge to BS prices. This is unsurprising since learning is the only source of non-stationarity in our model. The first panel of Figure 10 shows that differences previously of the order of 1-2 dollars, now decline to a quarter of that range. Panels two and three show that BL implied volatilities continue to display a systematic pattern over moneyness, although the implied volatility surface is flatter than the one exhibited in Figure 6. As $N_t \to \infty$ and $\hat{\pi}_t \to \pi$ (from above) a smile is obtained instead of the smirk in figure 6. This indicates that at times when option markets are characterized by smiles, learning effects are weak in the sense that investors attach high precision to their initial beliefs. Smirks, on the other hand, are indicative of markets with strong learning effects and uncertain beliefs. Finally, the fourth panel shows that the SPD converges to the log-normal distribution.[16]

To ensure that learning effects do not disappear asymptotically, one can simplify the learning model to assume that the markets use a rolling window of the data to estimate π. This guarantees that the markets will not obtain an infinite precision of the fundamentals parameters as the sample grows. This is also consistent with widespread market practice and is a way of robustifying the estimate of π with respect to slowly moving non-stationarities in the fundamentals process. An additional benefit from this assumption is that we can deduce from the options price data the effective memory (or window length, N_t) applied by the market, which is of separate interest. As long as the implied window is not too small, the Bayesian model will provide a good approximation to an exact solution which explicitly incorporates investors' use of a rolling window.[17]

[16] We omit the discussion of the impact of a large N_t on the volatility term structure. Though the patterns are unchanged, the implied volatility curves get flatter and their level moves down towards the BS value. Nevertheless, a precision, N_t, in excess of 16 months of observations implies effects that are still close to those observed in the data.

[17] Simulation experiments confirmed that for the values of N_t typically encountered in our sample, namely around 180-220, the Bayesian price under an expanding and a rolling window were quite similar.

6. **Option price dynamics and learning effects**

Thus far we have studied the ability of the BL model to match typical cross-sections of option prices on individual days. However, the Bayesian updating algorithm implies a set of dynamic restrictions on how implied volatility surfaces and term structures evolve over time as investors update $\widehat{\pi}_t$. Such testable restrictions do not have a counterpart in the BS model which does not consider the effect of changing probability beliefs. By tracking option prices on several consecutive days, not only do we get insights into how investors change their beliefs but we also get a more precise estimate of the initial beliefs.

Estimating the dynamics of beliefs from observed option prices is a new exercise that needs to be put in perspective. When asset markets are (dynamically) complete, equilibrium asset prices contain information about preferences and beliefs. Rubinstein (1985) observes that any pair of the following implies the third: (1) the preferences of a representative agent; (2) agents' beliefs; and (3) the state-price density (SPD). A vast literature has attempted to use the observed prices of risky assets to infer preferences, the stochastic process of prices, or both. For instance, Bick (1990) and He and Leland (1993) impose parametric restrictions on the generating process of asset prices and infer the preferences of a representative agent. Rosenberg and Engle (1997) develop a nonparametric estimator of the empirical pricing kernel which is based on the ratio of the estimated state prices and the estimated physical (objective) probability beliefs defined on a discrete grid of possible future returns. They use their pricing kernel estimates to document time-variation in the risk attitudes of the market. Furthermore, Bates (1991), Jackwerth and Rubinstein (1996), and Aït-Sahalia and Lo (1998) back out the perceived risk neutral stochastic process of asset prices from observed option prices.

Bayesian learning provides an as far unexplored possibility to expand Rubinstein's list to a fourth and separate item: the dynamics of the learning process followed by a representative agent in an equilibrium model. As in Rubinstein (1994) and Jackwerth and Rubinstein (1996), we fix preferences and infer a vector of unknown parameters from observed option prices. Since the dynamics of beliefs on a learning path determine the evolution in the SPD and therefore also option prices, the parameters on the entire learning path can be backed out from the following general program[18]

[18] Our approach is similar to Bates (1991), who imposes CRRA preferences to estimate by NLS

$$\min_{\{\pi_t\}_{t=1}^T, N, m} \sum_{t=1}^{T} \sum_{\tau=\underline{\tau}_t}^{\bar{\tau}_t} \sum_{K_{\tau_t}=\underline{K}_{\tau_t}}^{\overline{K}_{\tau_t}} g\left(C^{BL}(f_t, \pi_t, N, m; \gamma, \beta), C(f_t)\right)$$

$$s.t. \quad \frac{\pi_t N}{N+m} \leq \pi_{t+1} \leq \frac{\pi_t N + m}{N+m}$$

$$0 \leq \pi_t \leq 1 \quad t = 1, ..., T \tag{7}$$

$$N > 0, \quad m > 0$$

where $f_t = [\tau_t \; K_{\tau_t} \; S_t \; r_t^f]'$ denotes the option contract features and the underlying asset price, S_t, which we condition on, $C^{BL}(f_t, N, m; \gamma, \beta)$ is the theoretical price of the call option on a learning path and $C(f_t)$ is the observed call price at time t for an option with strike K_{τ_t} expiring in τ_t days. The indexes appended to τ_t and K_{τ_t} reflect the fact that maturities and strike prices change over time, following the dynamics of the underlying price and the financial cycle. Finally, $g(\cdot)$ is a function that measures the distance between the observed and theoretical option price. For instance, we might minimize the sum of squared pricing errors across days, strikes or maturities.

Again we assume that the annual volatility of the fundamentals, σ, is 5%. Given σ, g_h and g_l can be determined from Proposition 3. We set $\gamma = 0.9$ and $\beta = \frac{1}{1.1} \simeq 0.91$ on an annual basis. These choices are based either on the features of our data on the S&P 500 index and index options, or on what seems a priori plausible.

The estimation procedure provides an estimate $\hat{N}$ which represents how much precision the market assigns to its beliefs, an estimate $\hat{m}$ of the frequency with which these beliefs are updated over the sample period, and a $T \times 1$ vector $\hat{\pi}$ whose dynamics is constrained by Bayes rule. $\hat{N}$ can also be interpreted as an estimate of the length of the time window investors use to form their beliefs about fundamentals.

6.1. *Inferring learning from daily cross-sections*

The objective of our first exercise is to infer from option prices the belief, π_t, its precision level, N_t, and the frequency with which beliefs are updated and new

the parameters of an asymmetric jump diffusion model with systematic jump risk.

information arrives, m_t, *for each day in the sample*:

$$\min_{\pi_t, N_t, m_t} \sum_{\tau=\underline{\tau}_t}^{\overline{\tau}_t} \sum_{K_{\tau_t}=\underline{K}_{\tau_t}}^{\overline{K}_{\tau_t}} \left[C^{BL}(f_t, \pi_t, m_t, N_t) - C(f_t)\right]^2 \tag{8}$$

for $t = 1, 2, ..., T$. This amounts to minimizing the in-sample squared pricing errors produced by the model for each daily cross-section of option prices. $\hat{\pi}_t, \hat{N}_t$, and $\hat{m}_t$ can be viewed as non-standard NLS estimates.[19] The problem is solved by a combination of grid search and a polytope method, details of which are provided in Appendix B.

This exercise ignores the intertemporal restrictions imposed by our model on the updating of investors' beliefs and therefore does not provide the strongest possible test. We do not rule out large changes across days in the estimated belief $\hat{\pi}_t$ or unbounded variation in the estimated precision level $\hat{N}_t$. On the other hand, the exercise is quite simple, requiring the estimation of only three parameters on a data set with a cross-sectional size equal to the number of traded contracts on day t.

To better gauge the plausibility of the learning path implied by S&P 500 option prices, our empirical tests compare the relative performance of the BL model to the following alternatives:

$$(i)\ BS:\ C^{BS}(f_t; \hat{\sigma}_t),$$

where $\hat{\sigma}_t$ is the implied volatility that minimizes the sum of the squared deviations

$$\sum_{\tau_t=\underline{\tau}_t}^{\overline{\tau}_t} \sum_{K_{\tau_t}=\underline{K}_{\tau_t}}^{\overline{K}_{\tau_t}} \left[C^{BS}(f_t; \hat{\sigma}_t) - C(f_t)\right]^2 \qquad s.t.\ \hat{\sigma}_t > 0, t = 1, ..., T \tag{9}$$

$\hat{\sigma}_t$ is the implied volatility estimated from the cross-section of option prices at time t.

The second model is Black-Scholes generalized to allow the volatility input to depend on the moneyness of the priced option, that is, a BS with a step-wise modification to accomodate the 'smile' bias along moneyness:

[19] The problem is nonstandard both because of the presence of constraints on the parameters and because N_t is a positive integer.

$$(ii)\ BS(m): C^{BS}(f_t; \hat{\sigma}_{K_{\tau_t}})$$

where $\hat{\sigma}_{K_{\tau_t}} = \left\{\hat{\sigma}_{K^{ITM}_{\tau_t}}, \hat{\sigma}_{K^{ATM}_{\tau_t}}, \hat{\sigma}_{K^{OTM}_{\tau_t}}\right\}$ is a volatility index function of moneyness.[20] We allow $\hat{\sigma}_{K_{\tau_t}}$ to take three separate values depending on the moneyness of the option. A contract is ITM if moneyness is above +2%, it is ATM is moneyness is between -2 and +2%, and it is OTM if moneyness is below -2%. The three volatility indices are chosen to minimize the sum of squared deviations

$$\sum_{\tau_t=\underline{\tau}_t}^{\overline{\overline{\tau}}_t} \sum_{K_{\tau_t}=\underline{K}_{\tau_t}}^{\overline{K}_{\tau_t}} \left[C^{BS}(f_t; \hat{\sigma}_{K^{Mon}_{\tau_t}}) - C(f_t)\right]^2 \qquad s.t.\ \hat{\sigma}_{K^{Mon}_{\tau_t}} > 0 \tag{10}$$

$$\hat{\sigma}_{K^{Mon}_{\tau_t}} = \left\{\hat{\sigma}_{K^{ITM}_{\tau_t}}, \hat{\sigma}_{K^{ATM}_{\tau_t}}, \hat{\sigma}_{K^{OTM}_{\tau_t}}\right\}.$$

The third model is a Black-Scholes formula generalized to allow the volatility input to depend on the time-to-maturity of the priced option:

$$(iii)\ BS(\tau): C^{BS}(f_t; \hat{\sigma}_{\tau^{Mat}_t}),$$

where $\hat{\sigma}_{\tau^{Mat}_t} = \left\{\hat{\sigma}_{\tau^{short}_t}, \hat{\sigma}_{\tau^{med}_t}, \hat{\sigma}_{\tau^{long}_t}\right\}$ is a volatility index function of time-to-maturity which reflects whether the option is close-to-expiration (less than 40 calendar days to expiration), medium term (between 40 and 70 days to maturity), or has a long (more than 70 days to expiration) time-to-expiration. The parameters are chosen to minimize the sum of squared deviations

$$\sum_{\tau_t=\underline{\tau}_t}^{\overline{\overline{\tau}}_t} \sum_{K_{\tau_t}=\underline{K}_{\tau_t}}^{\overline{K}_{\tau_t}} \left[C^{BS}(f_t; \hat{\sigma}_{\tau^{Mat}_t}) - C(f_t)\right]^2 \qquad s.t.\ \hat{\sigma}_{\tau^{Mat}_t} > 0 \tag{11}$$

$$\hat{\sigma}_{\tau^{Mat}_t} = \left\{\hat{\sigma}_{\tau^{short}_t}, \hat{\sigma}_{\tau^{med}_t}, \hat{\sigma}_{\tau^{long}_t}\right\}.$$

Finally we consider a deterministic volatility BS model modified to allow σ to be quadratic in strikes (to obtain smile shapes) and linear in time-to-maturity:[21]

[20] Moneyness is defined as $100\left(\frac{F_{t,\tau}}{K} - 1\right)$, where K is the strike price and $F_{t,\tau}$ is the price of a futures contract expiring at τ.

