

**Option Prices under Bayesian Learning:  
Implied Volatility Dynamics and Predictive  
Densities**

**By**

**Massimo Guidolin and Allan Timmermann**

**DISCUSSION PAPER 397**

**November 2001**

---

**FINANCIAL MARKETS GROUP**  
AN ESRC RESEARCH CENTRE

---

**LONDON SCHOOL OF ECONOMICS**



**Any opinions expressed are those of the author and not necessarily those of the Financial Markets Group.**

ISSN 0956-8549-397

# Option Prices under Bayesian Learning: Implied Volatility Dynamics and Predictive Densities\*

Massimo Guidolin<sup>†</sup>  
University of Virginia

Allan Timmermann  
University of California, San Diego

JEL codes: G12, D83.

## 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 and a GARCH model.

---

\*We wish to thank four anonymous referees for their extensive and thoughtful comments that greatly improved the paper. We also thank Alexander David, José Campa, Bernard Dumas, Wake Epps, Stewart Hodges, Claudio Michelacci, Enrique Sentana and seminar participants at Bocconi University, CEMFI, University of Copenhagen, Econometric Society World Congress in Seattle, August 2000, the North American Summer meetings of the Econometric Society in College Park, June 2001, the European Finance Association meetings in Barcelona, August 2001, Federal Reserve Bank of St. Louis, INSEAD, McGill, UCSD, Université de Montreal, and University of Virginia for discussions and helpful comments.

<sup>†</sup>Correspondence to: Massimo Guidolin, Department of Economics, University of Virginia, Charlottesville - 114 Rouss Hall, Charlottesville, VA 22903. Tel: (434) 924-7654; Fax: (434) 982-2904; e-mail: mg8g@virginia.edu

## 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 GARCH models (Duan (1995), Heston and Nandi (2000)); 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,<sup>1</sup> the only other studies that explicitly consider its derivative pricing implications appear to be David and Veronesi (1999) and Yan (2000). In an important contribution, David and Veronesi (1999) have investors facing a filtering problem which leads them to update the probability of which of two states fundamentals are currently in. In their model, shocks to fundamentals are drawn from a non-Gaussian mixture distribution and hence the BS model is not obtained in the limit even when there is full information about the state. In contrast, we assume that fundamentals follow a binomial process, whose limit is Gaussian. This allows us to isolate the effect of

---

<sup>1</sup>E.g., Bossaerts (1995, 1999) and Timmermann (1993).

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 BS. The importance of such exercises for evaluating different option pricing models has recently been emphasized by Dumas et al. (1998). In Yan (2000) the unobservable rate of growth of dividends follows a mean-reverting diffusion process. Again shocks to fundamentals are drawn from a non-Gaussian mixture distribution, so that BS does not obtain under full information,<sup>2</sup> which makes it more difficult to extract the effects of recursive learning on option prices. Yan presents calibrations which indicate that the model can generate a wide variety of implied volatility surfaces but presents no empirical estimates of the model's fit and does not explore the possibility of implying beliefs from observed asset prices.

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.<sup>3</sup> 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 a good in-sample fit, and especially an excellent out-of-sample performance in addition to generating smaller out-of-sample hedging errors than the BS model and various alternatives proposed in the literature (Dumas et al.'s (1998) ad hoc strawman and Heston and Nandi's (2000) option NGARCH).

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

---

<sup>2</sup>However Yan (2000, p. 22) shows that under log-utility a special version of BS that incorporates stochastic interest rates and volatility obtains.

<sup>3</sup>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.

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 weekly S&P 500 index option prices<sup>4</sup> covering a five and a half year period from June 1988 through December 1993 (a total of 30,461 observations) and is identical to the one used in Dumas et al. (1998).<sup>5</sup> We explicitly take into account that the index pays out daily dividends and adjust the reported index level by subtracting the present value of the cash dividends to be paid during the life of the option. We follow the same procedures for filtering the data as in Ait-Sahalia and Lo (1998) and thus differ slightly from Dumas et al. (1998) in the application of some exclusion criteria. First, we eliminate options with fewer than 6 and more than 200 days to expiration.<sup>6</sup> Second, since it is well known that 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 are translated into call prices using put-call parity. All information contained in liquid put option prices has thus been extracted and converted into call prices. The remaining put options are discarded from the data set without loss of information. The application of these criteria reduces the size of the sample to a total of 9,679 observations.<sup>7</sup>

### 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

---

<sup>4</sup>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.

<sup>5</sup>We thank Bernard Dumas for making this data set available to us. Data are sampled on Wednesdays and only options with bid/ask price quotes between 2:45 and 3:15 p.m. (CST) are used. Wednesdays are chosen in order to minimize the incidence of the number of holidays. Option prices correspond to bid/ask midpoints. The risk-free rate is proxied by the average of bid and ask discounts reported in the *Wall Street Journal*. Daily cash dividends are collected from the *S&P 500 Information Bulletin*.

<sup>6</sup>Dumas et al. (1998) exclude all contracts with more than 100 days to expiration.

<sup>7</sup>One observation had to be excluded since it violated the lower bound condition and led to a negative implied volatility.

the six December maturities represented in our data sample.<sup>8</sup> 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 lognormal. Figure 3 shows the SPD inferred from option contracts with at least 7 calendar days to expiration and averaged across calendar days and maturities (a total of 765 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, cf. Bates (1991). Therefore we plot in the bottom of Figure 3 the estimated tail behavior of 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 1 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.

---

<sup>8</sup>The finding of skews in implied volatility is generic and holds across different maturities.

### 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  $\{D_{t+k}\}_{k=1}^{\infty}$ .<sup>9</sup> 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  $\pi$  or  $g_l$  with probability  $1 - \pi$ :

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

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  $\frac{dt}{v} \rightarrow 0$ , the distribution of dividends converges weakly to a geometric Brownian motion with constant drift and diffusion  $(\mu, \sigma)$ .

To price assets we assume a perfect capital market. There are unlimited short sales possibilities, 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.

---

<sup>9</sup>Dividends in our model really refer to all information on (cash and non cash) earnings produced by companies in the stock index. David and Veronesi (2000) compare beliefs implicit in observed option prices and beliefs filtered out by a sensible definition of fundamentals (real earnings). They find that the two series are quite different, a sign that a much more complex notion of fundamentals than earnings or dividends is likely to be used by investors.

The representative investor has power utility

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

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, see Abel (1988).<sup>10</sup> 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:

$$\begin{aligned} & \max_{\{Z_{t+k}, w_{t+k}^s, w_{t+k}^b\}_{k=0}^{\infty}} E_t \left[ \sum_{k=0}^{\infty} \beta^k u(Z_{t+k}) \right] \\ & \text{s.t. } Z_{t+k} + w_{t+k}^s S_{t+k} + w_{t+k}^b P_{t+k} = w_{t+k-1}^s (S_{t+k-1} + D_{t+k-1}) + w_{t+k-1}^b, \end{aligned} \quad (3)$$

where  $\beta = \frac{1}{1+\rho}$ ,  $\rho$  is the rate of impatience and  $w_{t+k}^s$  and  $w_{t+k}^b$  represent the number of stocks and bonds in the agent's portfolio as of period  $t+k$ .<sup>11</sup>

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

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

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 (2001) price the underlying stock and risk-free bond in this setting subject to a transversality condition.<sup>12</sup> For convenience, we state asset prices using the

---

<sup>10</sup>The restriction  $\gamma \leq 1$  is not as implausible as it might appear in the light of the voluminous literature that has either estimated or used values of  $\gamma$  well above 1 in order to match known stylized facts concerning asset prices and returns. The existing literature has in general performed these empirical exercises under the assumption of full information rational expectations. Guidolin and Timmermann (2000) show that on a Bayesian learning path, many stylized facts concerning asset prices (such as high volatility and serial correlation) can easily be rationalized for  $\gamma \leq 1$ . A referee also pointed out that in the framework of Abel (1999), an economy with leverage can be well approximated by our endowment economy with the coefficient of relative risk aversion replaced by the ratio between  $\gamma$  and the leverage index.

<sup>11</sup>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.

<sup>12</sup>The transversality condition is:

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



transformed parameters  $g_l^* = (1 + g_l)^{1-\gamma} - 1$  and  $g_h^* = (1 + g_h)^{1-\gamma} - 1$ .

**Proposition 1 (Guidolin and Timmermann (2001))** *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

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

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.<sup>13</sup> 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 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 per unit of calendar time goes to infinity. Let  $r$  be the risk-free rate,  $\delta$  the dividend yield, and  $\mu$  and  $\sigma$  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:*

$$\begin{aligned} g_h^{(m)} &= e^{\sigma\sqrt{dt/v}} - 1, 1 + g_l^{(m)} = \left(1 + g_h^{(m)}\right)^{-1}, \pi^{(m)} = \frac{1}{2} + \frac{1}{2} \frac{\mu}{\sigma} \sqrt{\frac{dt}{v}} \\ \rho^{(m)} &= (1 + r)^{\frac{dt}{v}} \left\{ (1 + g_l^{(m)})^{-\gamma} + \pi^{(m)} \left[ (1 + g_h^{(m)})^{-\gamma} - (1 + g_l^{(m)})^{-\gamma} \right] \right\} - 1 > 0 \\ \delta^{(m)} &= \frac{(1 + r) \left\{ (1 + g_l^{(m)})^{-\gamma} + \pi^{(m)} \left[ (1 + g_h^{(m)})^{-\gamma} - (1 + g_l^{(m)})^{-\gamma} \right] \right\}^{\frac{v}{dt}}}{\left\{ (1 + g_l)^{1-\gamma} + \pi^{(m)} \left[ (1 + g_h^{(m)})^{1-\gamma} - (1 + g_l^{(m)})^{1-\gamma} \right] \right\}^{\frac{v}{dt}}} - 1. \end{aligned}$$

<sup>13</sup>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, as conjectured.

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

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

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

**Proof.** See Appendix A. ■

As  $m \rightarrow \infty$ ,  $\frac{dt}{v}$  goes to zero, and the binomial lattice converges weakly to a geometric Brownian motion with parameters  $(\mu, \sigma)$ , the distributional assumption required by BS in continuous time.<sup>14</sup> This proposition shows that the results of Cox 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 also be 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,  $\pi$ , is unknown to investors who recursively estimate it through the simple maximum likelihood esti-

---

<sup>14</sup>This point can also be shown by noting that as  $m, v \rightarrow \infty$  the discrete state price density converges to a transformation of the lognormal:

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

mator:

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

where  $I_{\{m(t+k)+i=g_h\}}$  is an indicator function which is one when at the  $m(t+k)+i$ -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  $m(t+k)+j$  ( $j = 0, 1, \dots, m-1$ ) 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** *Suppose that the stock price is homogeneous of degree one in the level of real dividends, that a transversality condition holds, and that  $\rho > g_h^*$ . Then the stock price under Bayesian learning (BL) is*

$$\begin{aligned} S_t^{BL} &= D_t v \rightarrow \infty \lim \left\{ \sum_{i=1}^v \beta^i \sum_{j=0}^i (1+g_h^*)^j (1+g_l^*)^{i-j} \Pr_t^{BL} \left( D_{t+i}^j | n_t, N_t \right) \right\} \\ &\equiv D_t \Psi_t^{BL}(n_t, N_t) \end{aligned}$$

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

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

**Proof.** See Appendix A. ■

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  $\hat{\pi}_{t+k}$ . 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*.<sup>15</sup> 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.<sup>16</sup>

This path dependence means that no-arbitrage methods are more complicated in an equilibrium model with learning. Nevertheless, European call options on a stock index 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  $\tau$  and strike price  $K$  is*

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

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

<sup>15</sup>This can most easily be seen by comparing the value as of period  $t + 1$  of one dollar in period  $t + 3$  when both the high and low dividend growth rates occur. When high growth is followed by low growth we have

$$\begin{aligned} & \left[ 1 + r_{t+1}^{BL}(\hat{\pi}_{t+1}^u) \right]^{-1} \left[ 1 + r_{t+2}^{BL}(\hat{\pi}_{t+2}^{ud}) \right]^{-1} \\ &= \frac{1 + \rho}{(1 + g_l)^{-\gamma} + \frac{(n_t+1)}{(N_t+1)} [(1 + g_h)^{-\gamma} - (1 + g_l)^{-\gamma}]} \times \frac{1 + \rho}{(1 + g_l)^{-\gamma} + \frac{(n_t+1)}{(N_t+2)} [(1 + g_h)^{-\gamma} - (1 + g_l)^{-\gamma}]}, \end{aligned}$$

where  $\hat{\pi}_{t+2}^{ud} = \frac{n_t+1}{N_t+2} = \hat{\pi}_{t+2}^{du}$ .