[21] This echoes the "ad hoc strawman" model of Dumas et al. (1998, pp. 2085-2087) who argue that market makers simply smooth the implied volatility surface using polynomials and predict option prices by using a BS in which σ is a function of the strike and maturity.

$$(iv)\ BS-spline: C^{spline}(f_t) = C^{BS}(f_t; \sigma^{spline}(\tau_t, K_{\tau_t}))$$
$$\sigma^{spline}(\tau_t, K_{\tau_t}) = \max\left\{.01, \alpha_0 + \alpha_1 K_{\tau_t} + \alpha_2 K_{\tau_t}^2 + \alpha_3 \tau_t + \alpha_4 K_{\tau_t}^2 \tau_t\right\} \tag{12}$$

For each sample day we fit a model for the volatility surface $\sigma^{spline}(\tau_t, K_{\tau_t})$ by solving:

$$\min_{\boldsymbol{\alpha}_t} SSR(\boldsymbol{\alpha}_t) \equiv \sum_{\tau_t=\underline{\tau}_t}^{\overline{\overline{\tau}}_t} \sum_{K_{\tau_t}=\underline{K}_{\tau_t}}^{\overline{K}_{\tau_t}} \left[\sigma^{spline}(f_t; \boldsymbol{\alpha}_t) - \sigma_I(f_t)\right]^2 \tag{13}$$

where σ_I denotes observed market implied volatilities (calculated by inverting the BS formula). We then plug the estimated $\hat{\boldsymbol{\alpha}}_t$ into BS to obtain option prices.

6.2. *Goodness-of-fit Measures*

To measure the fit of any given option pricing model, $\widehat{C}(f_t)$, we adopt three indcators. The first is the average root mean squared valuation error (RMSVE),

$$\begin{aligned} RMSVE &= T^{-1} \sum_{t=1}^{T} \left\{ W_t^{-1} \sum_{\tau_t=\underline{\tau}_t}^{\overline{\overline{\tau}}_t} \sum_{K_{\tau_t}=\underline{K}_{\tau_t}}^{\bar{K}_{\tau_t}} \left[\widehat{C}(f_t) - C(f_t)\right]^2 \right\}^{1/2} \\ &= T^{-1} \sum_{t=1}^{T} \left\{ W_t^{-1} \widehat{SSR}_t^M \right\}^{1/2} \end{aligned} \tag{14}$$

where W_t measures the total number of contracts for which closing prices were available as of day t.

Likewise, we consider the average mean absolute valuation error (MAVE),

$$MAVE = T^{-1} \sum_{t=1}^{T} W_t^{-1} \sum_{\tau_t=\underline{\tau}_t}^{\overline{\overline{\tau}}_t} \sum_{K_{\tau_t}=\underline{K}_{\tau_t}}^{\bar{K}_{\tau_t}} \left|\widehat{C}(f_t) - C(f_t)\right|. \tag{15}$$

Finally, to assess the potential for overfitting, we report Akaike's information criterion (AIC),[22]

$$AIC = T^{-1} \sum_{t=1}^{T} \left[-2 \ln L(\hat{\theta}_t) + 2p\right], \tag{16}$$

[22]For simplicity, we call θ the vector of parameters to be estimated with reference to any possible pricing model. For instance, in the case of model (iv) $\theta = \alpha$, for the BL model $\theta = [\pi_t\ N_t\ m_t]'$, and for the variants on the BS model θ collects the volatility parameters.

where p is the number of parameters estimated for a given model and $L(\hat{\theta}_t)$ is its likelihood. This criterion trades off fit, as measured by $L(\hat{\theta}_t)$, against parsimony.

To compare the results across models, we report the proportion of days in the sample for which a given model displays the lowest value of a particular performance measure. This is also done to avoid situations where a certain model dominates most of the time, although its average performance is very poor as a result of a few days with extreme performance.

Table II contains the average RMSVE, AIC, and MAVE across the 251 days during the period Jan. 4, 1993 - Dec. 31, 1993. On average, and across the different measures of goodness-of-fit, our daily implementation of the BL model strongly outperforms the proposed alternatives (Panel A). It leads to an improvement of 58 cents over Black-Scholes and also outperforms the three-parameter versions of BS that let the implied volatility be a step-function over strike or maturity levels. The average value of the AIC is -2.03 under BL, much lower than the BS value of -0.32. Finally, the fit provided by a five-parameter smoothed implied volatility surface ($BS - spline$) is not particularly impressive and is severely penalized in terms of AIC. In relative terms, BL leads to the best goodness-of fit in over 80% of the sample days along the three criteria.

Following Dumas et al. (1998), these performance indicators are broken down according to both moneyness and the time-to-expiration of each traded contract. When the goodness-of-fit indicators are decomposed according to either moneyness or time-to-maturity (Panels B and C) the picture barely changes. While BL uniformly fits option prices across all traded maturities, Black-Scholes and its empirical variants do well for short-term options although their performance dramatically worsens with time-to-maturity. Overall, the fact that BL can fit option prices with average dollar errors between 30 and 40 cents for all traded strikes and maturities is quite remarkable.

Table III presents summary statistics for the parameter estimates obtained for the five models. While BS and $BS(m)$ are characterized by estimated volatilities that are quite stable over time, the same cannot be said for $BS(\tau)$ and $BS-spline$. For the latter two models the SPD implied by the estimates also shows significant instability. In the case of the BL model, we show results for two sets of risk aversion parameters, $\gamma = 0.9$ and $\gamma = 0.5$. In the first case, while the π_t and N_t estimates are quite stable, there is more variation in the estimate of the rate of information

arrival m_t.

To get a better impression of parameter stability for the different models, Figure 11 shows the time series of the estimated parameters for BS, $BS-spline$, and BL. The constant volatility assumption underlying BS is clearly rejected as volatility varies systematically over time. On some days $\hat{\sigma}_t$ changes by several percentage points. Likewise, the intercept and coefficients associated with K and K^2 in the $BS-spline$ model fluctuate widely from one day to the next. The sharp revisions in these parameters suggests significant instability in the implied volatility surface. The BL model generates relatively stable parameter estimates, with $\hat{\pi}_t$ steadily in the narrow range of [.52, .56], $\hat{N}_t$ normally around 180 trading days, and $\hat{m}_t$ most of the time in the interval [.3, .7]. The flow of information seems to peak between February and mid-April and slows down between June and September.[23] In general option markets display weak optimism ($\hat{\pi}_t \geq \frac{1}{2}$). For comparison the last row plots the estimates under BL when $\gamma = 0.5$. Interestingly, the parameter estimates are now very stable, suggesting again the relatively smooth evolution in the market's learning.

6.3. *Out-of-sample predictions*

The true economic and statistical value of a good option pricing model depends not on its in-sample fit, which will improve the more parameters are used, but on the precision with which the model predicts future option prices out-of-sample. We therefore extend the previous analysis and report statistical measures of out-of-sample prediction performance. Subsequently we report the hedging performance to measure the economic tracking errors of the models.

For each day t in the sample we estimate the relevant vector of parameters $\hat{\theta}_t$ in the manner described above and then use this to forecast the cross-section of option prices on the following trading day.[24]

Table IV provides summary statistics for the one-step-ahead prediction errors. Unsurprisingly, these are somewhat larger than the errors in Table II. Out-of-

[23] Since $\hat{m}$ is on average around 0.5, $\hat{N} = 170$ corresponds to about 340 trading days.

[24] When forecasting option prices one-step ahead, we follow Dumas et al. (1998) and condition on the stock index level and the risk-free rate at close of the following trading day. This way we test the predictive properties of the option pricing models independently of the ability to predict the future stock price.

sample, the BL model outperforms the proposed alternatives in the aggregate (Panel A). The root mean squared prediction error of the BL model is 54 cents against 1 dollar for Black-Scholes and 75 cents or higher for the empirical modifications to BS. In relative terms, BL is the best model out-of-sample in over 60% of the sample days. The superior forecasting performance of the BL model suggests that this model does not fit better purely as a result of daily variation in the learning parameters. Furthermore, the better aggregate predictive performance is not the result of a few outliers.[25] When the forecast indicators are decomposed according to either moneyness or time-to-maturity (Panels B and C) the picture remains the same. The BL model is far better than the other models across moneyness and maturity levels.

6.4. *Hedging performance*

A common economic measure of the precision of an option pricing model is its ability to assist in setting up a hedge against changes in the value of the underlying asset. Such a delta hedge is attained by selling short an amount $\widehat{\Delta}(f_t;\hat{\theta}_t)$ of the stock index, where Δ measures the sensitivity of the option price to the value of the underlying asset.[26] $\widehat{\Delta}(f_t;\hat{\theta}_t)$ is model dependent and an option pricing model performs well if it allows accurate calculation of $\widehat{\Delta}(f_t;\hat{\theta}_t)$ over time. Allowing for possible misspecification in the option model, the dollar return on the delta neutral position in excess of the risk-free rate can be shown to be:

$$\eta_{t+1}(\hat{\theta}_t) = \Delta C_{t+1} - \widehat{\Delta} C_{t+1}(\hat{\theta}_t)$$

which is the difference between the change in the actual price of the call minus the change predicted by the model based on the estimated vector of parameters $\hat{\theta}_t$. A good model should reduce the excess returns from hedging to zero since

[25] Some of the findings in Table IV differ from those reported in Dumas et al (1998) which can be explained by the difference in the two data sets. Dumas et al use options with moneyness between -10 and $+10$ % and between 6 and 100 days to expiration, while our data covers a wider range. If we were to apply the same criteria as Dumas et al. (1998) our data set would lose 28.5% of the original observations. They also perform their prediction and hedging tests on a weekly basis while ours are adopted on a daily basis.

[26] For simplicity we disregard the problem of the impossibility of buying or selling the S&P 500 index, which usually forces market operators to trade in S&P 500 futures.

$\eta_t(\hat{\theta}_t) \neq 0$ can result from misspecification. This is the same concept of error from a delta-hedging strategy used by Dumas et al. (1998, 2088-2089).

Table V reports statistics on hedging errors. When it comes to hedging a portfolio of options, the simple Black-Scholes model turns out to be a viable alternative, with a root mean squared hedging error of only 36 cents. This finding is consistent with those reported by Dumas et al. (1998) and Bakshi et al. (1997). Despite this good performance of the BS model, the BL model still outperforms all the proposed alternatives in the aggregate (Panel A).

In relative terms, BL provides the smallest average hedging errors in almost 50% of the sample days, although the performance of the Black-Scholes models clearly improves when a hedging rather than a statistical metric is used. Panels B and C show that BL improves hedging especially for portfolios that include long term, in-the-money calls or out-of-the-money puts which are important for risk management purposes.[27] Interestingly, the simple BS model outperforms the three-parameter BS models, a sign that BS provides a more robust and reliable performance than its complex generalizations.

6.5. *Learning dynamics with an expanding window*

We next infer the dynamics of learning from observed option prices by imposing the intertemporal restrictions on $\widehat{\pi}_t$ implied by the BL model. The model in section 4 assumes that beliefs evolve as follows:

$$\widehat{\pi}_{m(t+k)+j} = \widehat{\pi}_{m(t+k)+j-1} + \frac{I_{\{m(t+k)+j=g_h\}} - \widehat{\pi}_{m(t+k)+j-1}}{N_{m(t+k)+j}}. \tag{17}$$

If only one piece of news arrives every period, this intertemporal structure rules out large revisions in the belief parameters.

We impose these restrictions on half-monthly blocks of option prices, and thus consider a panel of option prices comprising the first 15 days of January, followed by a panel of the last 16 days of January and so on. For each block we solve the

[27] Using the West (1996) test, we found that the BL model outperforms the best alternative model at a statistically significant level in the out-of-sample forecasting experiment, while the better performance of the BL model was not significant at conventional levels in the hedging experiment.

program:

$$\min_{\{\pi_t\}_{t=1}^{T},N} \sum_{t=1}^{T} \sum_{\tau=\underline{\tau}_t}^{\bar{\tau}_t} \sum_{K_{\tau_t}=\underline{K}_{\tau_t}}^{\overline{K}_{\tau_t}} \left[C^{BL}(f_t, \pi_t, N; \gamma, \beta) - C(f_t)\right]^2$$

$$s.t. \quad \frac{\pi_t N}{N+1} \leq \pi_{t+1} \leq \frac{\pi_t N + 1}{N+1}$$

$$0 \leq \pi_t \leq 1 \quad t = 1, ..., T-1 \qquad N > 0 \tag{18}$$

The program is limited to blocks of 13-16 calendar days to avoid the curse of dimensionality implicit in this NLS estimation program: as $T \to \infty$, the dimension of the parameter vector $\theta = [\{\pi_t\}_{t=1}^{T}, N]'$ tends to infinity which makes a solution practically impossible. Limiting ourselves to a bi-weekly period means estimating a vector with 10-13 parameters to fit a sample of well over 500 observations on average. This setup therefore imposes much stronger restrictions on our model than in the previous sub-section and narrowly constrains the temporal dynamics of investors' beliefs with regard to the probability of good states.

Figure 12 reports results over the three-month period January - March 1993. The estimated sequence of beliefs $\{\hat{\pi}_t\}_{t=1}^{T}$ remains very stable and is confined to the interval [0.48,0.56]. The implied size of the data window used by the markets to estimate the unknown parameters of the fundamentals process is consistent with the levels uncovered in the cross-sectional exercises. Although there is a kink in the window size, N_t after the first half of January, from mid-January to mid-March the window expands smoothly over time in a manner fully consistent with our learning-based explanation of BS mispricings. While other even more restrictive tests can be designed, we consider these findings *prima facie* evidence that the BL model adequately fits the sequence of observed option prices.

7. **Conclusion**

This paper has proposed a simple stylized equilibrium model for asset prices under Bayesian learning and investigated its ability to explain a variety of empirical biases in the Black-Scholes option pricing model.

Despite its simplicity, the model with Bayesian learning proved to be able to match both skews in implied volatilities and a non-constant term-structure in implied volatility. Thus the model offers a new economic explanation of BS biases.

This is important because standard models in the literature that incorporate jump-diffusion and stochastic volatility effects, have been found by Das and Sundaram (1999) to be unable to correctly fit both stylized facts for plausible parameters values.

The parameters of the learning process implied by the cross-section of option prices appear reasonably stable over time. Episodes of sharp revisions in the beliefs implied by options prices are rare and tend to involve the precision of these beliefs rather than their level, which we find plausible. When we impose the intertemporal restrictions built in through our maintained Bayesian learning scheme, we find again that the estimated parameters are stable across time and do not deviate much from the results obtained through the daily cross-sections. This stability means that the model performs well in delta hedging experiments. It also provides a very different learning model than that recently proposed by David and Veronesi (1999). In their filtering algorithm the updated state probabilities have frequent discrete jumps from near zero to near one. Our results suggest that a smoother learning process may be better suited for out-of-sample predictions of the evolution in the cross-section of option prices.

References

[1] Abel A., 1988, "Stock Prices Under Time-Varying Dividend Risk. An Exact Solution in an Infinite-Horizon General Equilibrium Model", *Journal of Monetary Economics*, 22, 375-393.

[2] Aït-Sahalia, Y. and A. Lo, 1998, "Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices", *Journal of Finance*, 53, 499-547.

[3] Amin K., 1993, "Jump Diffusion Option Valuation in Discrete Time", *Journal of Finance,* 48, 1833-1863.

[4] Ball C., and W. Torous, 1985, "On Jumps in Common Stock Prices and Their Impact on Call Option Pricing", *Journal of Finance,* 40, 155-173.

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

[6] Bates, D., 1991, "The Crash of '87: Was It Expected? The Evidence from Options Markets", *Journal of Finance*, 46, 1009-1044.

[7] Bates, D., 1996, "Testing Option pricing Models", in G. S., Maddala and C. R. Rao, eds., *Handbook of Statistics, Vol. 15: Statistical methods in Finance*, North Holland, Amsterdam, 567-611.

[8] Bick A., 1990, "On Viable Diffusion Price Processes of the Market Portfolio", *Journal of Finance,* 45, 673-689.

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

[10] Bossaert, P., 1995, "The Econometrics of Learning in Financial Markets", *Econometric Theory* 11, 151-189.

[11] Bossaert, P., 1999, "Learning-Induced Price Volatility", Mimeo, Caltech.

[12] Campa J., and K. Chang, 1995, "Testing the Expectations Hypothesis on the Term Structure of Volatilities", *Journal of Finance*, 50, 529-547.

[13] Campbell, J., 1996, "Consumption and the Stock Market: Interpreting the International Experience", *NBER W.P. 5610*, June 1996.

[14] Chriss N., 1997, *Black-Scholes and Beyond. Option Pricing Models,* Irwin.

[15] Cox, J., S. Ross and M. Rubinstein, 1979, "Option Pricing : A Simplified Approach", *Journal of Financial Economics,* 7, 229-263.

[16] Das, S., and R. Sundaram, 1999, "Of Smiles and Smirks: A Term Structure Perspective", *Journal of Financial and Quantitative Analysis,* 34, 211-239.

[17] David, A., and P. Veronesi, 1999, "Option Prices with Uncertain Fundamentals: Theory and Evidence on the Dynamics of Implied Volatilities and Over-/Underreaction in the Options Market", mimeo.

[18] Duan, J.-C., 1995, "The GARCH Option Pricing Model", *Mathematical Finance,* 5, 13-32.

[19] Dumas, B., J. Fleming and R., Whaley, 1998, "Implied Volatility Functions: Empirical Tests", *Journal of Finance,* 53, 2059-2106.

[20] Guidolin M., and A. Timmermann, 1999, "Asset Prices on a Learning Path", *mimeo*, University of California, San Diego.

[21] Harrison, J. and D., Kreps, 1979, "Martingales and Arbitrage in Multiperiod Securities Markets", *Journal of Economic Theory,* 20, 381-408.

[22] He H., and Leland, H., 1993, "On Equilibrium Asset Price Processes", *Review of Financial Studies*, 6, 593-617.

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

[24] Hull, J. and A., White, 1987, "The pricing of Options on Assets with Stochastic Volatilities", *Journal of Finance,* 42, 281-300.

[25] Jackwerth, J. and M. Rubinstein, 1996, "Recovering probability Distributions from Option Prices", *Journal of Finance*, 51, 1611-1631.

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

[27] Leland, H., 1985, "Option Pricing and Replication with Transaction Costs", *Journal of Finance,* 40, 1283-1302.

[28] Lucas, R., 1978, "Asset prices in an Exchange Economy", *Econometrica,* 46, 1429-1445.

[29] Melino A., and S. Turnbull, 1990, "Pricing Foreign Currency Options with Stochastic Volatility", *Journal of Econometrics,* 45, 239-265.

[30] Merton, R., 1976, "Option Pricing When Underlying Stock returns are Discontinuous", *Journal of Financial Economics,* 3, 125-144.

[31] Pliska S., 1997, *Introduction to Mathematical Finance*, Blackwell Publishers.

[32] Rosenberg, J., and R. Engle, 1997, "Option Hedging Using Empirical Pricing Kernels", Discussion paper 97-20, UCSD.

[33] Rubinstein, M., 1985, "Nonparametric Tests of Alternative Option Pricing Models Using All Reported Trades and Quotes on the 30 Most Active {CBOE} Option Classes from August 23, 1976 Through August 31, 1978", *Journal of Finance,* 40, 455-480.

[34] Rubinstein, M., 1994, "Implied Binomial Trees", *Journal of Finance,* 49, 771-818.

[35] Stapleton, R., and M. Subrahmanyam, 1984, "The Valuation of Options When Asset Returns Are Generated by a Binomial Process", *Journal of Finance,* 39, 1525-1539.

[36] Timmermann, A., 1993, "How Learning in Financial Markets Generates Volatility and Predictability in Stock Prices", *Quarterly Journal of Economics*, 108, 1135-1145.

[37] Walters F., L. Parker, S. Morgan, and S. Deming, 1991, *Sequential Simplex Optimization*, CRC Press.

[38] West, K., 1996, "Asymptotic Inference about Predictive Ability". *Econometrica*, 64, 1067-1084.

[39] Wiggins, J., 1987, "Option Values Under Stochastic Volatility: Theory and Empirical Estimates", *Journal of Financial Economics,* 19, 351-372.

Appendix A

Proof of Proposition 2. We first show that adjusting $g_h^{(m)}$, $g_l^{(m)}$, $\pi^{(m)}$, $\rho^{(m)}$ as a function of the number of steps per period m, and choosing an appropriate coefficient of relative risk aversion γ gives the assumed annual risk-free rate r and dividend yield δ, as well as the correct annual mean and standard deviation of the dividend growth rate. $g_h^{(m)}$, $g_l^{(m)}$, $\pi^{(m)}$, and $\rho^{(m)}$ are functions of m, μ, σ, the risk-free rate r and the dividend yield δ. The restrictions on the process for dividend growth involving $g_h^{(m)}$, $g_l^{(m)}$, and $\pi^{(m)}$ are the same as in Cox et al. (1979, 246-251). $\rho^{(m)}$ and γ jointly depend on r and δ. A subjective discount rate and a coefficient of relative risk aversion can be found to keep the period $[t, t+\tau]$ risk-free and dividend rates constant per unit of calendar time and independent of m:

$$1 + r^{(m)} = \frac{1 + \rho^{(m)}}{(1 + g_l^{(m)})^{-\gamma} + \pi^{(m)} \left[(1 + g_h^{(m)})^{-\gamma} - (1 + g_l^{(m)})^{-\gamma} \right]}.$$

As $m \to \infty$, $\rho^{(m)}$ decreases since $(1+r)^{\frac{dt}{v}}$ goes to zero and $\{(1 + g_l^{(m)})^{-\gamma} + \pi[(1 + g_h^{(m)})^{-\gamma} - (1 + g_l^{(m)})^{-\gamma}]\}$ goes to one. The decrease in $r^{(m)}$ cancels out against the increase in v so the annual risk-free rate is constant as assumed. As for the dividend yield,

$$1 + \delta^{(m)} = \frac{(1+r)^{\frac{dt}{v}} \left\{ (1 + g_l^{(m)})^{-\gamma} + \pi^{(m)} \left[(1 + g_h^{(m)})^{-\gamma} - (1 + g_l^{(m)})^{-\gamma} \right] \right\}}{(1 + g_l)^{1-\gamma} + \pi^{(m)} \left[(1 + g_h^{(m)})^{1-\gamma} - (1 + g_l^{(m)})^{1-\gamma} \right]} = (1 + \delta)^{\frac{dt}{v}}$$

which over the period $[t, t+\tau]$, is $[(1+\delta)^{\frac{dt}{v}}]^v = (1+\delta)^{dt}$, or $(1+\delta)$ per year ($dt = 1$). Notice that γ does not need adjustment as a function of m since this parameters does not depend on time. For the proof of convergence of this model to BS, we refer to Cox et al. (1979, 246-251).

Proof of Proposition 4. The equations follow from the expression for the no-arbitrage price of a contingent claim that pays off $\max\left\{0, S_{t+v}^{BL,j} - K\right\}$ once the probabilities perceived on a learning path are used.