When the sequence is reversed and low growth is followed by high growth, we get

$$\begin{aligned} & \left[ 1 + r_{t+1}^{BL}(\hat{\pi}_{t+1}^d) \right]^{-1} \left[ 1 + r_{t+2}^{BL}(\hat{\pi}_{t+2}^{du}) \right]^{-1} = \\ &= \frac{1 + \rho}{(1 + g_l)^{-\gamma} + \frac{n_t}{N_t+1} [(1 + g_h)^{-\gamma} - (1 + g_l)^{-\gamma}]} \times \frac{1 + \rho}{(1 + g_l)^{-\gamma} + \frac{n_t+1}{N_t+2} [(1 + g_h)^{-\gamma} - (1 + g_l)^{-\gamma}]}. \end{aligned}$$

Taking the ratio of these two expressions and letting  $G \equiv [(1 + g_h)^{-\gamma} - (1 + g_l)^{-\gamma}]$ , we have

$$\frac{(1 + g_l)^{-2\gamma} + \frac{n_t(n_t+1)}{(N_t+1)(N_t+2)} G^2 + \frac{n_t(N_t+2) + (n_t+1)(N_t+1)}{(N_t+1)(N_t+2)} (1 + g_l)^{-\gamma} G}{(1 + g_l)^{-2\gamma} + \frac{(n_t+1)^2}{(N_t+1)(N_t+2)} G^2 + \frac{(n_t+1)(2N_t+3)}{(N_t+1)(N_t+2)} (1 + g_l)^{-\gamma} G}$$

which in general differs from unity. For discounting purposes, the exact sequence of dividend realizations matters.

<sup>16</sup>This is another important difference between our model and that of David and Veronesi (2000) which assumes multiple states for dividends. In a Lucas-type general equilibrium 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 of a partial equilibrium nature.

$D_{t+v}^j = (1 + g_h)^j (1 + g_l)^{v-j} D_t$  ( $j = 0, 1, \dots, v$ ), and

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

**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

$$\begin{aligned} &\left[ \tilde{P}^{BL} \{X_{t+k+1} = n_t + j | X_{t+k} = n_t + i\} \right] = \tilde{M}_{t+k}(i + 1, j + 1) \quad (6) \\ =\beta &\begin{bmatrix} R_{t+k}^l (1 + g_l)^{-\gamma \frac{N_{t+k} - n_t}{N_{t+k}}} & R_{t+k}^h (1 + g_h)^{-\gamma \frac{n_t}{N_{t+k}}} & 0 & \dots & 0 \\ 0 & R_{t+k}^l (1 + g_l)^{-\gamma \frac{N_{t+k} - n_t - 1}{N_{t+k}}} & R_{t+k}^h (1 + g_h)^{-\gamma \frac{n_t + 1}{N_{t+k}}} & \dots & 0 \\ \dots & \dots & \dots & \dots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix} \end{aligned}$$

where  $R_{t+k}^l \equiv 1 + r_{t+k}^l$  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 ( $\hat{\pi}_t$ ). When risk aversion is not ‘too high’ ( $\gamma < 1$ ) the second effect dominates. A similar ranking across asset prices can be established by comparing option prices under learning to BS option prices:

**Proposition 5** *If the mean dividend growth rate is nonnegative and investors have optimistic beliefs ( $\hat{\pi}_t \geq \pi$ ) then*

$$C_t^{BL}(K) \geq C_t^{BS}(K) \quad \forall K.$$

**Proof.** See Appendix A. ■

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. Optimistic beliefs ( $\hat{\pi}_t > \pi$ ) are sufficient for the difference between option prices  $C_t^{BL}(K) - C_t^{BS}(K)$  to be monotonically increasing for low strike prices. For unrestricted beliefs  $\hat{\pi}_t$ , there will be intervals for the strike price over which  $C_t^{BL}(K) \geq C_t^{BS}(K)$  and others over which this inequality is reversed. For instance, pessimism makes it more likely that out of the money calls be underpriced under BL relative to BS; however, the mean preserving spread caused by the presence of learning effects keeps assigning positive valuation to deep out of the money calls that otherwise receive near-zero prices under BS. Furthermore, although optimistic beliefs are sufficient, they are not a necessary condition for learning to systematically affect option prices. Numerical results confirm that even when  $\hat{\pi}_t \leq \pi$ , provided that  $\pi - \hat{\pi}_t$  is not too large, it is possible to obtain  $C_t^{BL}(K) \geq C_t^{BS}(K)$  for many configurations of the remaining parameters.

Guidolin and Timmermann (2001) show that in this setup the price-dividend ratio is a monotonically increasing and convex function of beliefs,  $\hat{\pi}_t$ . The intuition is that the presence of learning effects shifts risk-neutral probability mass towards the tails relative to the lognormal benchmark underlying BS. In general, this could either depress or increase the equilibrium price of a call option. However, optimistic beliefs are sufficient to guarantee that the mass shifted towards the right tail dominates the mass shifted towards the left tail. Also, because of the self-enforcing effects of dividend changes, learning in a lattice framework widens (in both directions) the support of the perceived risk-neutral distribution of time  $t + \tau$  stock prices relative to the lognormal case. Convexity of the pricing kernel guarantees that this effect increases the equilibrium option price.

Effectively the cross-section of option prices allows us to infer the market's perception of fundamentals once the model is tested 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 set the parameters of the dividend process and the risk-free rate at plausible levels. 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 ( $\mu$ ) is set to 3%, to match the average dividend growth rate around our sample period (June 1992 - June 1995). Volatility ( $\sigma$ ) is set at 5%, slightly lower than the point estimate of the annual dividend growth rate whildD-0.9weillpe i( □

ally.<sup>17</sup> Finally the annualized risk-free rate ( $r$ ) is set to 4%, while the dividend yield is fixed at

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 sub-periods 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  $\hat{\pi}_t = 0.538$  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 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  $\pi$  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  $\hat{\pi}_t$  decreases towards  $\pi$  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 ( $\tau$ ) 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.<sup>20</sup> Figure 10 presents both dimensions of the implied volatility surface under BL. Skews dominate at all maturities, with exception of very short term options, for which an asymmetric smile obtains. While for OTM and ATM contracts the term structure is upward sloping, a richer variety of shapes obtains for ITM contracts.

#### 5.4. *Vanishing learning effects*

In our setup investors use a consistent estimator of  $\pi$ , and as  $N_t \rightarrow \infty$ ,  $\hat{\pi}_t \rightarrow \pi$  and learning effects vanish. 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 dividend realizations. Figure 11 shows the resulting option prices.

When learning effects are weak and investors have a more accurate estimate of  $\pi$ , BL option prices converge to BS prices. This is not surprising since learning is the only source of non-stationarity in our model. The first panel of Figure 11 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  $N_t \rightarrow \infty$  and  $\hat{\pi}_t \rightarrow \pi$  (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.<sup>21</sup>

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  $\pi$ . 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  $\pi$  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. In practice, the use of a rolling window of observations has two consequences. The first is obvious: agents now form an estimate of the unknown parameter

---

<sup>20</sup>See Campa and Chang (1995).

<sup>21</sup>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.

$\pi$  based only on the most recent  $R$  observations, where  $R$  is the length of the rolling window,  $\widehat{\pi}_t^R = \frac{\sum_{j=t-R+1}^t I_{\{g_{t+j}=g_h\}}}{R} = \frac{n_t^R}{R}$  (assuming  $R < t$ ). There is a second and subtler implication: When agents form perceptions of the probability distribution of future dividend levels on the binomial lattice, they have to integrate over all the possible future perceptions of  $\widehat{\pi}_t^R$ , accounting for the fact that moving into the future they will have to keep rolling the window of observations used in the estimation, replacing past observations with future realizations. The resulting formula to calculate the perceived probability distribution over future dividend levels is:

$$\begin{aligned} \Pr_t^{RW} \left\{ D_{t+i}^j | n_t^R, R \right\} &= \frac{1}{R^i} \sum_{l=1}^{\binom{i}{j}} \left\{ \left[ I_{\{g_{t+l}^l=g_h\}} n_t^R + \left( 1 - I_{\{g_{t+l}^l=g_h\}} \right) (R - n_t^R) \right] \times \right. \\ &\times \left. \prod_{k=2}^i I_{\{g_{t+k}^l=g_h\}} \left( n_t^R + \sum_{s=1}^{k-1} I_{\{g_{t+s}=g_h\}} - f_k \right) + \left( 1 - I_{\{g_{t+k}^l=g_h\}} \right) \left( R - n_t^R - \sum_{s=1}^{k-1} I_{\{g_{t+s}=g_h\}} + f_k \right) \right\} \end{aligned} \quad (7)$$

where  $f_k = \sum_{s=0}^{k-2} I_{\{g_{t-N+s}=g_h\}}$  is a ‘forget factor’ that counts the number of times in which dividends grew at a high rate between  $t - R$  and  $t - R + k - 1$ . Notice that the summation  $\sum_{l=1}^{\binom{i}{j}}$  iterates over all of the  $\binom{i}{j}$  independent paths leading to the same final dividend level  $D_{t+i}^j$ ,  $0 \leq j \leq i$ , with  $l$  indexing each of these paths. Indeed, the simple binomial coefficient  $\binom{i}{j}$  in Proposition 3 is now replaced by  $\sum_{l=1}^{\binom{i}{j}}$ . However, although its theoretical limiting properties are interestingly different, as long as the rolling window is not too small, the Bayesian model described in Section 4 will provide a good approximation to an exact solution which explicitly incorporates investors’ use of a rolling window.<sup>22</sup>

## 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 at a point in time. 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

---

<sup>22</sup>Simulation experiments (approximating rolling window option prices with Monte Carlo techniques) confirmed that for the values of  $N_t$  typically encountered in our sample, namely around 300 observations, the Bayesian price under an expanding and a rolling window were quite similar.

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<sup>23</sup>

$$\begin{aligned}
\min_{\{\pi_t\}_{t=1}^T, N, m} & \sum_{t=1}^T \sum_{\tau=\underline{\tau}_t}^{\bar{\tau}_t} \sum_{K_{\tau_t}=\underline{K}_{\tau_t}}^{\bar{K}_{\tau_t}} g(C^{BL}(f_t, \pi_t, N, m; \gamma, \beta), C(f_t)) \\
s.t. & \quad \frac{\pi_t N}{N+m} \leq \pi_{t+1} \leq \frac{\pi_t N + m}{N+m} \\
& \quad 0 \leq \pi_t \leq 1 \quad t = 1, \dots, T \\
& \quad N > 0, \quad m > 0
\end{aligned} \tag{8}$$

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  $K_{\tau_t}$  expiring in  $\tau_t$  days. The indexes appended to  $\tau_t$  and  $K_{\tau_t}$  reflect the fact that maturities and strike prices change over time, following the dynamics of the underlying price and the financial cycle. The first constraint arises from the fact that when dividends change  $m$  times in a unit interval,  $\pi_t$  can be updated to any value between  $\frac{\pi_t N}{N+m}$  (when dividends grow  $m$  times at the rate  $g_l$ ) and  $\frac{\pi_t N + m}{N+m}$  (when dividends grow  $m$  times at the rate  $g_h$ ). Finally,  $g(\cdot)$  is a function

---

<sup>23</sup>Our approach is similar to Bates (1991), who imposes CRRA preferences to estimate by NLS the parameters of an asymmetric jump diffusion model with systematic jump risk.

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,  $\sigma$ , is 5%. Given  $\sigma$ ,  $g_h$  and  $g_l$  can be determined from Proposition 3. We set  $\gamma = 0.9$  and  $\beta = \frac{1}{1.02} \simeq 0.98$  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 initial beliefs, an estimate  $\hat{m}$  of the frequency with which these beliefs are updated over the sample period,<sup>24</sup> and a  $T \times 1$  vector  $\hat{\pi}$  whose dynamics is constrained by Bayes rule.  $\hat{N}/\hat{m}$  can also be interpreted as an estimate of the length of the time window investors use to form their beliefs about fundamentals. For instance, if news on fundamentals arrive every three days ( $\hat{m} = \frac{1}{3}$ ) and  $\hat{N} = 300$ , this implies that agents are using a data window of 900 days.

### 6.1. Inferring learning from daily cross-sections

The objective of our first exercise is to infer from option prices the belief,  $\pi_t$ , its precision level,  $N_t$ , and the frequency with which beliefs are updated and new information arrives,  $m_t$ , for each week in the sample:

$$\min_{\pi_t, N_t, m_t} \sum_{\tau=\tau_t}^{\bar{\tau}_t} \sum_{K_{\tau_t}=\underline{K}_{\tau_t}}^{\bar{K}_{\tau_t}} [C^{BL}(f_t, \pi_t, m_t, N_t) - C(f_t)]^2 \quad (9)$$

for  $t = 1, 2, \dots, T$ . This amounts to minimizing the in-sample squared pricing errors produced by the model for each weekly cross-section of option prices.  $\hat{\pi}_t$ ,  $\hat{N}_t$ , and  $\hat{m}_t$  can be viewed as non-standard NLS estimates.<sup>25</sup> 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

<sup>24</sup>The perceived frequency of dividend news,  $m$ , is another structural parameter that agents are unlikely to know. In reality investors will probably use an estimate of its most likely value,  $\hat{m}$ . Also, we refer to  $m$  as a number of the form  $m = \frac{1}{d}$ ,  $d \in N$ , i.e. a rational number and not necessarily an integer.  $m < 1$  ( $d > 1$ ) would then imply that news are perceived to hit the market every  $d$  periods, on average. For instance  $\hat{m} = \frac{1}{2}$  implies that news on fundamentals are perceived to arrive every two days.

<sup>25</sup>The problem is nonstandard both because of the presence of constraints on the parameters and because  $N_t$  is a positive integer while  $m_t$  has structure  $\frac{1}{d}$ , where  $d \in N$ . As far as the metric  $g(\cdot)$  is concerned, we tried to estimate the six competing models listed below by minimizing instead the sum of squared implied volatility errors, i.e. deviation of theoretical implied volatilities from the empirical implied volatilities underlying observed prices. We obtain in-sample results roughly similar to the one presented below.

only three parameters on a data set with a cross-sectional size equal to the number of traded contracts on week  $t$ . Furthermore, this setup is fully consistent with the presence of infrequent and unpredictable structural breaks in the distribution of the fundamentals (cf. Section 5.4).  $\hat{N}_t$  then represents an estimate of the size of the window used by the market at time  $t$  to form its estimate of  $\hat{\pi}_t$ . When structural breaks are assessed to occur more frequently, the optimal reaction is to shorten the time window of observations used for estimation purposes, and vice versa. This is also our rationale for allowing sizeable ‘jumps’ in  $\hat{\pi}_t$ , since changing the size of the observation window can have quite a strong impact on the resulting estimate of  $\pi$ , especially when beliefs on the frequency of breaks are drastically revised.

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:

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

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

$$\sum_{\tau_t=\underline{\tau}_t}^{\bar{\tau}_t} \sum_{K_{\tau_t}=\underline{K}_{\tau_t}}^{\bar{K}_{\tau_t}} [C^{BS}(f_t; \hat{\sigma}_t) - C(f_t)]^2 \quad \text{s.t. } \hat{\sigma}_t > 0, \quad t = 1, \dots, T \quad (10)$$

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

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 accommodate the ‘smile’ bias along moneyness:

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

where  $\hat{\sigma}_{K_{\tau_t}} = \left\{ \hat{\sigma}_{K_{\tau_t}^{ITM}}, \hat{\sigma}_{K_{\tau_t}^{ATM}}, \hat{\sigma}_{K_{\tau_t}^{OTM}} \right\}$  is a volatility index function of moneyness.<sup>26</sup> 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 if 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

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

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

---

<sup>26</sup>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  $\tau$ .

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:

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

where  $\hat{\sigma}_{\tau_t^{Mat}} = \left\{ \hat{\sigma}_{\tau_t^{short}}, \hat{\sigma}_{\tau_t^{med}}, \hat{\sigma}_{\tau_t^{long}} \right\}$  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

$$\sum_{\tau_t = \underline{\tau}_t}^{\bar{\tau}_t} \sum_{K_{\tau_t} = \underline{K}_{\tau_t}}^{\bar{K}_{\tau_t}} \left[ C^{BS}(f_t; \hat{\sigma}_{\tau_t^{Mat}}) - C(f_t) \right]^2 \quad s.t. \quad \hat{\sigma}_{\tau_t^{Mat}} > 0 \quad (12)$$

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

The fourth model we consider is a deterministic volatility BS model modified to allow  $\sigma$  to be quadratic in strikes (to obtain smile shapes) and linear in time-to-maturity:<sup>27</sup>

$$(iv) \text{ BS-spline: } C^{spline}(f_t) = C^{BS}(f_t; \sigma^{spline}(\tau_t, K_{\tau_t}))$$

$$\sigma^{spline}(\tau_t, K_{\tau_t}) = \max \{ .01, \alpha_0 + \alpha_1 K_{\tau_t} + \alpha_2 K_{\tau_t}^2 + \alpha_3 \tau_t + \alpha_4 K_{\tau_t}^2 \tau_t \} \quad (13)$$

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

$$\min_{\alpha_t} SSR(\alpha_t) \equiv \sum_{\tau_t = \underline{\tau}_t}^{\bar{\tau}_t} \sum_{K_{\tau_t} = \underline{K}_{\tau_t}}^{\bar{K}_{\tau_t}} \left[ \sigma^{spline}(f_t; \alpha_t) - \sigma_I(f_t) \right]^2 \quad (14)$$

where  $\sigma_I$  denotes observed market implied volatilities (calculated by inverting the BS formula). We then plug the estimated  $\hat{\alpha}_t$  into BS to obtain option prices. Notice that although similar in spirit, (13) differs from models two and three because it is fitted to the implied volatility surface and not to observed option prices.

Finally, we estimate Heston and Nandi's (2000) NGARCH(1,1) model on the sequence of weekly cross sections of option prices. Assume that the continuously compounded returns on the underlying asset follow a nonlinear asymmetric GARCH (1,1)-in-mean process over time steps of fixed length equal to one day:

$$r(t) = r^f + \lambda h(t) + \sqrt{h(t)} z(t) \quad (15)$$

$$h(t) = \omega + \alpha \left[ z(t-1) - \xi \sqrt{h(t-1)} \right]^2 + \phi h(t-1), \quad (16)$$

---

<sup>27</sup>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  $\sigma$  is a function of the strike and maturity.

where  $r(t) \equiv \ln S(t) - \ln S(t - \Delta)$ ,  $r^f$  is the continuously compounded and constant risk-free interest rate,  $z(t) \sim N(0, 1)$ , and  $h(t)$  is the volatility of the underlying daily returns for the time interval  $[t - 1, t)$  conditional on the information available at time  $t - 1$ . Let

$$z^*(t) = z(t) + \left( \lambda + \frac{1}{2} \right) \sqrt{h(t)} \quad \text{and} \quad \xi^* = \xi + \lambda + \frac{1}{2}. \quad (17)$$

Substituting these definitions into (15) - (16) we obtain:

$$r(t) = r^f - \frac{1}{2} \sqrt{h(t)} + \sqrt{h(t)} z^*(t) \quad (18)$$

$$h(t) = \omega + \alpha \left[ z^*(t-1) - \xi^* \sqrt{h(t-1)} \right]^2 + \phi h(t-1). \quad (19)$$

Under the transformed measure  $z^*(t)$ ,  $r_t$  has conditional mean  $r^f - \frac{1}{2} \sqrt{h(t)}$  and conditional variance  $h_t$ . Under assumptions ensuring that a call option with one period to expiration satisfies the Black-Scholes formula (see, e.g., Duan (1995)), Heston and Nandi (2000) show that a local risk neutral probability measure  $Q$  exists, that it is unique and that it is characterized by another NGARCH (1,1) process with  $E^Q [r_t | \Phi_{t-1}] = r^f - \frac{1}{2} h_t$ . They also show how to solve for the conditional generating function  $f_t(\phi)$  of the final spot price  $S_T$  under the process in (15) - (16) using the method of undetermined coefficients.<sup>28</sup> The current equilibrium price of a call option with strike  $K$  and  $\tau$  periods to maturity can be calculated by inverting the risk neutral conditional characteristic function  $f_t^*(i\phi)$  according to the formula:

$$P_K^* = F^*(\ln K) = \frac{1}{2} + \frac{1}{\pi} \int_0^\infty \text{Re} \left[ \frac{e^{-i\phi \ln K} f_t^*(i\phi)}{i\phi} \right] d\phi, \quad (20)$$

We can therefore write our final benchmark model as follows:

$$(v) \text{ NGARCH}(1, 1) : C(S_t, \tau, K; \omega, \alpha, \phi, \xi, \theta) = S_t P_S^* - e^{-r^f \tau} K P_K^*$$

where  $P_S^* = \frac{1}{2} + \frac{1}{\pi} \int_0^\infty \text{Re} \left[ \frac{e^{-i\phi \ln K} f_t^*(i\phi+1)}{i\phi f_t^*(1)} \right] d\phi$ .<sup>29</sup>

<sup>28</sup> $f_t(\phi)$  is log-linear with coefficients that depend on the parameters  $\theta$ ,  $\omega$ ,  $\alpha$ ,  $\phi$ , and  $\xi$ . These coefficients can be calculated in a recursive fashion starting from the terminal conditions where all of the coefficients must equal zero so that  $f_T(\phi) = S_T^\phi$ .

<sup>29</sup>Our implementation follows the same procedures detailed in the paper by Heston and Nandi. We (i) use daily S&P 500 index returns to model the evolution of volatility; (ii) update the variance  $h(t+1)$  each week by using the time series of index returns over the previous 252 days.

## 6.2. Goodness-of-fit Measures

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

$$\begin{aligned} RMSVE &= T^{-1} \sum_{t=1}^T \left\{ W_t^{-1} \sum_{\tau_t=\mathcal{I}_t}^{\bar{\tau}_t} \sum_{K_{\tau_t}=\underline{K}_{\tau_t}}^{\bar{K}_{\tau_t}} \left[ \widehat{C}(f_t) - C(f_t) \right]^2 \right\}^{1/2} \\ &= T^{-1} \sum_{t=1}^T \left\{ W_t^{-1} \widehat{SSR}_t^M \right\}^{1/2} \end{aligned} \quad (21)$$

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

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

$$MAVE = T^{-1} \sum_{t=1}^T W_t^{-1} \sum_{\tau_t=\mathcal{I}_t}^{\bar{\tau}_t} \sum_{K_{\tau_t}=\underline{K}_{\tau_t}}^{\bar{K}_{\tau_t}} \left| \widehat{C}(f_t) - C(f_t) \right|. \quad (22)$$

To assess the potential for overfitting, we report Akaike's information criterion (AIC),<sup>30</sup>

$$AIC = T^{-1} \sum_{t=1}^T \left[ -2 \ln L(\hat{\theta}_t) + 2p \right], \quad (23)$$

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.

Finally, we follow Dumas et al. (1998) and calculate the mean absolute error outside the bid-ask spread (MOE):<sup>31</sup>

$$MOE = \max \left\{ \max[\widehat{C}(f_t) - C^a(f_t), 0], \max[C^b(f_t) - \widehat{C}(f_t), 0], 0 \right\}. \quad (24)$$

According to this definition,  $MOE > 0$  if and only if either the theoretical option price exceeds the ask price ( $C^a(f_t)$ ) or the theoretical option price is below the bid quote ( $C^b(f_t)$ ).

To compare the results across models, we also report the proportion of weeks 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.

<sup>30</sup>For simplicity, we call  $\theta$  the vector of parameters to be estimated with reference to any possible pricing model  $M$ . For instance, in the case of model 4,  $\theta = \alpha$ , for the BL model  $\theta = [\pi_t \ N_t \ m_t]'$ , and for the variants on the BS model  $\theta$  collects the contract-specific volatility parameters.

<sup>31</sup>Our definition differs from that in Dumas et al. (1998, p. 2072) since our  $MOE \geq 0$ , while their measure can be either positive or negative. We want MOE to be able to detect all situations where the theoretical option price falls outside the bid-ask spread and thus do not distinguish between cases where the price is too low and cases where it is too high.



Table 2 contains the in-sample averages of RMSVE, AIC, and MAVE across the 292 weeks during the period June 1988 - December 1993. On average, and across the different measures of goodness-of-fit, the *BS – spline* model seems to outperform all alternatives, followed by the NGARCH model and the BL model.<sup>32</sup> Although the average distance between the BL model and BS-spline is not excessive (less than 10 cents when measured by the RMSVE), the AIC estimates suggest that the larger number of parameters of the BS-spline model does not completely explain its superior performance. At the same time, our weekly implementation of the BL model leads to an improvement of 78 cents over Black-Scholes. Though far from being the best fitting model in-sample, the BL approach provides quite precise estimates of option prices, often inside the bid-ask spread. Indeed the MOE of the BL model (35 cents on average) is quite close to the BS-spline minimum of 30 cents and better than any of the other competing models.

Following Dumas et al. (1998), these performance indicators are broken down according to both moneyness and the time-to-expiration of each traded contract. Interesting information can be extracted when the goodness-of-fit indicators are decomposed according to either moneyness or time-to-maturity (Panels B and C). The BL model fits as well as the BS-spline for ATM contracts. For these contracts, BL has a much lower MOE than BS-spline, signalling that the latter — although surely superior on average — can also lead to gross mispricings outside the bid-ask spread for these options. While the NGARCH model is particularly effective at pricing ITM contracts, BS and its variants  $BS(\tau)$  and  $BS(m)$  only provide good in-sample fits for short-term options.