We only need to check that the probabilities in the expression for the SPD are risk-neutralized in all the underlying single period models associated with the

information structure. From the Euler equations under Bayesian learning, we know that $S_{t+k} = E\{Q_{t+k+1}(D_{t+k+1} + S_{t+k+1})|\hat{\pi}_{t+k}\}$. Dividing through by the price of the one-period zero coupon bond issued at time $t+k$ we have:

$$\frac{(1+\rho)S_{t+k}}{(1+g_l)^{-\gamma} + \widehat{\pi}_t\left[(1+g_h)^{-\gamma} - (1+g_l)^{-\gamma}\right]}$$

$$= E\left\{Q_{t+k+1}(D_{t+k+1} + S_{t+k+1})\frac{(1+\rho)}{(1+g_l)^{-\gamma} + \widehat{\pi}_{t+k}\left[(1+g_h)^{-\gamma} - (1+g_l)^{-\gamma}\right]}|\hat{\pi}_{t+k}\right\},$$

where we have used that a zero coupon bond has unit price at expiration. Let the discounted price and the discounted cumulative dividend process be

$$\begin{aligned} S^*_{t+k} &= \frac{(1+\rho)S_{t+k}}{(1+g_l)^{-\gamma} + \widehat{\pi}_{t+k}\left[(1+g_h)^{-\gamma} - (1+g_l)^{-\gamma}\right]}, \\ D^*_{t+k} &= \sum_{s=0}^{k} D_{t+s}\frac{(1+\rho)}{(1+g_l)^{-\gamma} + \widehat{\pi}_{t+s}\left[(1+g_h)^{-\gamma} - (1+g_l)^{-\gamma}\right]}. \end{aligned}$$

Adding D^*_{t+k} to both sides and using that

$$E\left\{Q_{t+k+1}\frac{(1+\rho)}{(1+g_l)^{-\gamma} + \widehat{\pi}_{t+k}\left[(1+g_h)^{-\gamma} - (1+g_l)^{-\gamma}\right]}|\hat{\pi}_{t+k}\right\} = 1,$$

we obtain:

$$\begin{aligned} S^*_{t+k} + D^*_{t+k} &= E\left[Q_{t+k+1}(S^*_{t+k+1} + D^*_{t+k+1})\right. \\ &\quad \left.\times\frac{(1+\rho)}{(1+g_l)^{-\gamma} + \widehat{\pi}_{t+k}\left[(1+g_h)^{-\gamma} - (1+g_l)^{-\gamma}\right]}|\widehat{\pi}_{t+k}\right\}, \end{aligned}$$

which demonstrates that the process $S^*_{t+k} + D^*_{t+k}$ is a martingale under the (conditional) probability measure

$$\begin{aligned} \widehat{P}\left\{S^j_{t+k+1}|\widehat{\pi}_{t+k}\right\} &= Q_{t+k+1}\frac{(1+\rho)}{(1+g_l)^{-\gamma} + \widehat{\pi}_{t+k}\left[(1+g_h)^{-\gamma} - (1+g_l)^{-\gamma}\right]} \\ &\quad \times P\left\{D^j_{t+k+1}|\widehat{\pi}_{t+k}\right\} \\ &= \beta\left(\frac{D^j_{t+k+1}}{D_{t+k}}\right)^{-\gamma}[1 + r^{BL}_{t+k}(\widehat{\pi}_{t+k})]P\left\{D^j_{t+k+1}|\widehat{\pi}_{t+k}\right\}. \end{aligned}$$

This represents the one-period risk neutral density. The corresponding state-price density is simply

$$\widetilde{P}\left\{S_{t+k+1}^{j}|\widehat{\pi}_{t+k}\right\}=\frac{1}{1+r_{t+k}^{BL}(\widehat{\pi}_{t+k})}\widehat{P}\left\{S_{t+k+1}^{j}|\widehat{\pi}_{t+k}\right\}$$
$$=\beta\left(\frac{D_{t+k+1}^{j}}{D_{t+k}}\right)^{-\gamma}\left[I_{\{j=1\}}\widehat{\pi}_{t+k}+(1-I_{\{j=1\}})(1-\widehat{\pi}_{t+k})\right].$$

Well-known results in, e.g., Pliska (1997) guarantee that the existence and uniqueness of the risk neutral measure in the underlying single-period models is sufficient for the existence and uniqueness of the risk neutral measure in the infinite horizon model. This risk-neutral measure can be found by 'pasting' together all the paths leading to a certain state $t+v$ periods ahead and exploiting the independence of the realizations of the dividend growth rate. For instance, $\widetilde{\Pr}\left\{S_{t+v}^{j}|n_t,N_t\right\}$ corresponds to the product of the state-price densities that a high dividend growth occurs j out of v times multiplied by the number of sample paths that can lead to this final outcome, $\binom{v}{j}$:

$$\widetilde{\Pr}_t\left\{S_{t+v}^{j}\right\}=\binom{v}{j}\prod_{k=1}^{v}\beta\left(\frac{D_{t+k+1}^{j}}{D_{t+k}}\right)^{-\gamma}\left[I_{\{j_k=1\}}\widehat{\pi}_{t+k}+(1-I_{\{j_k=1\}})(1-\widehat{\pi}_{t+k})\right]$$
$$=\binom{v}{j}\beta^{v}\frac{n_t...(n_t+j-1)(N_t-n_t)...(N_t+v-j-n_t+1)}{N_t(N_t+1)...(N_t+j-1)...(N_t+v)}\times$$
$$\times\prod_{k=1}^{v}\left(\frac{D_{t+k+1}^{j}}{D_{t+k}}\right)^{-\gamma}$$
$$=\beta^{v}\left(\frac{D_{t+v}^{j}}{D_t}\right)^{-\gamma}\binom{v}{j}\frac{\prod_{k=0}^{j-1}(n_t+k)\prod_{k=0}^{v-j-1}(N_t-n_t+k)}{\prod_{k=0}^{v-1}(N_t+k)}.$$

It is easily shown that

$$\frac{\prod_{k=0}^{j-1}(n_t+k)\prod_{k=0}^{v-j-1}(N_t-n_t+k)}{\prod_{k=0}^{v-1}(N_t+k)}=$$
$$=\frac{n_t...(n_t+j-1)(N_t-n_t)...(N_t+v-j-n_t+1)}{N_t(N_t+1)...(N_t+v)}.$$

This is the desired multiperiod risk neutral measure. Notice that the time t risk-neutral distribution of the time $t+v$ stock prices depends on the entire sequence of possible future probability beliefs.

Appendix B
Estimating learning parameters through the polytope method

Suppose we are interested in minimizing the sum of squared pricing errors computed across strikes, K_{τ_t}, and maturities, τ_t:

$$\min_{\pi_t, N_t, m_t} SSR(\theta_t) \equiv \sum_{\tau_t=\underline{\tau}_t}^{\overline{\overline{\tau}}_t} \sum_{K_{\tau_t}=\underline{K}_{\tau_t}}^{\overline{K}_{\tau_t}} \left[\epsilon(\tau_t, K_{\tau_t})\right]^2$$
$$0 \leq \pi_t \leq 1 \quad N_t > 0 \quad N_t \in \mathcal{N}, \quad m_t > 0.$$

Here the pricing error is defined as

$$\epsilon(\tau_t, K_{\tau_t}) \equiv C^{BL}(\tau_t, K_{\tau_t}, S_t, \pi_t, m_t, N_t) - C(\tau_t, K_{\tau_t}, S_t).$$

We accomplish this by a combination of the *polytope method* and a grid search over a restricted region of the parameter space:

$$0 \leq \pi_t \leq 1 \quad N_t > 0 \quad N_t \in \mathcal{N}, \quad m_t > 0$$

The polytope method is a multidimensional comparison method that first constructs a simplex in R^n (in our case $n = 3$, the dimension of θ_t).[28] The simplex comprises four vertices $\left\{\theta^a\ \theta^b\ \theta^c\ \theta^d\right\}$.[29] The initial simplex is composed of four vectors that lie on a three-dimensional plane. At each iteration the vertex that gives the highest value of the objective function is replaced with a new vertex that is likely to give a lower value.

[28] A more complete treatment of polytope methods can be found in Judd (1998) and Walters et al. (1991).

[29] We choose $\left\{\theta^a\ \theta^b\ \theta^c\ \theta^d\right\}$ such that:

$$\begin{aligned}
\theta^a &= [(\hat{\pi}_{t-1} - .05)\ (\hat{N}_{t-1} - 50)\ (\hat{m}_{t-1} + 0.3)]' \\
\theta^b &= [(\hat{\pi}_{t-1} - .05)\ (\hat{N}_{t-1} + 50)\ (\hat{m}_{t-1} - 0.3)]' \\
\theta^c &= [(\hat{\pi}_{t-1} + .05)\ (\hat{N}_{t-1} - 50)\ (\hat{m}_{t-1} + 0.3)]' \\
\theta^d &= [(\hat{\pi}_{t-1} + .05)\ (\hat{N}_{t-1} + 50)\ (\hat{m}_{t-1} - 0.3)]'
\end{aligned}$$

and thus make the starting values a function of the results obtained the day before.

A reason for the adoption of this comparison method is that one of the parameters, N_t, can only take positive, integer values. Furthermore, the layered objective function has systematic 'flats' and kinks since the total number of forward steps v_τ on the lattice is given by the integer part of τm. Gradient methods are not useful since there exists an infinite number of points in the parameter space where the objective function is not differentiable. Polytope methods do not impose smoothness conditions on the objective function and can handle simple discontinuities.

Two additional tests are conducted to see if the polytope search uncovers a local as opposed to a global minimum of the objective function. First, each day the polytope is started from a different initial simplex which is not a function of the estimation results from day $t-1$ and is chosen to be particularly wide.[30] Second, we supplement the combined polytope and grid search with a rough grid search and require that the mimimum sum of squared residuals over the grid exceeds the polytope solution.

If any of these conditions is not met, we resort to extensive grid search in order to obtain an optimal estimate of the parameters. Specifically, we implement a two-stage, three-layer grid search over the following wide region of the parameter space:

$$\Theta^{grid} = \{\theta : \pi \in [.4, .6]\,,\ N \in [100, 240],\ m \in [.2, 1.2]\}$$

We limit ourselves to values of π in the interval $[.4, .6]$ since we found that $\hat{\pi}_t$ never falls outside this region. $N \geq 250$ corresponds more or less to a constant volatility Black-Scholes model so we concentrate on cases where $N < 250$. $N < 100$, on the other hand, results in too strong skews and such values are therefore not considered. The first stage of the grid search involves 825 grid points, from which the three best estimates are selected prior to a more extensive neighborhood search around

[30] We use the following values for $\left\{\theta^a\ \theta^b\ \theta^c\ \theta^d\right\}$:

$$\begin{aligned} \theta^a &= [.40\ 20\ 1]',\ \theta^b = [.40\ 300\ .50]' \\ \theta^c &= [.60\ 20\ 1]',\ \text{and } \theta^d = [.60\ 300\ 0.50]'. \end{aligned}$$

$\hat{\theta}^{(1)}$:

$$\left\{\theta : \pi \in \left[\hat{\pi}^{(1)} - .02, \hat{\pi}^{(1)} + .02\right],\ N \in [\hat{N}^{(1)} - 15,\right.$$
$$\left.\hat{N}^{(1)} + 15],\ m \in [\hat{m}^{(1)} - .15, \hat{m}^{(1)} + .15]\right\}$$

In the second stage each search involves 3,225 grid points. This procedure thus searches over more than 10,000 distinct points in the parameter space. Up to this point we have performed two polytope optimizations (from two alternative starting simplices), and two grid searches (the second in two steps and two layers), so we have considerable confidence in the results.

Figure 1. Implied volatility as a function of moneyness for S&P 500 index options maturing in August 1993 - January 1994. Each symbol in the plots corresponds to a particular day in the sample where a given maturity was traded.