Table 3 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 the estimates under  $BS(\tau)$  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, while the  $\pi_t$  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 12 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 week to the next. The case of highest instability concerns the NGARCH model, as already reported by Heston and Nandi (2000, p. 605). In particular, the estimates of  $\xi$  and  $\phi$  oscillate

---

<sup>32</sup>This finding differs slightly from Heston and Nandi’s (2000) result that the NGARCH(1,1) model outperforms the BS-spline by a few cents in-sample. This difference is likely to reflect the different sample periods and data selection procedures.

dramatically: while  $\hat{\phi}$  is frequently observed to approach extreme values such as 0 and 1,  $\hat{\xi}$  even switches sign, thus implying the occasional presence of a negative leverage effect that could be associated with atypical shapes of the implied volatility surface. This instability must be carefully considered when evaluating the out-of-sample performance of the NGARCH model. In all cases, such sharp revisions in the parameter estimates suggest significant instability in the implied volatility surface. The BL model generates relatively stable parameter estimates, with  $\hat{\pi}_t$  steadily in the range of [.52, .65],  $\hat{N}_t$  normally around 300 observations, and  $\hat{m}_t$  most of the time in the narrow interval [.3, .5]. The speed of flow of information seems to peak in the first months of 1991 in correspondence with the recovery of the US economy from the recession of 1990.<sup>33</sup> In general option markets display weak optimism ( $\hat{\pi}_t \geq \frac{1}{2}$ ). The stability of the parameter estimates suggests a 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 uniformly improve as more parameters are introduced, but on the precision with which the model predicts future option prices out-of-sample. Moreover, despite the superior in-sample fit of the BS-spline and NGARCH models uncovered in the previous subsection, we also found some results that are potentially problematic for their empirical performance, such as the instability of the parameter estimates and the large MOEs. 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 week  $t$  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 week.<sup>34</sup>

Table 4 provides summary statistics for the one-step-ahead prediction errors. Unsurprisingly, these are somewhat larger than the errors in Table 2. Out-of-sample, the BL model outperforms the proposed alternatives in the aggregate (Panel A), albeit only marginally in the case of the BS-spline model. The root mean squared prediction error of the BL model is 1 dollar against 1.68 dollar for Black-Scholes and 1.06 dollars or higher for the empirical modifications to BS. The NGARCH does relatively poorly, outperforming Black-Scholes (by 42 cents) but also predicting with average errors 16 cents larger than BL. In relative RMSPE terms, BL is the

---

<sup>33</sup>Since  $\hat{m}$  is on average around 0.3,  $\hat{N} = 300$  corresponds to about 1000 trading days, i.e. 4 years.  $\hat{N}_t$  and  $\hat{m}_t$  also have a high positive correlation (0.55) so the actual size of the data window implied by the estimates displays low volatility and for most of the sample period lies between 800 and 1,100 observations.

<sup>34</sup>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 week. This way we test the predictive properties of the option pricing models independently of the ability to predict the future stock price.

best model out-of-sample in 35% of the sample weeks. The superior forecasting performance of the BL model suggests that this model does not provide an accurate fit purely as a result of weekly variation in the learning parameters. On the other hand, it is natural to attribute the relatively weaker predictive performance of the BS-spline and NGARCH models to their parameter instability. When the forecast indicators are decomposed according to either moneyness or time-to-maturity (Panels B and C), we get a better picture of the relative strengths of the BL model: It outperforms the other benchmarks for OTM contracts and always compares closely with the best performing alternative 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  $\Delta$  measures the sensitivity of the option price to the value of the underlying asset.  $\widehat{\Delta}(f_t; \hat{\theta}_t)$  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:

$$\eta_{t+1}(\hat{\theta}_t) = \Delta C_{t+1} - \widehat{\Delta} C_{t+1}(\hat{\theta}_t) \quad (25)$$

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  $\eta_t(\hat{\theta}_t) \neq 0$  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 5 reports statistics on hedging errors. Results are quite similar to the out-of-sample prediction experiments, in the sense that the BL model still outperforms all the proposed

and (ii) the risk-free interest rate (from proposition 3). Figure 13 reports the two time series. Overall, the implied path for these variables is quite plausible and consistent with an economy moving through a slow learning path.

The expected rate of growth of fundamentals is 5% per annum on average. This is higher than the average rate of growth of real dividends (2.2%) reported by Shiller (2000) for the period 1988-1993. However the dynamics over time of the implied expected fundamental growth fits the NBER business cycles dates quite well. For instance, the average expected growth rate is 7.4% for the period June 1988 - July 1990 and then declines to a modest 3.5% for the period August 1990 - December 1993, which contains the last recorded recession of 1990-1991.<sup>35</sup> Moreover, the high average might be related to the high growth rates observed by agents on a learning path over the expansion period 1983-1989 (3.9%), in the sense that agents' beliefs might have adjusted slowly to the incipient recession and the exogenous uncertainty related to the Gulf War. Although the expected growth rate implied by the BL model is never negative, consistent with beliefs that the US economy is growing in real terms, annualized growth rates of real dividends as high as 9% appear in Shiller's data.

Another implied series that gives positive indications on the robustness of the BL model is the riskfree interest rate. Its average is 6.4% vs. a sample value of 5.6% for the 1988-1993 period. The implied interest rate appears to decline over time, consistent with the evidence and with the evolution in the US business cycle during our sample. The model also produces a time-series standard deviation of the risk-free rate (2.3%) that fits well with the sample value of 2.1%.

### 6.6. *Learning dynamics with an expanding window*

The empirical exercise of the previous subsections ignores the intertemporal restrictions imposed by our model on the updating of investors' beliefs and therefore does not provide the strongest possible test. Therefore we try next to infer the dynamics of learning from observed option prices by imposing these restrictions on  $\hat{\pi}_t$ . The model in section 4 assumes that beliefs evolve as follows:

$$\hat{\pi}_{m(t+k)+j} = \hat{\pi}_{m(t+k)+j-1} + \frac{I_{\{m(t+k)+j=g_h\}} - \hat{\pi}_{m(t+k)+j-1}}{N_{m(t+k)+j}}. \quad (26)$$

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 blocks of time each of which lasts for two months, and thus consider a panel of option prices comprising July and August 1988, followed by a panel of the September and October 1988 prices, and so on, up to December 1993. For each block we solve

---

<sup>35</sup>Of course, the identification of fundamentals with real dividends is problematic at best.

the program:

$$\begin{aligned}
& \min_{\{\pi_t\}_{t=1}^T, N} \sum_{t=1}^T \sum_{\tau=\underline{\tau}_t}^{\bar{\tau}_t} \sum_{K_{\tau_t}=\underline{K}_{\tau_t}}^{\bar{K}_{\tau_t}} [C^{BL}(f_t, \pi_t, N; \gamma, \beta) - C(f_t)]^2 \\
& \quad s.t. \quad \frac{\pi_t N}{N+1} \leq \pi_{t+1} \leq \frac{\pi_t N + 1}{N+1} \\
& \quad 0 \leq \pi_t \leq 1 \quad t = 1, \dots, T-1 \quad N > 0
\end{aligned} \tag{27}$$

The program is limited to blocks of 8-9 weeks to avoid the curse of dimensionality implicit in this NLS estimation program: as  $T \rightarrow \infty$ , 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-monthly period means estimating a vector with 9-10 parameters to fit a sample of well over 300 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. We make the same assumptions on preferences as in Section 6.3 and to simplify the estimation task also impose  $m = 1/7$ , i.e. fundamentals change at weekly frequency.

Figure 14 reports results over the period July 1988 - June 1991. We limit ourselves to this interval of time for computational reasons. The figure suggests that the estimated sequence of beliefs  $\{\hat{\pi}_t\}_{t=1}^T$  is roughly consistent with our results above, as  $\hat{\pi}_t$  varies in the interval  $[0.3, 0.7]$ . We notice the same phenomena that were identified above: During 1990 agents drastically revised downwards their average beliefs on the likelihood of a high growth rate. This is consistent with a rational reaction to an incipient recession as well as to the additional uncertainty created by the Gulf War. While other even more restrictive tests can be designed, we consider these findings *prima facie* evidence that the BL model is indeed capable of extracting measures of agents' beliefs that are economically meaningful.

## 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 out-of-sample prediction and 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 well 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] Abel, A., 1999. Risk premia and term premia in general equilibrium. *Journal of Monetary Economics* 43, 3-33.
- [3] Aït-Sahalia, Y., Lo, A., 1998. Nonparametric estimation of state-price densities implicit in financial asset prices. *Journal of Finance* 53, 499-547.
- [4] Amin, K., 1993. Jump diffusion option valuation in discrete time. *Journal of Finance* 48, 1833-1863.
- [5] Ball, C., Torous, W., 1985. On jumps in common stock prices and their impact on call option pricing. *Journal of Finance* 40, 155-173.
- [6] Bakshi G., Cao, C., Chen, Z., 1997. Empirical performance of alternative option pricing models. *Journal of Finance* 52, 2003-2049.
- [7] Bates, D., 1991. The crash of '87: was it expected? The evidence from options markets. *Journal of Finance* 46, 1009-1044.
- [8] Bates, D., 1996. Testing option pricing models, in: Maddala G. S., Rao, C. R. (Eds.), *Handbook of Statistics, Vol. 15: Statistical methods in Finance*. North Holland, Amsterdam, pp. 567-611.

- [9] Bick, A., 1990. On viable diffusion price processes of the market portfolio. *Journal of Finance* 45, 673-689.
- [10] Black, F., Scholes, M., 1973. The pricing of options and corporate liabilities. *Journal of Political Economy* 81, 637-659.
- [11] Bossaerts, P., 1995. The econometrics of learning in financial markets. *Econometric Theory* 11, 151-189.
- [12] Bossaerts, P., 1999. Learning-induced price volatility. Mimeo, Caltech.
- [13] Campa J., Chang, K., 1995. Testing the expectations hypothesis on the term structure of volatilities. *Journal of Finance* 50, 529-547.
- [14] Campbell, J., 1996. Consumption and the stock market: interpreting the international experience. National Bureau of Economic Research, working paper *5610*.
- [15] Chriss N., 1997. *Black-Scholes and Beyond. Option Pricing Models*, Irwin.
- [16] Cox, J., Ross, S., Rubinstein, M., 1979. Option pricing : a simplified approach. *Journal of Financial Economics* 7, 229-263.
- [17] Das, S., Sundaram, R., 1999. Of smiles and smirks: a term structure perspective. *Journal of Financial and Quantitative Analysis* 34, 211-239.
- [18] David, A., Veronesi, P., 1999. Option prices with uncertain fundamentals: theory and evidence on the dynamics of implied volatilities and over-/underreaction in the options market. Mimeo, Federal Reserve Board of Governors and University of Chicago.
- [19] Duan, J.-C., 1995. The GARCH option pricing model. *Mathematical Finance* 5, 13-32.
- [20] Dumas, B., Fleming, J., Whaley, R., 1998. Implied volatility functions: empirical tests. *Journal of Finance* 53, 2059-2106.
- [21] Guidolin M., Timmermann, A., 2001. Asset prices in a binomial lattice equilibrium model with rational learning. Mimeo, University of California San Diego and University of Virginia.
- [22] Harrison, J., Kreps, D., 1979. Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory* 20, 381-408.
- [23] He H., Leland, H., 1993. On equilibrium asset price processes. *Review of Financial Studies* 6, 593-617.

- [24] 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.
- [25] Heston S., Nandi, S., 2000. A closed-form GARCH option valuation model. *Review of Financial Studies* 13, 585-625.
- [26] Hull, J., White, A., 1987. The pricing of options on assets with stochastic volatilities. *Journal of Finance* 42, 281-300.
- [27] Jackwerth, J. Rubinstein, M., 1996. Recovering probability distributions from option prices. *Journal of Finance* 51, 1611-1631.
- [28] Judd, K., 1998. *Numerical Methods in Economics*. MIT Press, Cambridge, Mass.
- [29] Leland, H., 1985. Option pricing and replication with transaction costs. *Journal of Finance* 40, 1283-1302.
- [30] Lucas, R., 1978. Asset prices in an exchange economy. *Econometrica* 46, 1429-1445.
- [31] Melino A., Turnbull, S., 1990. Pricing foreign currency options with stochastic volatility. *Journal of Econometrics* 45, 239-265.
- [32] Merton, R., 1976. Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics* 3, 125-144.
- [33] Pliska S., 1997. *Introduction to Mathematical Finance*. Blackwell Publishers.
- [34] Rosenberg, J., Engle, R., 1997. Option hedging using empirical pricing kernels. University of California San Diego, discussion paper 97-20.
- [35] 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.
- [36] Rubinstein, M., 1994. Implied binomial trees. *Journal of Finance* 49, 771-818.
- [37] Schachermayer, W., 1994. Martingale measures for discrete-time processes with infinite horizon. *Mathematical Finance* 4, 25-56.
- [38] Shiller R., 2000. *Irrational Exuberance*. Princeton University Press, Princeton, NJ.
- [39] Stapleton, R., Subrahmanyam, M., 1984. The valuation of options when asset returns are generated by a binomial process. *Journal of Finance* 39, 1525-1539.