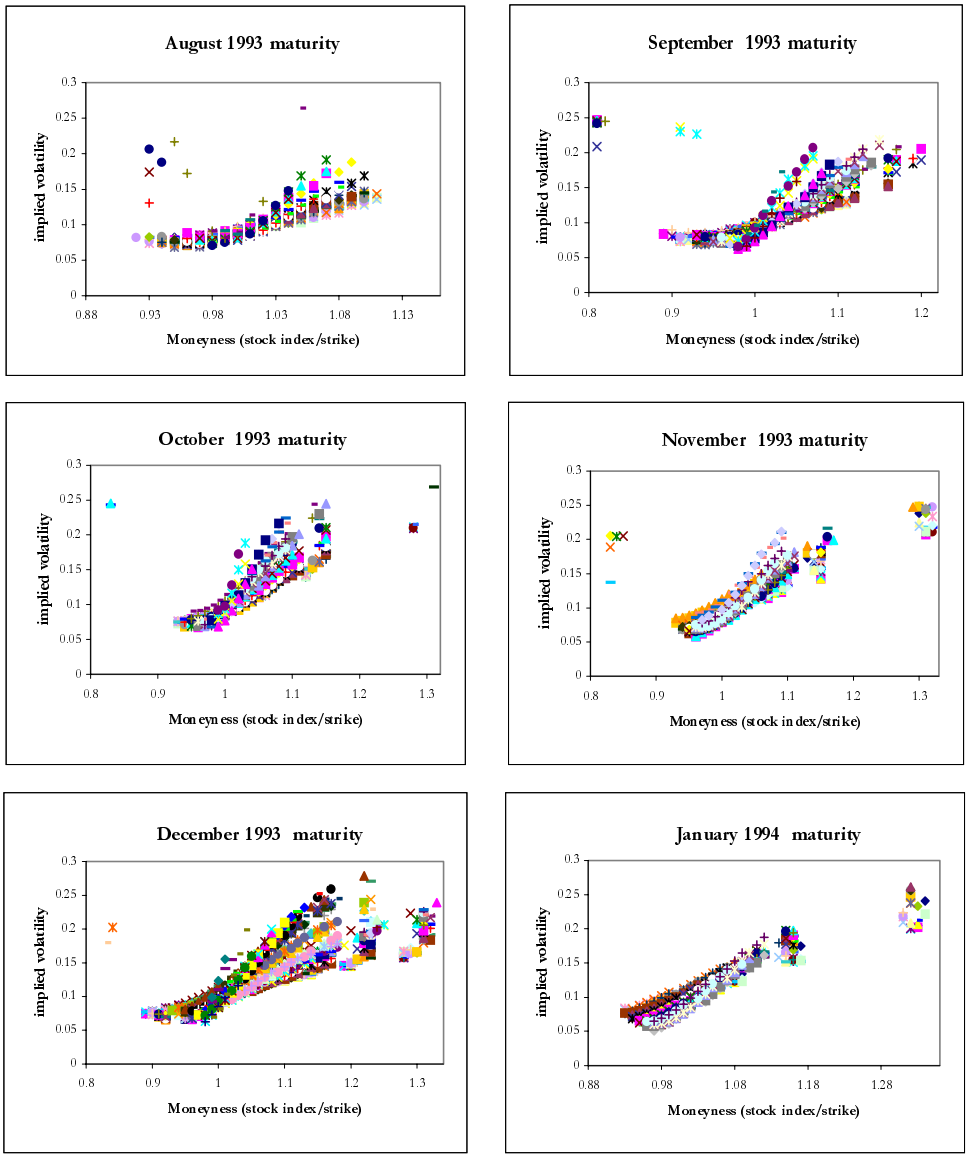


Figure 2. Term structure of implied volatility. The three graphs plot implied volatility as a function of maturity for S&P 500 index options, over the period Jan. 18 -25, 1993. Three different moneyness levels are used: 0.96 (in the money), 1 (at-the-money), and 1.04 (out-of-the-money).

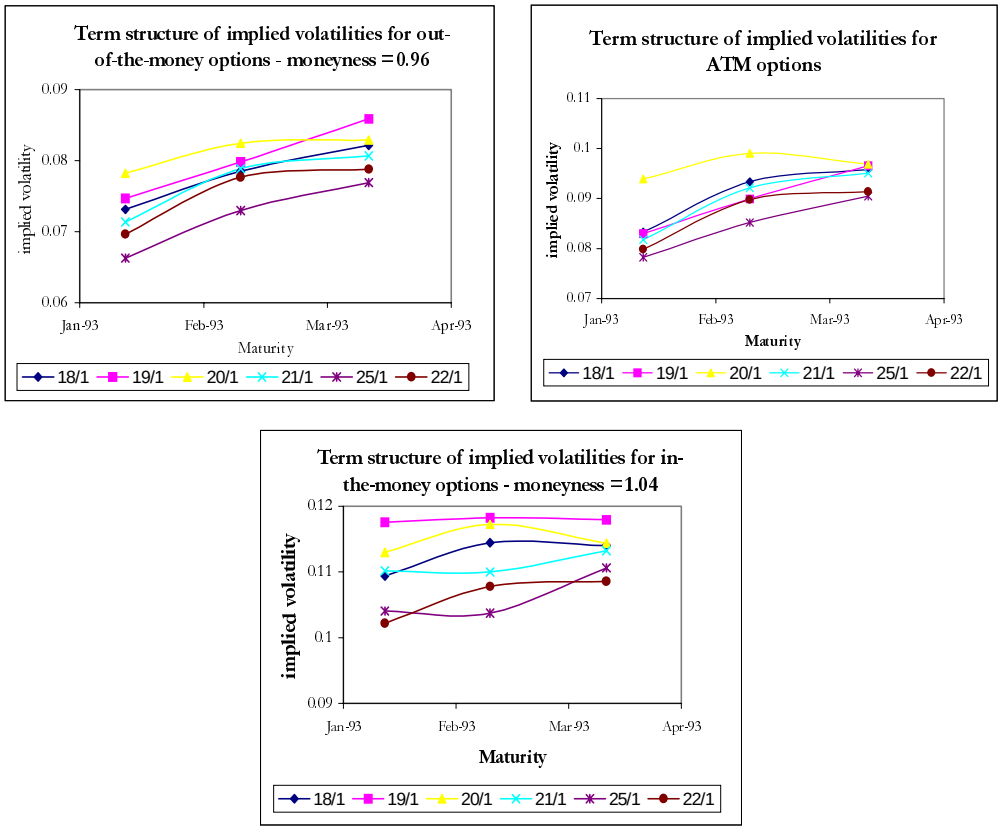
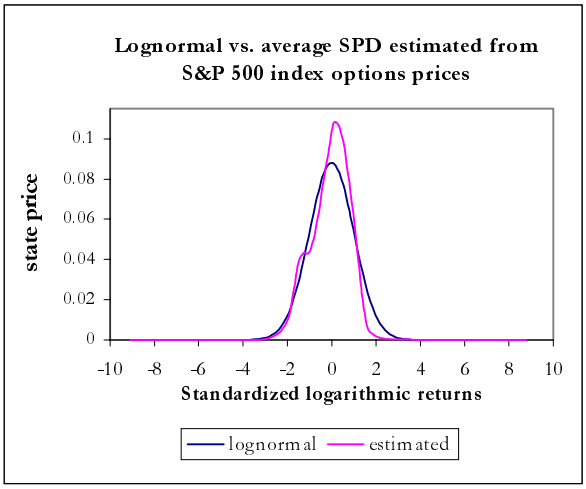
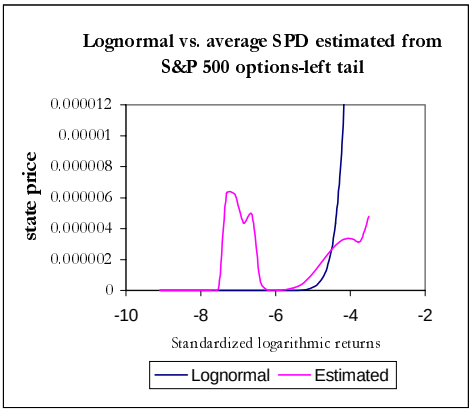


Figure 3. Average state-price density estimated from S&P 500 index options and the S&P 500 cash index over the period Jan. 4 - Dec. 31, 1993 compared to a lognormal SPD. The estimated SPD is the average of 334 SPDs obtained from options data with more than 50 calendar days to expiration using the nonparametric, implied binomial tree method of Rubinstein (1994) and Jackwerth and Rubinstein (1996). The objective function is the maximum smoothness function:

$$\sum_{j=0}^{v}(P_{j-1}-2P_j+P_{j+1})^2 \quad P_{-1}=P_{v+1}=0$$

For every day in the sample and for each cross section of contract, we minimize the objective function subject to martingale constraints on the option prices and the underlying index. The constraints are imposed by a penalty method that progressively raises the penalty parameter over various steps of the numerical optimization (see Judd (1998, 123-125)).





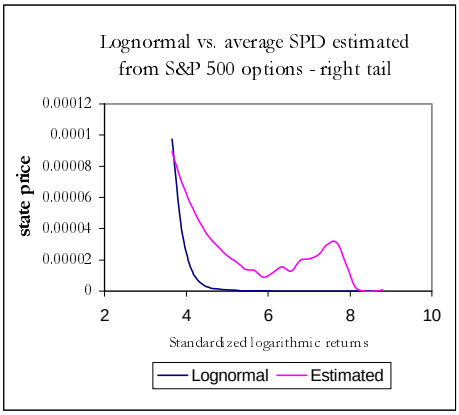


Figure 4. Difference between the price of a European call with 50 days to expiration (τ=50) calculated under full information and Bayesian learning. The assumed parameters are: m=1, g_h=0.00315, g_l=0.00314, π=0.519, ρ=0.02 (annual), γ=0.999. For BL prices, we take n_t=42 and N_t=80, implying a marginally biased initial belief $\hat{\pi}_t$ =0.525. The current price of the S&P 500 is $436.38, the closing price on Feb. 22, 1993.

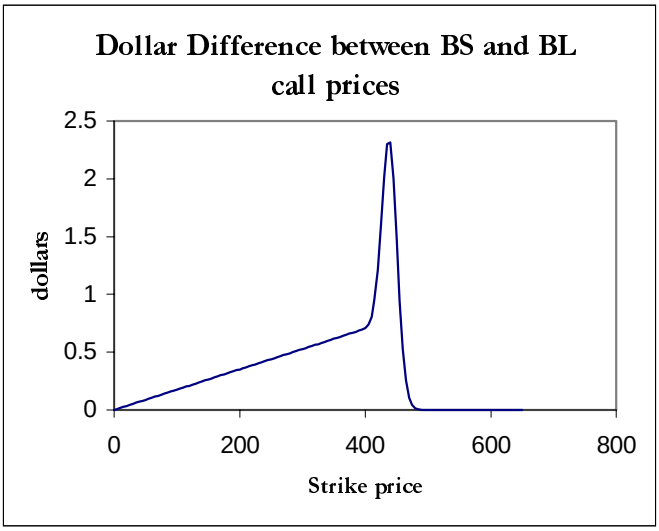


Figure 5. Implied Black-Scholes volatilities as a function of moneyness for a European call with 50 days to expiration (τ=50) calculated under full information. The parameters are set as follows: m=1, g_h=0.00315, g_l=0.00314, π=0.519, ρ=0.02 (annual), and γ=0.999. The price of the S&P 500 is $436.38, the closing price on Feb. 22, 1993. For comparison, the left panel reports implied volatilities as a function of moneyness for options expiring in April 1993.

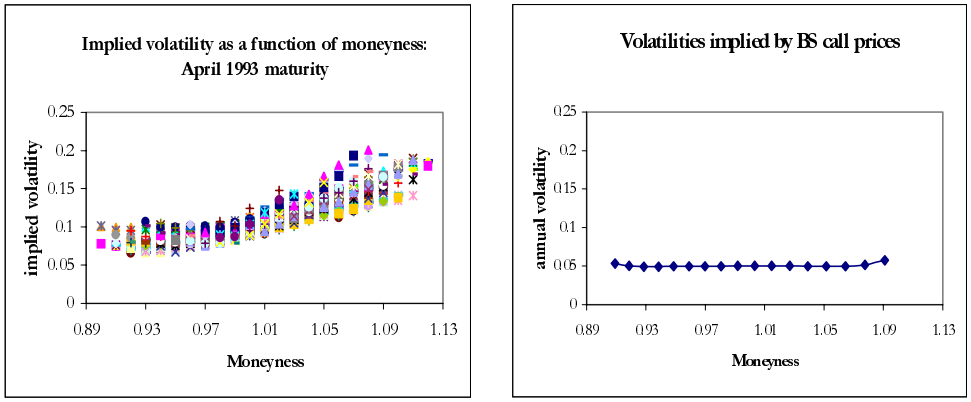


Figure 6. Implied Black-Scholes volatilities as a function of moneyness for a European call with 50 days to expiration (τ=50) calculated on a Bayesian learning path. The parameters are set as follows: m=1, g_h=0.00315, g_l=0.00314, π=0.519, ρ=0.02 (annual), γ=0.999, n_t=42, and N_t=80, implying a marginally biased initial belief $\hat{\pi}_t$=0.525. The price of the S&P 500 is $436.38, the closing price on Feb. 22, 1993. For comparison, the left panel reports implied volatilities as a function of moneyness for options expiring in April 1993.