- [40] Timmermann, A., 1993. How learning in financial markets generates volatility and predictability in stock prices. *Quarterly Journal of Economics* 108, 1135-1145.
- [41] Walters F., Parker, L., Morgan, S., Deming, S., 1991. *Sequential Simplex Optimization*. CRC Press.
- [42] West, K., 1996. Asymptotic inference about predictive ability. *Econometrica* 64, 1067-1084.
- [43] Wiggins, J., 1987. Option values under stochastic volatility: theory and empirical estimates. *Journal of Financial Economics* 19, 351-372.
- [44] Yan, H., 2000. Uncertain growth prospects, estimation risk, and asset prices. Mimeo, University of Texas at Austin.

## Appendix A

**Proof.** [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  $\gamma$  gives the assumed annual risk-free rate  $r$  and dividend yield  $\delta$ , 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  $m, \mu, \sigma$ , the risk-free rate  $r$  and the dividend yield  $\delta$ . 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  $\gamma$  jointly depend on  $r$  and  $\delta$ . 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  $m \rightarrow \infty$ ,  $\rho^{(m)}$  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  $\gamma$  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.** [Proposition 3.] Start from the first of the two Euler equations in (4). Taking the limit as  $T \rightarrow \infty$ , assuming that the sum converges, using the transversality condition, and dividing and multiplying all the terms in the summation by the current dividend level, we obtain:

$$\begin{aligned} S_t^{BL} &= D_t \left\{ \widehat{E}_t [Q_{t+1}(1 + g_{t+1})] + \sum_{s=2}^{\infty} \widehat{E}_t \left[ Q_{t+1}(\dots \widehat{E}_{t+s-1}(Q_{t+s}D_{t+s})\dots) \right] \right\} \\ &= D_t \left\{ \widehat{E}_t [\beta(1 + g_{t+1})^{1-\gamma}] + \widehat{E}_t \left[ \beta(1 + g_{t+1})^{1-\gamma} \widehat{E}_{t+1} \left( \beta \left( \frac{D_{t+2}}{D_{t+1}} \right)^{-\gamma} \times \right. \right. \right. \\ &\quad \left. \left. \left. + \sum_{s=3}^{\infty} \widehat{E}_t \left[ Q_{t+1}(\dots \widehat{E}_{t+s-1}(Q_{t+s}D_{t+s})\dots) \right] \right) \right] \right\} \\ &= \dots = D_t \left\{ \beta \sum_{j=0}^1 (1 + g_h^*)^j (1 + g_l^*)^{1-j} P_t^{BL} \left( D_{t+1}^j | n_t, N_t \right) + \square \right. \end{aligned}$$

+

$$\begin{aligned}
& \times (1 + g_l^*)^{2-j} P_t^{BL} \left( D_{t+1}^j | n_t, N_t \right) + \dots + \sum_{s=3}^{\infty} \widehat{E}_t \left[ Q_{t+1} (\widehat{E}_{t+s-1} (Q_{t+s} D_{t+s}) \dots) \right] \Big\} \\
& = D_t \lim_{v \rightarrow \infty} \left\{ \sum_{i=1}^v \beta^j \sum_{j=0}^i (1 + g_h^*)^j (1 + g_l^*)^{i-j} P_t^{BL} \left( D_{t+i}^j | n_t, N_t \right) \right\} \\
& = D_t \Psi_t^{BL} (g_h, g_l, \gamma, \rho, n_t, N_t)
\end{aligned}$$

where we have used the fact that in equilibrium  $Q_{t+s} = \beta \left( \frac{D_{t+s}}{D_{t+s-1}} \right)^{-\gamma}$ . Notice that  $\Psi_t^{BL} > 0$  ensures positive stock prices if dividends are always positive. For this to happen it suffices that  $(1 + g_h^*)^j (1 + g_l^*)^{i-j} > 0 \forall i, j$ . These conditions follow on their turn from the fact that  $g_h > g_l > -1$  and  $\rho > -1$ .

As for the bond price, the second Euler equation in (4) directly gives the result by evaluating the conditional expectation taking  $\Pr_t^{BL} \left( D_{t+i}^j | n_t, N_t \right)$  as the relevant probability measure. Since

$$r_t^{BL}(\widehat{\pi}_t) = \frac{1}{B_t^{BL}(\widehat{\pi}_t)} - 1 = \frac{1 + \rho}{(1 + g_l)^{-\gamma} + \widehat{\pi}_t [(1 + g_h)^{-\gamma} - (1 + g_l)^{-\gamma}]} - 1$$

it is immediate to verify that  $\rho > -1$  is sufficient for  $r_t^{BL} > -1 \forall t$  as the denominator is always positive from  $g_h > g_l > -1$ .

A last aspect of the result is the convergence of the infinite sum in the equilibrium stock price, or equivalently the existence of the BL equilibrium. Since  $\gamma \leq 1$ ,  $\rho > g_h^*$  is necessary and sufficient for the equilibrium to exist. In this case  $g_h^* > g_l^*$  so that the worst possible case is  $\Pr_t^{BL} \left( D_{t+i}^i | n_t, N_t \right) = 1$ . Then:

$$\Psi_t^{BL} = \left\{ \sum_{i=1}^{\infty} \left( \frac{1 + g_h^*}{1 + \rho} \right)^i \right\}$$

converges to  $\frac{1 + g_h^*}{\rho - g_h^*}$  if and only if  $\rho > g_h^*$ . By a backward induction reasoning similar to the one made above  $\rho > g_h^*$  is necessary and sufficient for existence of the equilibrium when  $\gamma \leq 1$ . Given  $\rho$  this imposes a lower bound on  $\gamma$  or, conversely, given  $\gamma$  this implies an upper bound on  $\rho$ .<sup>36</sup> ■

**Proof.** [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

---

<sup>36</sup>Since  $\gamma = 1$  implies  $g_h^* = g_l^* = 0$ , in this case the BL pricing kernel simplifies to:

$$\begin{aligned}
\Psi^{BL} &= \sum_{i=1}^{\infty} \beta^j \sum_{j=0}^i \Pr_t^{BL} \left( D_{t+i}^j | n_t, N_t \right) \\
&= \sum_{i=1}^{\infty} \beta^j = \frac{1}{\rho} = \Psi^{FIRE},
\end{aligned}$$

i.e. the FI kernel. Then a  $\rho > 0$  is sufficient for the equilibrium to exist.

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:

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

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

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

Adding  $D_{t+k}^*$  to both sides and using that

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

we obtain:

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

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

$$\begin{aligned} \hat{P} \left\{ S_{t+k+1}^j \Big| \hat{\pi}_{t+k} \right\} &= Q_{t+k+1} \frac{(1+\rho)}{(1+g_l)^{-\gamma} + \hat{\pi}_{t+k} [(1+g_h)^{-\gamma} - (1+g_l)^{-\gamma}]} \times P \left\{ D_{t+k+1}^j \Big| \hat{\pi}_{t+k} \right\} \\ &= \beta \left( \frac{D_{t+k+1}^j}{D_{t+k}} \right)^{-\gamma} [1 + r_{t+k}^{BL}(\hat{\pi}_{t+k})] P \left\{ D_{t+k+1}^j \Big| \hat{\pi}_{t+k} \right\}. \end{aligned}$$

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

$$\begin{aligned} \tilde{P} \left\{ S_{t+k+1}^j \Big| \hat{\pi}_{t+k} \right\} &= \frac{1}{1 + r_{t+k}^{BL}(\hat{\pi}_{t+k})} \hat{P} \left\{ S_{t+k+1}^j \Big| \hat{\pi}_{t+k} \right\} \\ &= \beta \left( \frac{D_{t+k+1}^j}{D_{t+k}} \right)^{-\gamma} [I_{\{j=1\}} \hat{\pi}_{t+k} + (1 - I_{\{j=1\}})(1 - \hat{\pi}_{t+k})]. \end{aligned}$$

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,  $\binom{v}{j}$  :

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

It is easily shown that

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

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. ■

**Proof.** [Proposition 5] Let  $\nabla_{BS}^{BL}(K) \equiv C_{t+k}^{BL}(K) - C_{t+k}^{BS}(K)$ ,  $\Psi_{t+k+v}^{BL}(s) \equiv \Psi_{t+k+v}^{BL}(n_{t+k} + s, N_{t+k} + v)$ , and  $P_{t+k}^{BL}(s) \equiv P_{t+k}^{BL} \{ D_{t+k+v}^s | n_{t+k}, N_{t+k} \}$ . We start by observing that conditioning on a common, current stock price  $S_{t+k}$  has implications for the array of possible future stock prices in  $t + k + v$ . In fact  $S_{t+k+v}^{BL}(0) = \Psi_{t+k+v}^{BL}(0)(1 + g_l)^v D_t = \frac{\Psi_{t+k+v}^{BL}}{\Psi_{t+k}^{BL}}(1 + g_l)^v S_{t+k} < S_{t+k+v}^{BS}(0) = \Psi^{BS}(1 + g_l)^v D_t = (1 + g_l)^v S_{t+k}$  holds if and only if  $\frac{\Psi_{t+k+v}^{BL}}{\Psi_{t+k}^{BL}} < 1$  as  $S_{t+k}^{BS} = S_{t+k}^{BL}$  by assumption. It is easy to show that this is guaranteed by the fact that  $\widehat{\pi}_{t+k+v}(0) = \frac{n_{t+k}}{N_{t+k} + v} < \widehat{\pi}_{t+k} = \frac{n_{t+k}}{N_{t+k}}$  so that  $\Psi_{t+k+v}^{BL}(0) < \Psi_{t+k}^{BL}(n_{t+k}, N_{t+k})$ . Similarly,  $S_{t+k+v}^{BL}(v) = \frac{\Psi_{t+k+v}^{BL}}{\Psi_{t+k}^{BL}}(1 + g_h)^v S_{t+k} > S_{t+k+v}^{BS}(v) = (1 + g_h)^v S_{t+k}$  holds if and only if  $\frac{\Psi_{t+k+v}^{BL}}{\Psi_{t+k}^{BL}} > 1$ .

We study  $\nabla_{BS}^{BL}(K)$  in five disjoint intervals for the strike price:  $[0, S_{t+k+v}^{BL}(0))$ ,  $[S_{t+k+v}^{BL}(0), S_{t+k+v}^{BS}(0))$ ,  $[S_{t+k+v}^{BS}(0), S_{t+k+v}^{BS}(v))$ ,  $[S_{t+k+v}^{BS}(v), S_{t+k+v}^{BL}(v))$ , and  $[S_{t+k+v}^{BL}(v), +\infty)$ . As for the

first interval, observe that when  $K = 0$

$$\begin{aligned}\nabla_{BS}^{BL}(0) &= \sum_{s=0}^v \beta^v \left( \frac{D_{t+k+v}^s}{D_{t+k}} \right)^{-\gamma} S_{t+k+v}^{BL}(s) P_{t+k}^{BL}(s) - \sum_{s=0}^v \beta^v \left( \frac{D_{t+k+v}^s}{D_{t+k}} \right)^{-\gamma} S_{t+k+v}^{BS} \binom{v}{s} \pi^s (1-\pi)^{v-s} \\ &= S_{t+k} \beta^v \widehat{E}_{t+k} \left\{ \widehat{E}_{t+k+1} \left[ \dots \left( \Psi_{t+k+v}^{BL} \left( \frac{D_{t+k+v}}{D_{t+k}} \right)^{1-\gamma} \right) \right] \right\} - S_{t+k} \beta^v E_{t+k} \left[ \left( \frac{D_{t+k+v}}{D_{t+k}} \right)^{1-\gamma} \right]\end{aligned}$$

is positive as  $\Psi_{t+k+v}^{BL} \left( \frac{D_{t+k+v}^s}{D_{t+k}} \right)^{1-\gamma}$  is a non-decreasing, convex function of  $\widehat{\pi}_{t+k+v}$ . Indeed both  $\Psi_{t+k+v}^{BL}$  and  $\left( \frac{D_{t+k+v}^s}{D_{t+k}} \right)^{1-\gamma}$  are positive, non-decreasing and convex functions of  $\widehat{\pi}_{t+k+v}$ , and it is straightforward to check that their product possesses the same properties. Then, for  $\widehat{\pi}_{t+k} = \pi$ , Jensen's inequality implies that

$$\begin{aligned}&\Psi_{t+k}^{BL} \widehat{E}_{t+k} \left\{ \frac{\Psi_{t+k+1}^{BL}}{\Psi_{t+k}^{BL}} \left( \frac{D_{t+k+1}}{D_t} \right)^{1-\gamma} \widehat{E}_{t+k+1} \left[ \frac{\Psi_{t+k+2}^{BL}}{\Psi_{t+k+1}^{BL}} \left( \frac{D_{t+k+2}}{D_{t+k+1}} \right)^{1-\gamma} \dots \right. \right. \\ &\quad \left. \left. \dots \widehat{E}_{t+k+v-1} \left( \frac{\Psi_{t+k+v}^{BL}}{\Psi_{t+k+v-1}^{BL}} \left( \frac{D_{t+k+v}}{D_{t+k+v-1}} \right)^{1-\gamma} \mid \pi \right) \dots \mid \pi \right] \mid \pi \right\} \\ &\geq \Psi_{t+k}^{BL} E_{t+k} \left[ \frac{\Psi_{t+k+1}^{BL}}{\Psi_{t+k}^{BL}} \left( \frac{D_{t+k+1}}{D_{t+k}} \right)^{1-\gamma} \frac{\Psi_{t+k+2}^{BL}}{\Psi_{t+k+1}^{BL}} \left( \frac{D_{t+k+2}}{D_{t+k+1}} \right)^{1-\gamma} \dots \frac{\Psi_{t+k+v}^{BL}}{\Psi_{t+k+v-1}^{BL}} \left( \frac{D_{t+k+v}}{D_{t+k+v-1}} \right)^{1-\gamma} \right] \\ &= \Psi_{t+k}^{BL} E_{t+k} \left[ \frac{\Psi_{t+k+v}^{BL}}{\Psi_{t+k}^{BL}} \left( \frac{D_{t+k+v}}{D_{t+k}} \right)^{1-\gamma} \right] = \Psi_{t+k}^{BL} E_{t+k} \left[ \left( \frac{D_{t+k+v}}{D_{t+k}} \right)^{1-\gamma} \right]\end{aligned}$$

since when  $\pi$  is known  $\Psi_{t+k+v}^{BL} = \Psi_{t+k}^{BL}$ . For  $\widehat{\pi}_{t+k} > \pi$ ,  $C_{t+k}^{BL}(0) - C_{t+k}^{BS}(0) \geq 0$  holds as

$$\widehat{E}_{t+k} \left\{ \widehat{E}_{t+k+1} \left[ \dots \widehat{E}_{t+k+v-1} \left( \Psi_{t+k+v}^{BL} \left( \frac{D_{t+k+v}}{D_{t+k}} \right)^{1-\gamma} \mid \widehat{\pi}_{t+k} \right) \dots \mid \widehat{\pi}_{t+k} \right] \mid \widehat{\pi}_{t+k} \right\}$$

is always bigger than the first line of the previous expression since for  $\widehat{\pi}_{t+k} > \pi$  probability beliefs under Bayesian learning are no longer a pure mean-preserving spread of full information beliefs, but for  $\gamma < 1$  probability mass is transferred from bad to good states of the world in which  $\left( \frac{D_{t+k+v}^s}{D_{t+k}} \right)^{1-\gamma} \Psi_{t+k+v}^{BL}$  takes higher values.

Next consider any  $K \in (0, S_{t+k+v}^{BL}(0))$  :

$$\begin{aligned}\nabla_{BS}^{BL}(K) &= \sum_{s=0}^v \beta^v \left( \frac{D_{t+k+v}^s}{D_{t+k}} \right)^{-\gamma} S_{t+k+v}^{BL}(s) P_{t+k}^{BL}(s) - \sum_{s=0}^v \beta^v \left( \frac{D_{t+k+v}^s}{D_{t+k}} \right)^{-\gamma} S_{t+k+v}^{BS} \binom{v}{s} \pi^s (1-\pi)^{v-s} + \\ &\quad + K \beta^v \sum_{s=0}^v \left( \frac{D_{t+k+v}^s}{D_{t+k}} \right)^{-\gamma} \left[ \binom{v}{s} \pi^s (1-\pi)^{v-s} - P_{t+k}^{BL}(s) \right] > \\ &> K \beta^v \sum_{s=0}^v \left( \frac{D_{t+k+v}^s}{D_{t+k}} \right)^{-\gamma} \left[ \binom{v}{s} \pi^s (1-\pi)^{v-s} - P_{t+k}^{BL}(s) \right] > 0\end{aligned}$$

where we have used that the first two terms of  $\nabla_{BS}^{BL}(K)$  correspond to  $\nabla_{BS}^{BL}(0) > 0$ . The last inequality obtains if and only if  $\widehat{\pi}_{t+k} > \pi$ , as assumed.

It is convenient to proceed backwards for the rest of the proof. Suppose  $K \geq S_{t+k+v}^{BL}(v)$ . Since  $\max\{0, S_{t+k+v}^{BL}(v) - K\} = 0 \implies \max\{0, S_{t+k+v}^{BS}(v) - K\} = 0$  as  $S_{t+k+v}^{BL}(v) > S_{t+k+v}^{BS}(v)$ , the two call prices are both trivially zero and their difference is therefore nonnegative. Hence  $\nabla_{BS}^{BL}(K) = 0$  in the  $[S_{t+k+v}^{BL}(v), +\infty)$  interval.

Consider now the interval  $K \in [S_{t+k+v}^{BS}(v), S_{t+k+v}^{BL}(v))$ . Since  $\max\{0, S_{t+k+v}^{BS}(v) - K\} = 0$  while  $\max\{0, S_{t+k+v}^{BL}(v) - K\} \geq 0$  it follows that while  $C_{t+k}^{BS}(K) = 0$  everywhere,  $C_{t+k}^{BL}(K) > 0$ . Hence  $\nabla_{BS}^{BL}(K) > 0$ . Also,  $\nabla_{BS}^{BL}(K)$  is decreasing in  $K \in [S_{t+k+v}^{BS}(v), S_{t+k+v}^{BL}(v))$  as  $C_{t+k}^{BL}(K)$  is obviously decreasing in the strike price.

We can gain some insight by splitting the third interval for the strike price,  $[S_{t+k+v}^{BS}(0), S_{t+k+v}^{BS}(v))$ , into two sub-intervals. Define  $a^{BS}(K)$  as the smallest natural number s.t.  $(1 + g_h)^a(1 + g_l)^{v-a}S_t > K$  and  $a^{BL}(K)$  as the smallest natural number s.t.  $\frac{\Psi_{t+k+v}^{BL}(a)}{\Psi_{t+k}^{BL}(n_{t+k}, N_{t+k})}(1 + g_h)^a(1 + g_l)^{v-a}S_t > K$ . Observe that as for

$$\begin{aligned} s' &= \text{int}(v\widehat{\pi}_{t+k}) + I_{\{v\widehat{\pi}_{t+k} - \text{int}(v\widehat{\pi}_{t+k}) > 0\}} \geq v\widehat{\pi}_{t+k}\widehat{\pi}_{t+k+v} = \\ &= \frac{N_{t+k}}{N_{t+k} + v}\widehat{\pi}_{t+k} + \frac{1}{N_{t+k} + v}s \geq \frac{N_{t+k}}{N_{t+k} + v}\widehat{\pi}_{t+k} + \frac{1}{N_{t+k} + v}v\widehat{\pi}_{t+k} = \widehat{\pi}_{t+k}, \end{aligned}$$

so  $\frac{\Psi_{t+k+v}^{BL}}{\Psi_{t+k}^{BL}} \geq 1$ . Therefore,  $a^{BL}(K) \geq \text{int}(v\widehat{\pi}_{t+k}) + I_{\{v\widehat{\pi}_{t+k} - \text{int}(v\widehat{\pi}_{t+k}) > 0\}}$  implies that  $a^{BS}(K) \geq a^{BL}(K)$ , while for  $a^{BL}(K) < \text{int}(v\widehat{\pi}_{t+k}) + I_{\{v\widehat{\pi}_{t+k} - \text{int}(v\widehat{\pi}_{t+k}) > 0\}}$   $a^{BS}(K) < a^{BL}(K)$  as  $\frac{\Psi_{t+k+v}^{BL}}{\Psi_{t+k}^{BL}} < 1$ . In other words,  $\exists \overline{K} \in [S_{t+k+v}^{BS}(0), S_{t+k+v}^{BS}(v))$  s.t.  $a^{BL}(\overline{K}) \geq \text{int}(v\widehat{\pi}_{t+k}) + I_{\{v\widehat{\pi}_{t+k} - \text{int}(v\widehat{\pi}_{t+k}) > 0\}}$  so that  $\forall K \geq \overline{K}$   $a^{BS}(K) \geq a^{BL}(K)$  and  $\forall s \geq a^{BL}(K)$   $S_{t+k+v}^{BL}(s) \geq S_{t+k+v}^{BS}(s)$ , while  $\forall K < \overline{K}$   $a^{BS}(K) < a^{BL}(K)$ . Suppose first that  $K \in [\overline{K}, S_{t+k+v}^{BS}(v))$ . Then

$$\begin{aligned} \nabla_{BS}^{BL}(K) &= \sum_{s=a^{BL}(K)}^v \beta^v \left( \frac{D_{t+k+v}^s}{D_t} \right)^{-\gamma} [S_{t+k+v}^{BL} - K] P_{t+k}^{BL}(s) + \\ &\quad - \sum_{s=a^{BL}(K)}^v \beta^v \left( \frac{D_{t+k+v}^s}{D_t} \right)^{-\gamma} \max\{0, S_{t+k+v}^{BS} - K\} \binom{v}{s} \pi^s (1 - \pi)^v \end{aligned}$$

over  $[s_2, v]$  at least compensate the negative values between  $[v\widehat{\pi}_{t+k}, s_2)$ . For  $\widehat{\pi}_{t+k} > \pi$ , the two values of  $s$  such that  $f(s) > 0$  move ‘right’, to  $s'_1 > s_1$  and  $s'_2 > s_2$  as increasing probability mass is shifted from states with low final dividends to states with high final dividends. For the same reason, for  $\pi \geq f(s)$  contains an even larger (positive) weighting mass in the interval  $[v\widehat{\pi}_{t+k}, v]$ . Since the summation in the expression for  $\nabla_{BS}^{BL}(K)$  runs from  $a^{BL}(K)$  to  $v$ , for  $K$  s.t.  $a^{BL}(K) \geq s'_2 > v\widehat{\pi}_{t+k}$ ,  $P_{t+k}^{BL}(s) > \binom{v}{s}\pi^s(1-\pi)^{v-s} \forall s \geq a^{BL}(K)$ , the result obtains. Suppose instead that  $K \in [S_{t+k+v}^{BS}(0), \overline{K})$ . This implies  $a^{BS}(K) < a^{BL}(K)$  while  $\forall s < a^{BL}(\overline{K})$   $S_{t+k+v}^{BS}(s) > S_{t+k+v}^{BL}(s)$ . By the same argument made above,  $C_{t+k}^{BL}(K) - C_{t+k}^{BS}(K)$  can be either increasing or decreasing in  $K$  over this interval. First, note that  $a^{BL}(S_{t+k+v}^{BS}(0))$  is s.t.  $S_{t+k+v}^{BL}(a^{BL}(S_{t+k+v}^{BS}(0))) \geq S_{t+k+v}^{BS}(0)$ , while  $\forall s < a^{BL}(S_{t+k+v}^{BS}(0))$   $S_{t+k+v}^{BL}(s) < S_{t+k+v}^{BS}(0)$ . Given this, we check that  $\nabla_{BS}^{BL}(S_{t+k+v}^{BS}(0))$  is nonnegative:

$$\begin{aligned} \nabla_{BS}^{BL}(S_{t+k+v}^{FRE}(0)) &= \nabla_{BS}^{BL}(S_{t+k+v}^{BL}(0)) - \sum_{s=0}^{a^{BL}(S_{t+k+v}^{BS}(0))-1} \beta^v \left( \frac{D_{t+k+v}^s}{D_t} \right)^{-\gamma} S_{t+k+v}^{BL} P_{t+k}^{BL}(s) + \\ & S_{t+k+v}^{BL}(0) \sum_{s=0}^v \beta^v \left( \frac{D_{t+k+v}^s}{D_t} \right)^{-\gamma} P_{t+k}^{BL}(s) - S_{t+k+v}^{BS}(0) \sum_{s=a^{BL}(S_{t+k+v}^{BS}(0))}^v \beta^v \left( \frac{D_{t+k+v}^s}{D_t} \right)^{-\gamma} P_{t+k}^{BL}(s) + \\ & - S_{t+k+v}^{BL}(0) \sum_{s=0}^v \beta^v \left( \frac{D_{t+k+v}^s}{D_t} \right)^{-\gamma} \binom{v}{s} \pi^s (1-\pi)^{v-s} + S_{t+k+v}^{BS}(0) \sum_{s=0}^v \beta^v \left( \frac{D_{t+k+v}^s}{D_t} \right)^{-\gamma} \binom{v}{s} \pi^s (1-\pi)^{v-s} > \\ & > [S_{t+k+v}^{BS}(0) - S_{t+k+v}^{BL}(0)] \sum_{s=0}^v \beta^v \left( \frac{D_{t+k+v}^s}{D_t} \right)^{-\gamma} \left[ \binom{v}{s} \pi^s (1-\pi)^{v-s} - P_{t+k}^{BL}(s) \right] + \\ & + \sum_{s=0}^{a^{BL}(S_{t+k+v}^{BS}(0))-1} \beta^v \left( \frac{D_{t+k+v}^s}{D_t} \right)^{-\gamma} [S_{t+k+v}^{BS}(0) - S_{t+k+v}^{BL}(0)] P_{t+k}^{BL}(s) > 0 \end{aligned}$$