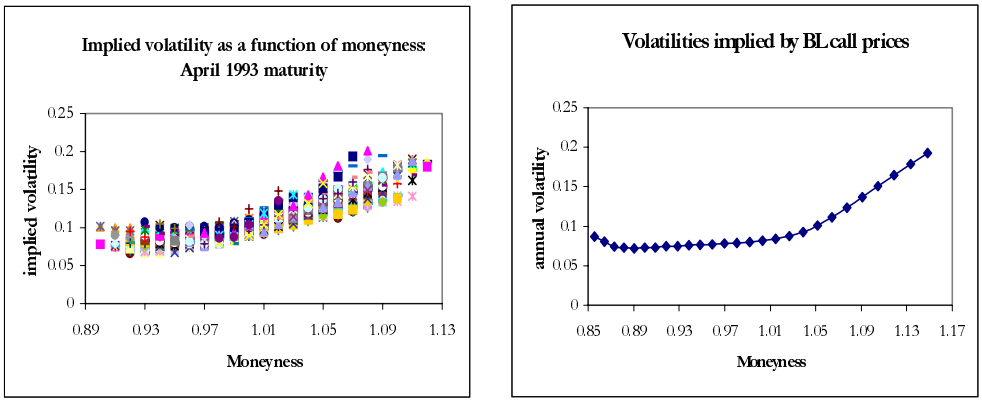


Figure 7. Implied Black-Scholes volatilities as a function of moneyness for a European call with 50 days to expiration (τ=50) calculated on a Bayesian learning path and using actual market data on February 22, 1993. The parameters are: m=1, g_h=0.00315, g_l=0.00314, π=0.519, ρ=0.02 (annual), γ=0.999, n_t=43, and N_t=80, implying an initial belief $\hat{\pi}_t$=0.538.

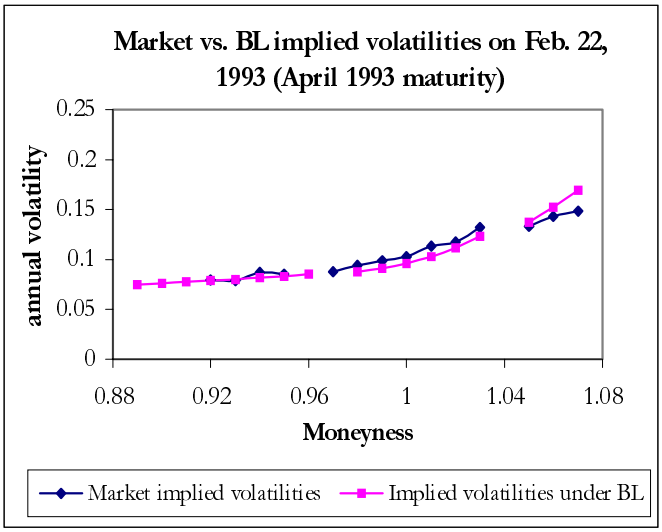


Figure 8. State-price densities for the 50 days ahead (τ=50) values of the S&P 500 index derived from a BS vs. a Bayesian learning model on Feb. 22, 1993. The parameters are set as follows: $m = 1$, $g_h = 0.00315$, $g_l = 0.00314$, $\pi = 0.519$, $\rho = 0.02$ (annual), $\gamma = 0.999$, $n_t = 42$, and $N_t = 80$, implying an initial belief $\hat{\pi}_t = 0.525$. The price of the S&P 500 is \$436.38, the closing price on Feb. 22, 1993. The third panel compares the BS with the BL state price density adjusting for differences in their respective supports.

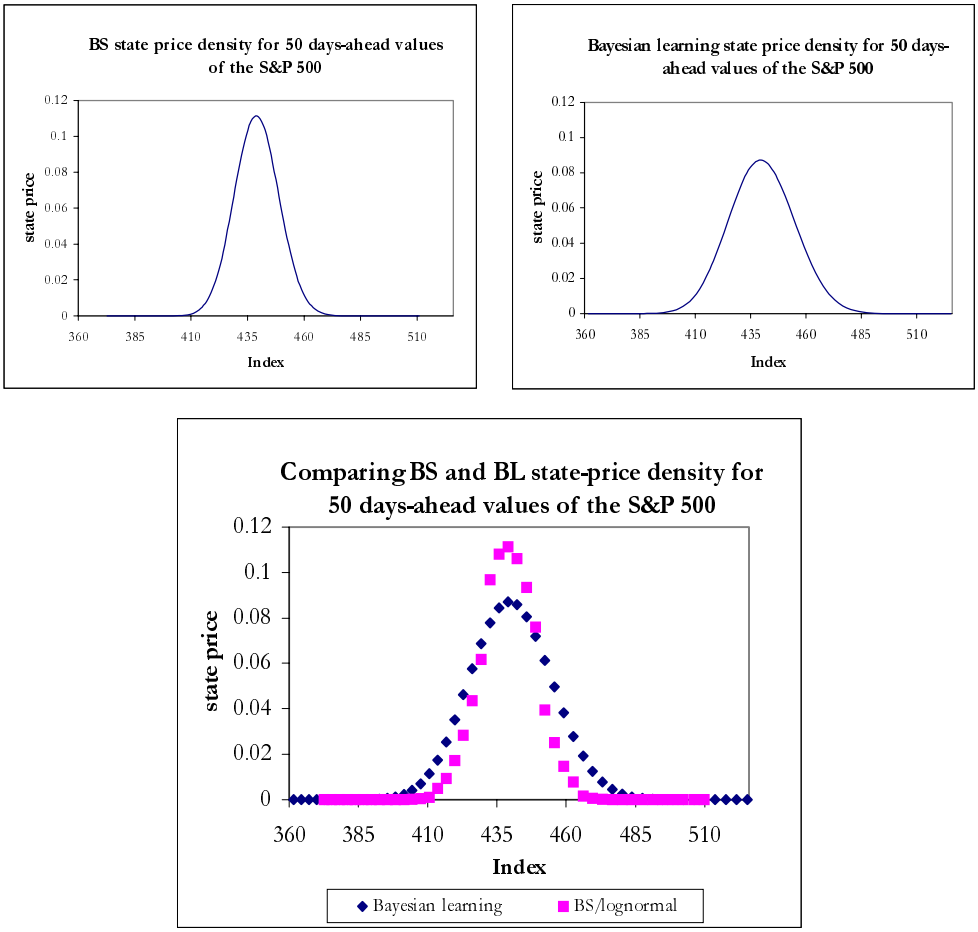


Figure 9. Implied volatility term structure plots under full information (BS) (squares) and with Bayesian learning (diamonds). We assume m=1 while the other 'deep' parameters (g_h, g_l, π, ρ, and γ) are adjusted according to Proposition 2 to ensure that the BS option price is an approximation to the Black-Scholes value with an annual risk-free rate of 4% and a dividend yield of 3%. In the case of BL prices, we take n_t=42 and N_t=80, implying an initial belief $\hat{\pi}_t$=0.525. The current price of the S&P 500 is taken to be 436.38 dollars, the closing price on Feb. 22, 1993. The first panel sets K=455, the second panel K=435, and the third K=420.

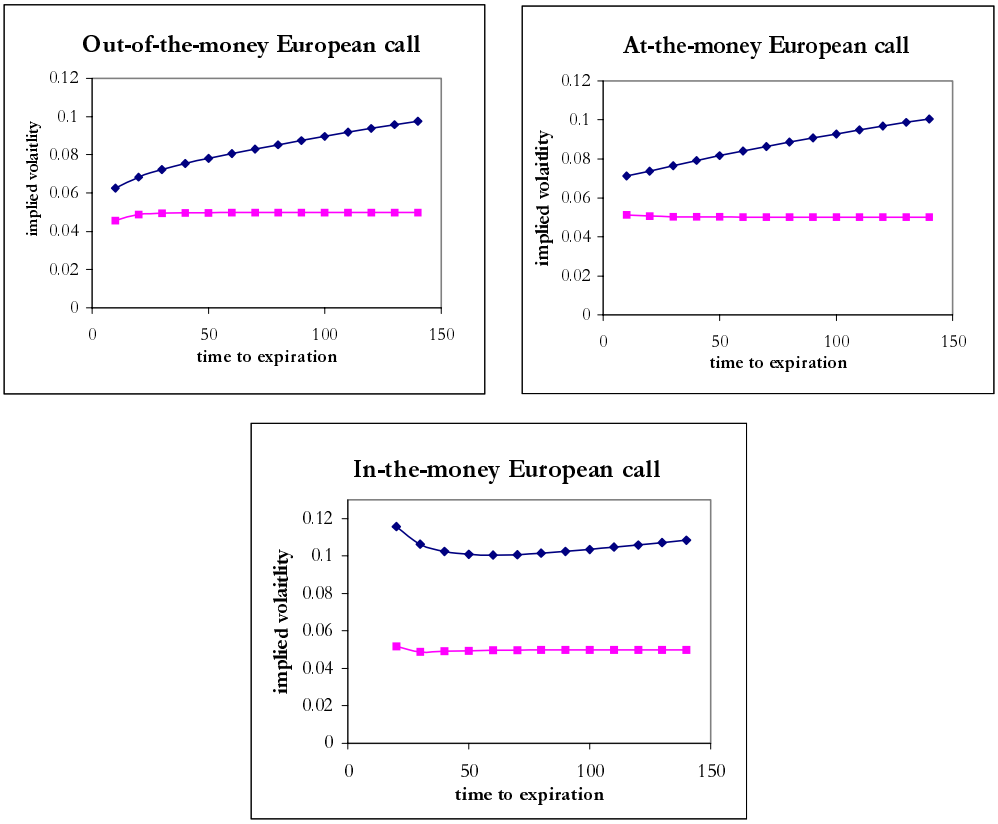


Figure 10. Weak learning effects — Plots of differences between call prices, implied volatilities as a function of the strike price, and 50-day-ahead state-price densities for the S&P 500 index on Feb. 22,1993 for the FIRE and BL asset pricing models. The first three graphs refer to a European call 50 days to expiration (τ=50). The fourth plot represents with squares the SPD calculated under full information rational expectations and with diamond the SPD calculated under Bayesian learning. The parameters are set as follows: m=1, g_h=0.00315, g_l=0.00314, π=0.519, ρ=0.02 (annual), γ=0.999, n_t=182, and N_t=350, implying an initial unbiased belief $\hat{\pi}_t$=0.52 ≅ π. The current price of the S&P 500 is $436.38 dollars, the closing price on Feb. 22, 1993.

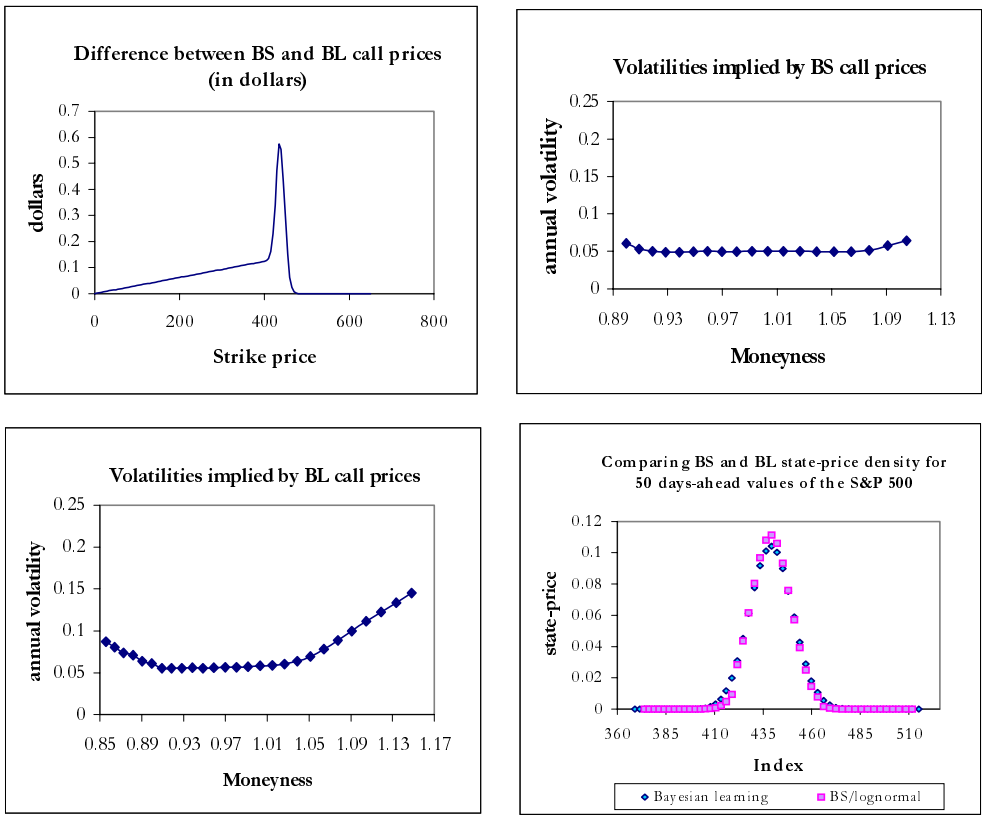


Figure 11. Parameter stability tests. Daily parameter estimates obtained by fitting to the cross section of S&P500 index option prices each day Black-Scholes (BS), BS with three moneyness parameters (BS(m)), BS with three maturity parameters (BS(τ)), a BS-spline model (BS-spline), and a Bayesian learning model (BL) under two sets of preference parameters. The sample consists of the daily data for the period January 4, 1993 - December 31, 1993.

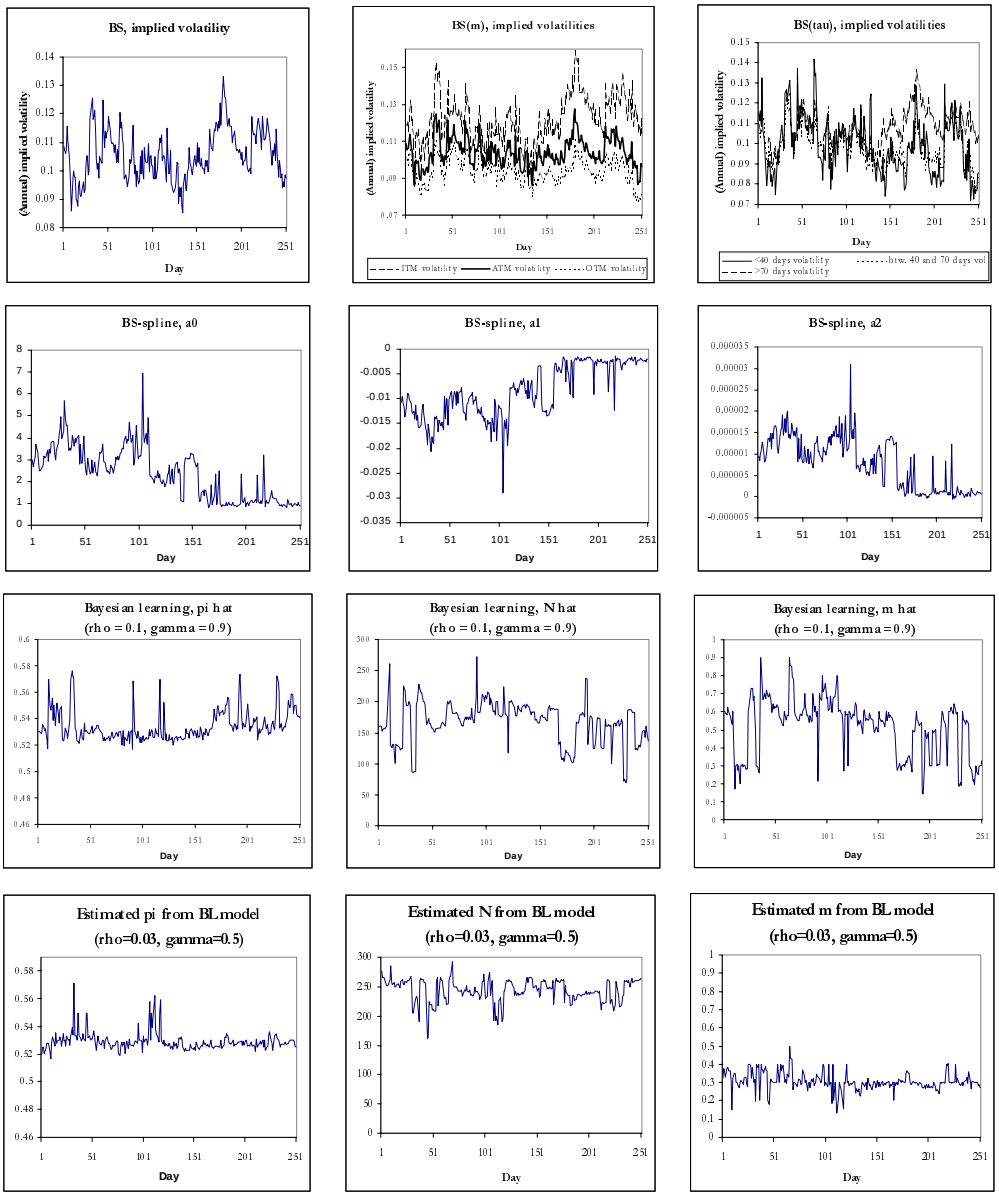


Figure 12. Time series for the probability of a high dividend growth state ($\hat{\pi}_t$) implied by observed S&P 500 index option prices during the period Jan. 4 - June. 30, 1993 (125 trading days). We also plot the precision of these beliefs (the number of previously observed states. $\{\hat{\pi}_t, \hat{N}_t, \hat{m}_t\}_{t=1}^{T}$ are estimated by solving the program:

$$\min_{\{\hat{\pi}_t\}_{t=1}^{T}, \hat{N}_1} \sum_{t=1}^{T} \left[C_t^{BL}(\tau_t, K_t, S_t, \pi_t, N_t, m_t) - C_t(\tau_t, K_t, S_t) \right]^2$$

$$s.t \quad \frac{\pi_t N_1}{N_1 + 1} \leq \pi_{t+1} \leq \frac{\pi_t N_1 + 1}{N_1 + 1}, \; 0 \leq \pi_t \leq 1 \; t \geq 1$$

$$N_1 > 0, \quad N_1 \in \mathrm{N}$$

i.e. by minimizing the (cross-sectional) sum of the squared pricing errors implied by the model under BL. The exercise assumes β =0.98 (annually) , CRRA-power preferences with constant relative risk aversion γ=0.9, and a volatility of fundamental news σ=5% (annualized). Future payoffs are discounted using the 3-month average T-bill rate measured on a daily basis.

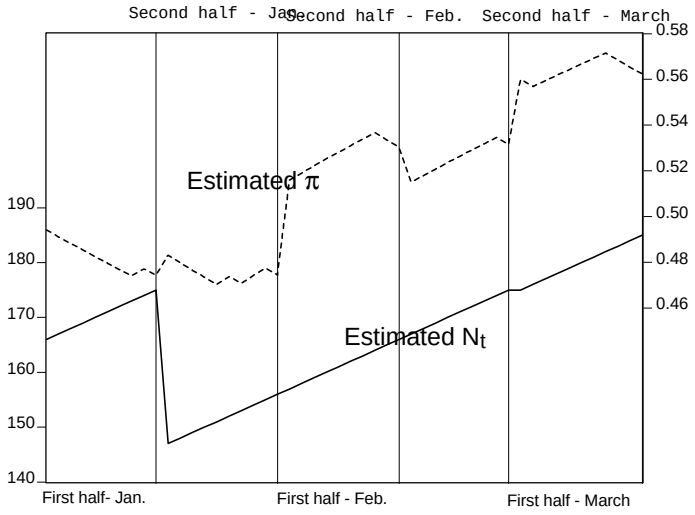


Table I

Implied State Price Densities.

The table reports the price of a state-contingent claim that pays out 1 dollar when (demeaned) S&P 500 returns fall below or above X standard deviations, calculated using the SPD estimated from option contracts with at least 50 calendar days to maturity. As a benchmark the table also reports the price of the same contingent claims based on a lognormal SPD. The SPDs have been demeaned and divided by $s\sqrt{t}$. The price of the contingent claims are calculated according to the formula:

$$Q^{-}=\sum_{j=0}^{v} l\left(\frac{1}{s\sqrt{t}}\left[\ln\left(\frac{S^{j}{}_{t+v}}{S_t}\right)-m\right]\right) I_{\left\{\frac{1}{s\sqrt{t}}\left[\ln\left(\frac{S^{j}{}_{t+v}}{S_t}\right)-m\right]<X\right\}} \text{ and } Q^{+}=\sum_{j=0}^{v} l\left(\frac{1}{s\sqrt{t}}\left[\ln\left(\frac{S^{j}{}_{t+v}}{S_t}\right)-m\right]\right) I_{\left\{\frac{1}{s\sqrt{t}}\left[\ln\left(\frac{S^{j}{}_{t+v}}{S_t}\right)-m\right]>X\right\}}$$

where the λ_js are the state-prices, I is the standard indicator function,

$$m=\sum_{j=0}^{v}\ln\left(\frac{S^{j}{}_{t+v}}{S_t}\right)\Pr ob\left\{S^{j}{}_{t+v}=u^{j}d^{v-j}S_t\right\},\ s^{2}=\sum_{j=0}^{v}\left[\ln\left(\frac{S^{j}{}_{t+v}}{S_t}\right)-m\right]^{2}\Pr ob\left\{S^{j}{}_{t+v}=u^{j}d^{v-j}S_t\right\}$$

and $\Pr ob\left\{S^{j}{}_{t+v}=u^{j}d^{v-j}S_t\right\}$ is a risk-neutral probability, based either on the normal density or implied by the options data. The estimated SPDs are either average or median state-price densities estimated from S&P 500 index options and the S&P 500 cash index over the period Jan. 4, 1993 - Dec. 31, 1993 using the nonparametric, implied binomial tree method of Rubinstein (1994) and Jackwerth and Rubinstein (1996). The objective is the maximum smoothness function:

$$\sum_{j=0}^{v}(P_{j-1}-2P_j+P_{j+1})^2 \quad P_{-1}=P_{v+1}=0$$

X	Average price		Median price	
	Lognormal SPD	Estimated SPD	Lognormal SPD	Estimated SPD
-7<	$3\cdot10^{-13}$	0.0000126	$8\cdot10^{-14}$	$1\cdot10^{-10}$
-6<	$1\cdot10^{-10}$	0.00002244	$1\cdot10^{-10}$	$4\cdot10^{-10}$
-5<	$2\cdot10^{-7}$	0.00002402	$2\cdot10^{-7}$	$9\cdot10^{-10}$
Q^{-} -4<	0.00001857	0.00003413	0.00001870	$2\cdot10^{-9}$
-3<	0.001505	0.00026	0.001515	$7\cdot10^{-9}$
-2<	0.01974	0.01264	0.01987	0.0000028
-1<	0.1744	0.1742	0.1747	0.1203
>1	0.1381	0.07197	0.1381	0.0679
>2	0.02334	0.00376	0.02334	0.00159
>3	0.000865	0.000849	0.000865	0.000173
Q^{+} >4	0.000235	0.000428	0.0000235	0.0000391
>5	$6\cdot10^{-8}$	0.000241	$6\cdot10^{-8}$	0.0000170
>6	$1\cdot10^{-10}$	0.000186	$1\cdot10^{-10}$	0.0000117
>7	$8\cdot10^{-14}$	0.000126	$8\cdot10^{-14}$	$7\cdot10^{-6}$

Table II

Average Dollar Valuation Errors Using the Bayesian Learning Model and Different Versions of Black-Scholes.