where the first inequality follows from  $\nabla_{BS}^{BL}(0) > 0$  and the second from the  $[S_{t+k+v}^{BS}(0) - S_{t+k+v}^{BL}(0)] > 0$  for every term in the summation, plus  $\sum_{s=0}^v \left( \frac{D_{t+k+v}^s}{D_{t+k}} \right)^{-\gamma} [\binom{v}{s} \pi^s (1-\pi)^{v-s} - P_{t+k}^{BL}(s)] > 0$ . Consider now a strike  $K = S_{t+k+v}^{BS}(0) + \delta$ ,  $\delta > 0$  such that  $a^{BL}(S_{t+k+v}^{BS}(0) + \delta) = a^{BL}(S_{t+k+v}^{BS}(0))$ . Then  $\nabla_{BS}^{BL}(K)$  equals

$$\begin{aligned} \nabla_{BS}^{BL}(S_{t+k+v}^{BS}(0) + \delta) &= \sum_{s=1}^{a^{BL}(K)-1} \beta^v \left( \frac{D_{t+k+v}^s}{D_{t+k}} \right)^{-\gamma} \binom{v}{s} \pi^s (1-\pi)^{v-s} + \\ & + \delta \sum_{s=a^{BL}(K)}^v \beta^v \left( \frac{D_{t+k+v}^s}{D_{t+k}} \right)^{-\gamma} \left\{ \binom{v}{s} \pi^s (1-\pi)^{v-s} - P_{t+k}^{BL}(s) \right\} \end{aligned}$$

which is positive and increasing in  $K$  since  $\sum_{s=a^{BL}(K)}^v \left( \frac{D_{t+k+v}^s}{D_{t+k}} \right)^{-\gamma} \left\{ \binom{v}{s} \pi^s (1-\pi)^{v-s} - P_{t+k}^{BL}(s) \right\} > 0$  from the fact that with  $a^{BL}(K) < v\widehat{\pi}_{t+k} - \sum_{s=a^{BL}(K)}^v f(s)$  contains a nonnegative probability mass. Repeating the same argument for all the subintervals of  $[S_{t+k+v}^{BS}(0), \overline{K})$  formed



by increasing the strike without letting  $a^{BL}(K)$  change and exploiting the continuity in  $K$  of  $\nabla_{BS}^{BL}(K)$ , we conclude that this function is everywhere positive and increasing in  $[S_{t+k+v}^{BS}(0), \bar{K}]$ . Finally, we are left with the interval  $[S_{t+k+v}^{BL}(0), S_{t+k+v}^{BS}(0)]$ . However,  $\nabla_{BS}^{BL}(K)$  is a continuous function and we have already proven that  $\nabla_{BS}^{BL}(S_{t+k+v}^{BL}(0)) > 0$ . By a similar reasoning it is possible to show that if  $\delta > 0$  is a small number such that  $a^{BL}(S_{t+k+v}^{BL}(0) + \delta) = 1$  and  $K' = S_{t+k+v}^{BL}(0) + \delta$ ,  $\nabla_{BS}^{BL}(K') > \nabla_{BS}^{BL}(S_{t+k+v}^{BL}(0)) > 0$  follows. Therefore  $\nabla_{BS}^{BL}(K)$  is increasing and positive in  $[S_{t+k+v}^{BL}(0), S_{t+k+v}^{BL}(0) + \delta]$ . We can repeat the same argument for all the subintervals of  $[S_{t+k+v}^{BL}(0), S_{t+k+v}^{BS}(0)]$  in which although the strike price increases,  $a^{BL}(K)$  does not change. Since  $\nabla_{BS}^{BL}(K)$  is a continuous function this implies that it is everywhere increasing over  $[S_{t+k+v}^{BL}(0), S_{t+k+v}^{BS}(0)]$ . Since  $\nabla_{BS}^{BL}(S_{t+k+v}^{BL}(0)) > 0$  positivity everywhere obtains. ■

## Appendix B

### Estimating learning parameters using the polytope method

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

$$\min_{\pi_t, N_t, m_t} SSR(\theta_t) \equiv \sum_{\tau_t = \underline{\tau}_t}^{\bar{\tau}_t} \sum_{K_{\tau_t} = \underline{K}_{\tau_t}}^{\bar{K}_{\tau_t}} [\epsilon(\tau_t, K_{\tau_t})]^2$$

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

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  $\theta_t$ ).<sup>37</sup> The simplex comprises four vertices  $\{\theta^a \theta^b \theta^c \theta^d\} \dots$ <sup>38</sup> The initial simplex is composed of four vectors that lie on a three-dimensional

<sup>37</sup>A more complete treatment of polytope methods can be found in Judd (1998) and Walters et al. (1991).

<sup>38</sup>We choose  $\{\theta^a \theta^b \theta^c \theta^d\}$  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.

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.

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  $v_\tau$  on the lattice is given by the integer part of  $\tau m$ . 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.<sup>39</sup> Second, we supplement the combined polytope and grid search with a rough grid search and require that the minimum 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  $\pi$  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  $\hat{\theta}^{(1)}$ :

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

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.

---

<sup>39</sup>We use the following values for  $\{\theta^a \theta^b \theta^c \theta^d\}$ :

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

Figure 1  
Implied Volatility Surface vs. Moneyness

fi

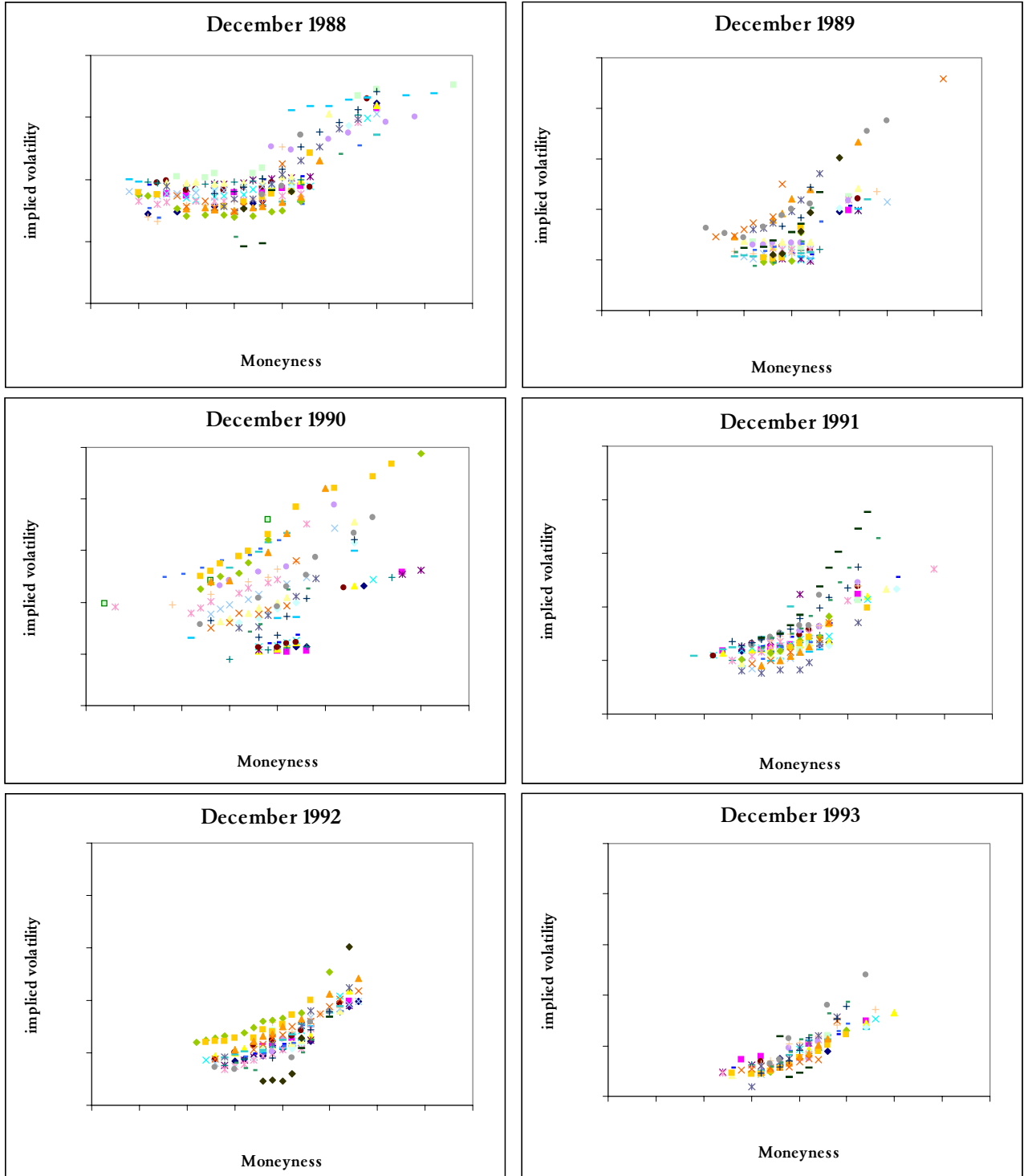
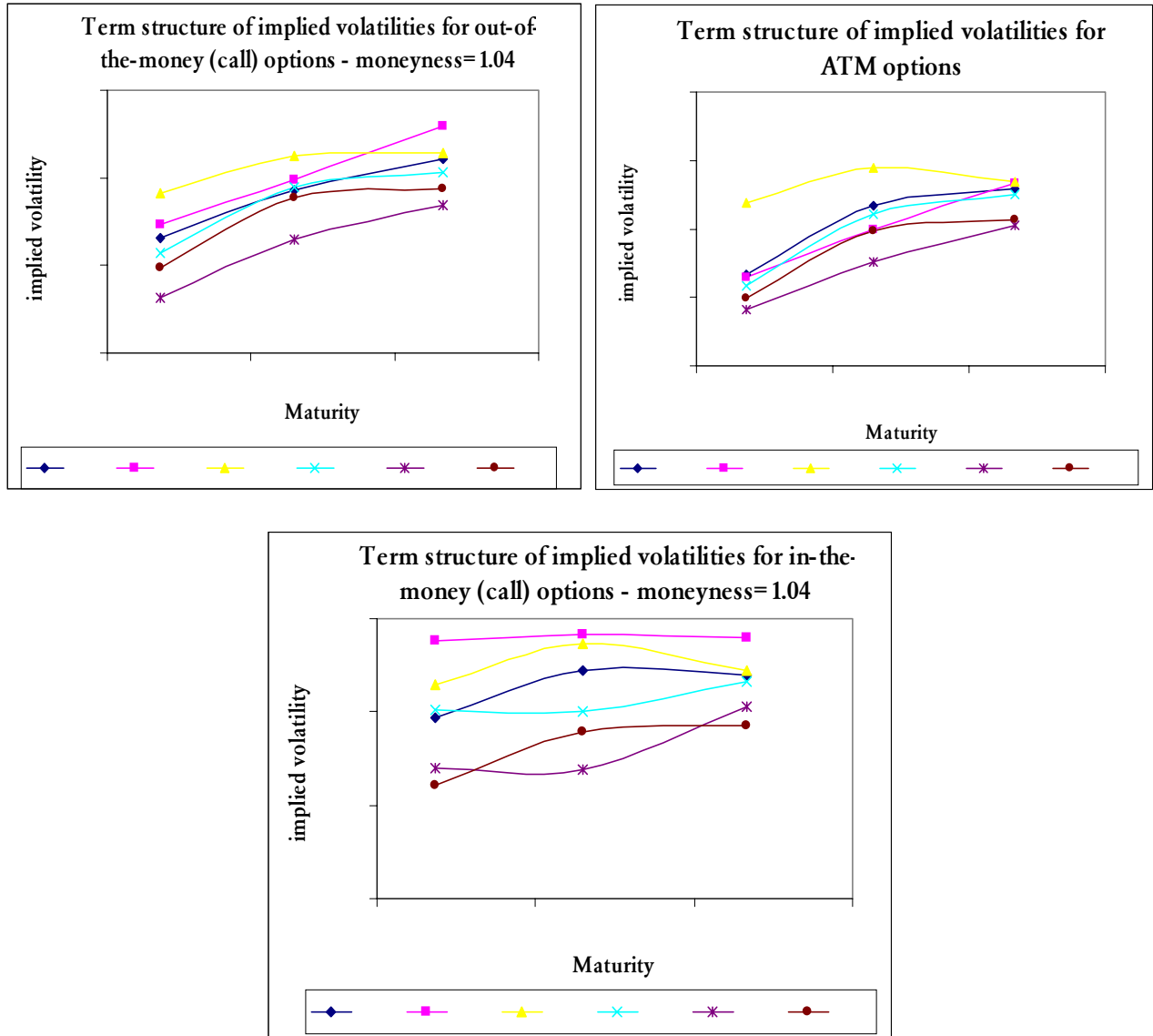