RMSVE is the root mean squared dollar valuation error averaged across all days and all traded contracts over the sample period January 4, 1993 - December 31, 1993. MAVE is the average of the mean absolute dollar evaluation error. AIC is the average of the Akaike Information Criterion. '% best-RMSVE' is the frequency of days, expressed as a ratio of the total number of days, on which a particular model has a lower daily RMSVE than any other. '% best-MAVE' and '% best-AIC' are defined analogously. BS is the Black-Scholes model, BS(τ) is a BS model that allows volatility to be a function of time-to-expiration, BS(m) is a BS model that allows volatility to be a function of moneyness, 'BS spline' fits a function of a polynomial in K (the strike price) and τ (time-to-maturity) to the implied volatility surface and then prices options using the fitted volatility levels. BL is a Bayesian learning model (assuming $\rho = 0.1$ and $\gamma = 0.9$) re-estimated on a daily basis. Moneyness is defined as $100\times(F/K-1)$, where F is the futures price for maturity identical to the option contract.

Panel A: Aggregate Results

Model	RMSVE	MAVE	AIC	% best-RMSVE	% best-AIC	% best-MAVE
BS	0.9463	0.7572	-0.3181	0	0	0
BS (τ)	0.8894	0.6998	-0.3743	0	0	0
BS(m)	0.6355	0.4972	-1.1377	15.93	15.93	17.13
BS spline	0.8927	0.6863	-0.1961	0	0	0
BL	0.3636	0.2890	-2.0253	84.07	84.07	82.87

Panel B: Results by Moneyness

	Moneyness (%)											
	Less than -2% (OTM)				-2% to +2% (ATM)				More than 2% (ITM)			
Model	RMSVE	MAVE	%best-RMSVE	%best-MAVE	RMSVE	MAVE	%best-RMSVE	%best-MAVE	RMSVE	MAVE	%best-RMSVE	%best-MAVE
BS	0.970	0.907	0.40	0.40	0.667	0.577	0.40	1.19	1.042	0.805	0	0
BS (τ)	0.984	0.868	0	0	0.516	0.407	21.03	17.86	0.981	0.782	0	0
BS(m)	0.402	0.324	41.27	42.46	0.622	0.520	6.75	6.75	0.700	0.562	20.32	19.12
BS spline	0.807	0.635	0.40	0.40	0.969	0.700	0	0	1.037	0.722	2.79	2.78
BL	0.322	0.272	57.94	56.75	0.327	0.269	71.83	74.21	0.389	0.318	76.89	78.09

Panel C: Results by Time-to-Maturity

	Days to Expiration											
	Less than 40				40 to 70				More than 70			
Model	RMSVE	MAVE	%best-RMSVE	%best-MAVE	RMSVE	MAVE	%best-RMSVE	%best-MAVE	RMSVE	MAVE	%best-RMSVE	%best-MAVE
BS	0.524	0.446	3.04	3.35	0.675	0.554	0	0	1.019	0.814	0	0
BS (τ)	0.425	0.365	10.98	12.80	0.521	0.442	2.45	2.86	0.951	0.735	0	0
BS(m)	0.449	0.383	23.78	27.13	0.520	0.427	19.18	20.41	0.719	0.573	11.21	12.46
BS spline	0.422	0.373	29.27	25.00	0.548	0.462	8.98	7.35	0.840	0.655	1.87	1.87
BL	0.395	0.336	32.93	31.71	0.347	0.283	69.39	69.39	0.393	0.317	86.92	85.67

Table III

Summary Statistics for Parameter Estimates.

The table reports summary statistics from fitting five different models to the cross-section of S&P 500 index option prices over the period Jan. 4 - Dec. 31, 1993. BS is the Black-Scholes model, BS (τ) is a BS model that allows volatility to be a function of time-to-expiration, BS(m) is a BS model that allows volatility to be a function of moneyness, 'BS spline' that fits a function of a polynomial in K (the strike price) and τ (time-to-maturity) to the implied volatility surface and then prices options using the fitted volatility levels. BL is a Bayesian learning model (assuming a fundamental volatility of 5% per year) re-estimated on a daily basis.

Model	Coefficient estimate	Mean	Median	Standard Dev.
BS	σ	0.1056	0.1050	0.0086
BS(*m*)	σ_{OTM}	0.0938	0.0935	0.0067
	σ_{ATM}	0.1038	0.1030	0.0078
	σ_{ITM}	0.1211	0.1200	0.0130
BS(*t*)	σ_{short}	0.0990	0.0985	0.5414
	σ_{medium}	0.0943	0.0970	0.0139
	σ_{long}	0.1089	0.1085	0.0216
BS-spline	α_0	2.3546	2.4126	1.1949
	α_1	-0.0085	-0.0091	0.0054
	α_2	7.77E-06	8.35E-06	6.07E-06
	α_3	-0.0047	-0.0034	0.0039
	α_4	1.03E-05	7.42E-06	8.6E-06
Bayesian learning (ρ=0.1, γ=0.9)	π_t	0.5343	0.5312	0.0115
	N_t	167.03	173.00	33.63
	M_t	0.5166	0.5556	0.1590
Bayesian learning (ρ=0.03, γ=0.5)	π_t	0.5292	0.5280	0.0065
	N_t	243.31	244.00	19.43
	m_t	0.3055	0.3000	0.0494

Table IV

Average Dollar Prediction Errors Using the Bayesian Learning Model and Different Versions of Black-Scholes formula.

RMSPE is the root mean squared dollar prediction error averaged across all days and traded contracts in the sample period January 4, 1993 - December 31, 1993. MAPE is the average of the mean absolute dollar prediction error. '% best-RMSPE' is the frequency of days, expressed as a ratio of the total number of days, on which a particular model has a lower daily RMSPE than any other. '% best-MAPE' is defined analogously. BS is the Black-Scholes model, BS(τ) is a BS model that allows volatility to be a function of time-to-expiration, BS(m) is a BS model that allows volatility to be a function of moneyness, 'BS spline' fits a function of a polynomial in K (the strike price) and τ (time-to-maturity) to the implied volatility surface and then prices options using the fitted volatility levels. BL is a Bayesian learning model (assuming $\rho = 0.1$ and $\gamma = 0.9$) re-estimated on a daily basis. Moneyness is defined as $100\times(F/K-1)$, where F is the futures price with maturity identical to the option contract.

Panel A: Aggregate Results

Model	RMSPE	MAPE	% best-RMSPE	% best-MAPE
BS	1.0079	0.8005	0.40	0.40
BS (τ)	1.0306	0.7897	1.20	0.80
BS(m)	0.7489	0.5882	22.00	25.60
BS spline	0.9164	0.7054	10.40	10.40
BL	0.5422	0.4422	66.00	62.80

Panel B: Results by Moneyness

	Moneyness (%)											
	Less than -2% (OTM)				-2% to +2% (ATM)				More than 2% (ITM)			
Model	RMSPE	MAPE	%best-RMSPE	%best-MAPE	RMSVE	MAVE	%best-RMSPE	%best-MAPE	RMSVE	MAVE	%best-RMSPE	%best-MAPE
BS	1.003	0.939	4.80	5.97	0.789	0.677	9.13	6.72	1.064	0.828	3.60	3.20
BS (τ)	1.068	0.930	2.00	1.20	0.985	0.625	7.94	14.62	1.044	0.823	9.20	10.80
BS(m)	0.476	0.383	24.00	1.04	0.753	0.626	6.74	9.49	0.824	0.667	16.40	16.80
BS spline	0.745	0.643	24.00	3.98	0.901	0.693	9.13	11.46	0.956	0.745	14.00	14.80
BL	0.404	0.347	45.20	47.81	0.496	0.424	67.06	57.71	0.599	0.512	56.80	54.40

Panel C: Results by Time-to-Maturity

	Days to Expiration											
	Less than 40				40 to 70				More than 70			
Model	RMSPE	MAPE	%best-RMSPE	%best-MAPE	RMSVE	MAVE	%best-RMSPE	%best-MAPE	RMSVE	MAVE	%best-RMSPE	%best-MAPE
BS	0.557	0.475	2.76	3.35	0.728	0.595	0.81	0.82	1.065	0.847	0	0
BS (τ)	0.497	0.424	12.58	16.16	0.635	0.519	8.54	9.43	1.067	0.810	0.63	1.25
BS(m)	0.494	0.425	24.23	25.30	0.605	0.504	24.39	25.41	0.804	0.642	17.19	20.00
BS spline	0.447	0.395	28.22	30.18	0.566	0.478	19.51	19.67	0.861	0.861	11.88	14.06
BL	0.470	0.407	32.82	25.00	0.478	0.398	45.93	44.67	0.549	0.450	70.31	64.69

Table V

Average Dollar Hedging Errors Using the Bayesian Learning Model and Different Versions of Black-Scholes.

RMSHE is the root mean squared dollar hedging error averaged across all days and all traded contracts in the sample period January 4, 1993 - December 31, 1993. MAHE is the average of the mean absolute dollar hedging error. '% best-RMSHE' is the frequency of days, expressed as a ratio of the total number of days, on which a particular model has a lower daily RMSHE than any other model. '% best-MAHE' is defined analogously. BS is the Black-Scholes model, BS (τ) is a BS model that allows volatility to be a function of time-to-expiration, BS(m) is a BS model that allows volatility to be a function of moneyness, 'BS spline' fits a function of a polynomial in K (the strike price) and τ (time-to-maturity) to the implied volatility surface and then prices options using the fitted volatility levels. BL is a Bayesian learning model (assuming ρ = 0.1 and γ = 0.9) re-estimated on a daily basis. Moneyness is defined as $100\times(F/K-1)$, where F is the futures price with maturity identical to the option contract.

Panel A: Aggregate Results

Model	RMSHE	MAHE	% best-RMSHE	% best-MAHE
BS	0.3626	0.2961	3.60	3.20
BS (τ)	0.3513	0.2855	19.20	20.80
BS(m)	0.3760	0.2839	14.80	15.20
BS spline	0.4237	0.3505	12.80	11.60
BL	0.3047	0.2416	49.60	49.20

Panel B: Results by Moneyness

	Moneyness (%)											
	Less than -2% (OTM)				-2% to +2% (ATM)				More than 2% (ITM)			
Model	RMSHE	MAHE	%best-RMSHE	%best-MAHE	RMSHE	MAHE	%best-RMSHE	%best-MAHE	RMSHE	MAHE	%best-RMSHE	%best-MAHE
BS	0.253	0.215	3.60	5.60	0.279	0.229	4.80	9.20	0.427	0.372	3.20	2.40
BS (τ)	0.266	0.228	10.80	11.60	0.250	0.205	47.60	45.20	0.412	0.356	17.20	17.20
BS(m)	0.195	0.155	43.20	42.0	0.359	0.287	13.60	9.60	0.415	0.342	14.00	14.80
BS spline	0.257	0.217	26.00	26.8	0.371	0.317	13.60	14.40	0.491	0.442	10.00	8.80
BL	0.230	0.189	16.40	14.0	0.294	0.243	20.40	21.60	0.329	0.272	55.60	56.80

Panel C: Results by Time-to-Maturity

	Days to Expiration											
	Less than 40				40 to 70				More than 70			
Model	RMSPE	MAPE	%best-RMSPE	%best-MAPE	RMSVE	MAVE	%best-RMSPE	%best-MAPE	RMSVE	MAVE	%best-RMSPE	%best-MAPE
BS	0.370	0.318	5.21	5.83	0.347	0.286	4.07	4.51	0.354	0.292	4.69	4.69
BS (τ)	0.343	0.296	29.14	31.90	0.335	0.276	26.83	25.82	0.349	0.285	18.44	18.75
BS(m)	0.368	0.310	11.35	11.04	0.353	0.274	13.01	15.16	0.378	0.285	15.94	16.25
BS spline	0.387	0.342	15.64	15.64	0.392	0.332	9.76	11.48	0.418	0.354	13.44	11.86
BL	0.351	0.306	38.65	35.58	0.293	0.237	45.53	43.03	0.304	0.244	47.50	48.44