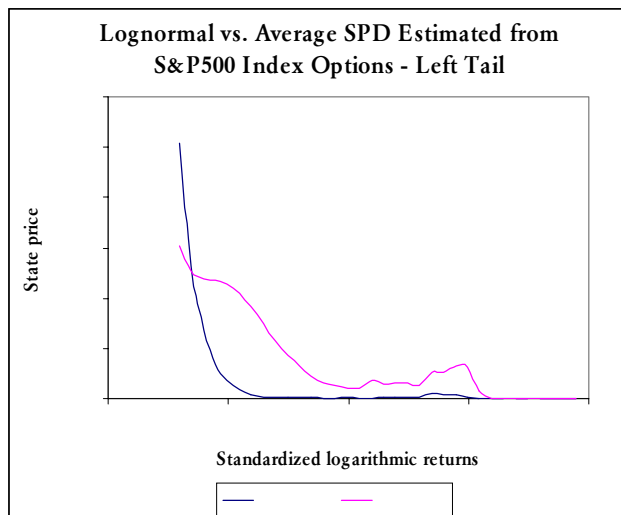
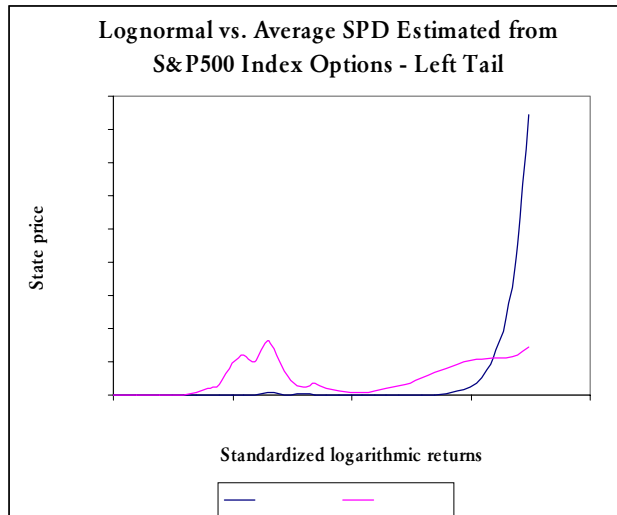
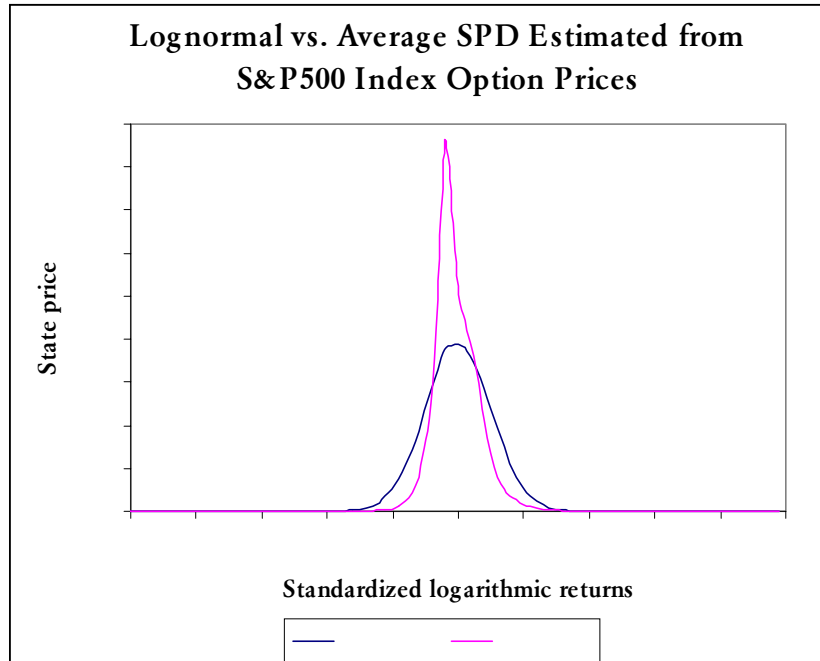
Figure 2  
Implied Volatility Surface vs. Term Structure



**Figure 3**  
**State Price Density Implied by S&P 500 Index Options**

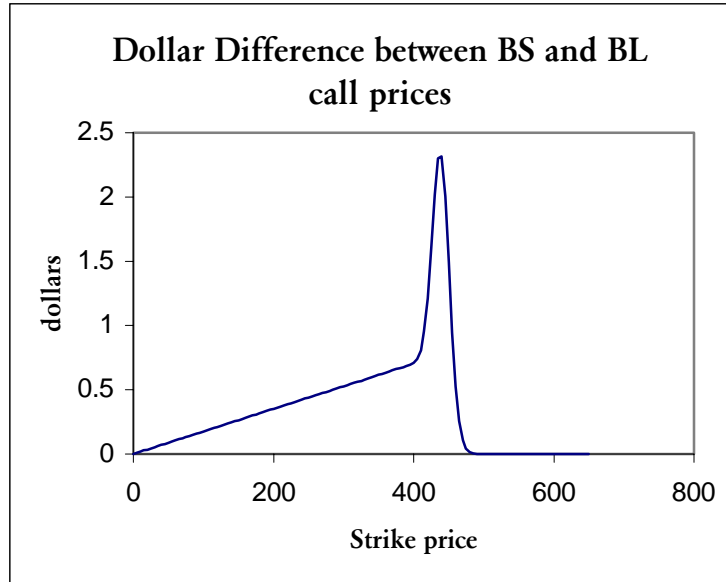
fi

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



**Figure 4**  
**Option Prices under Full Information and Bayesian Learning**

$m$   $\tau$   $\pi$   $\rho$   $\gamma$   
 $\hat{\pi}_i$



**Figure 5**  
**The Implied Volatility Surface under Full Information**

$\gamma$   $m$   $\pi$   $\tau$   $\rho$

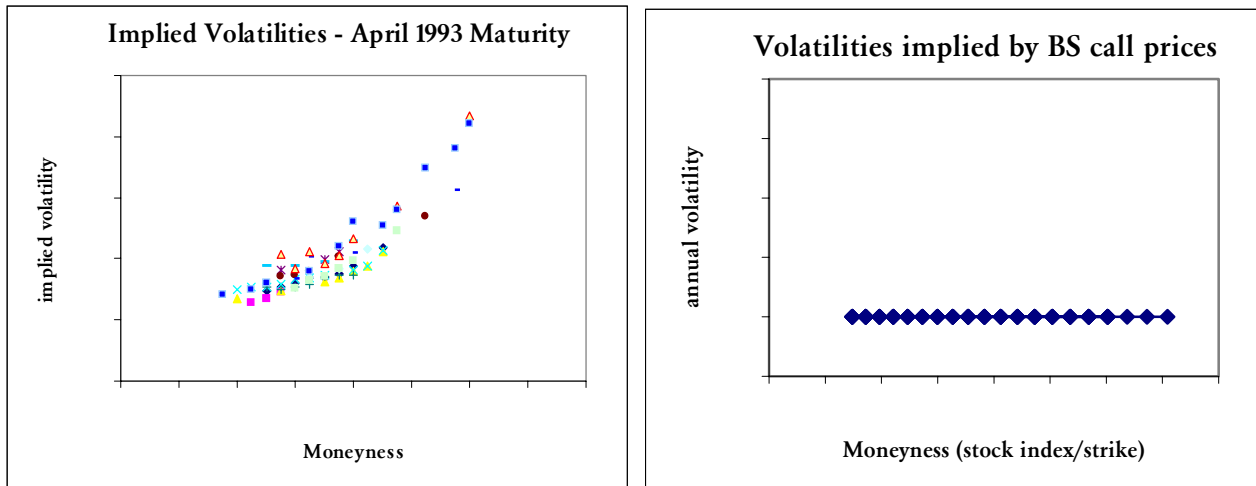


Figure 6  
The Implied Volatility Surface under Bayesian Learning

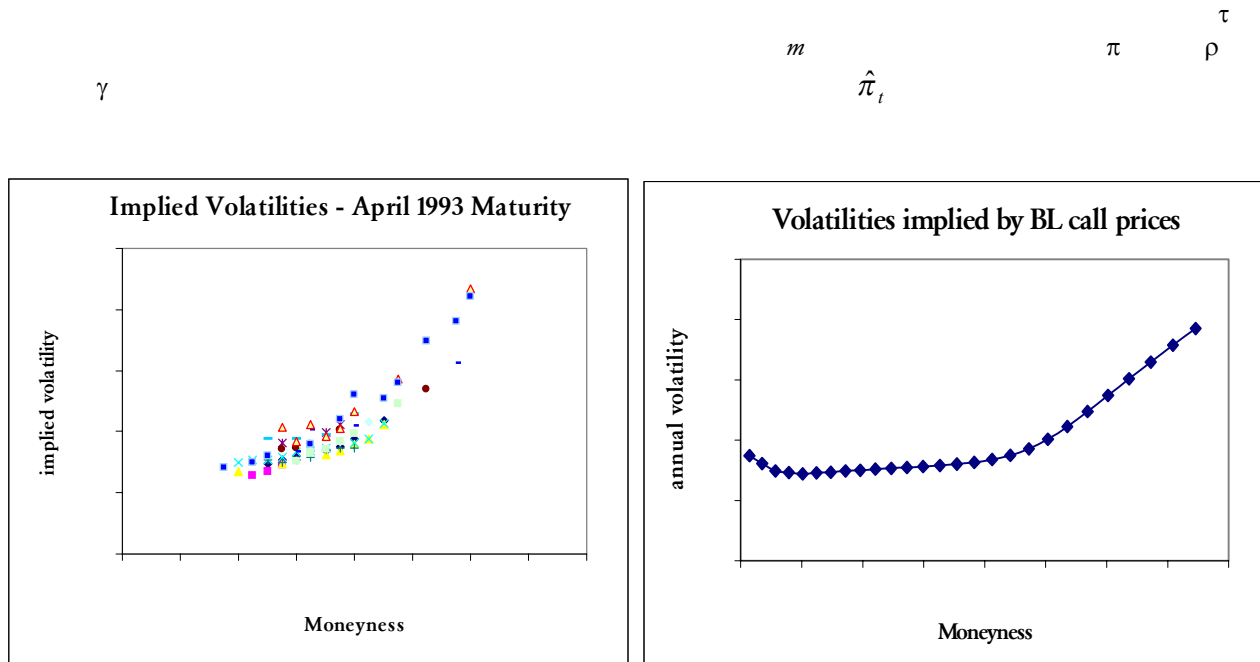
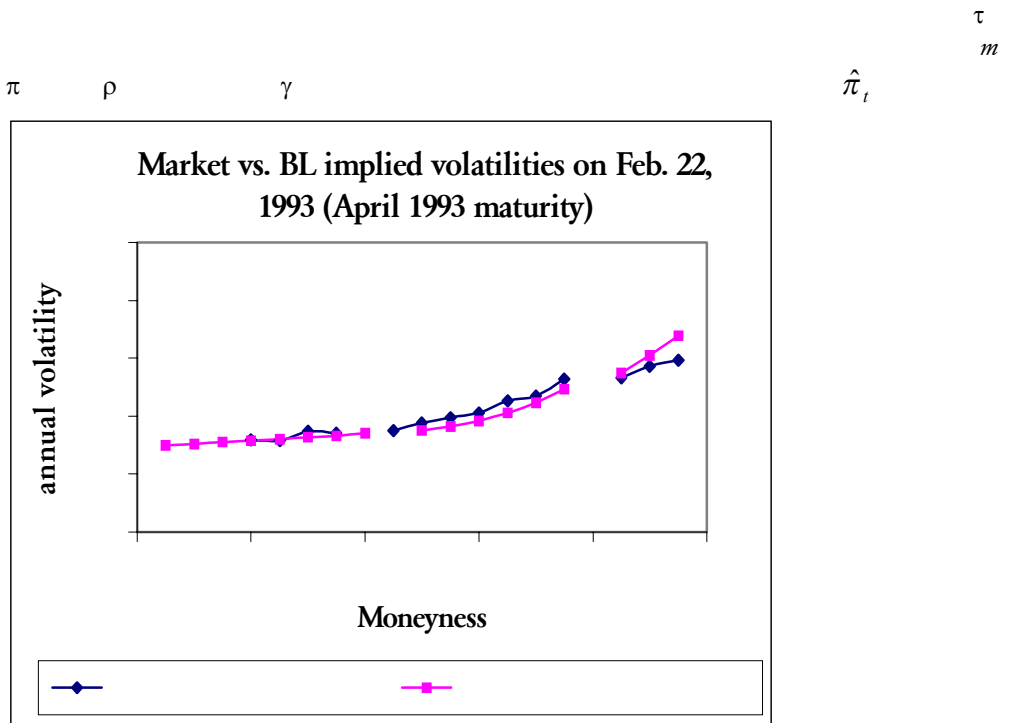


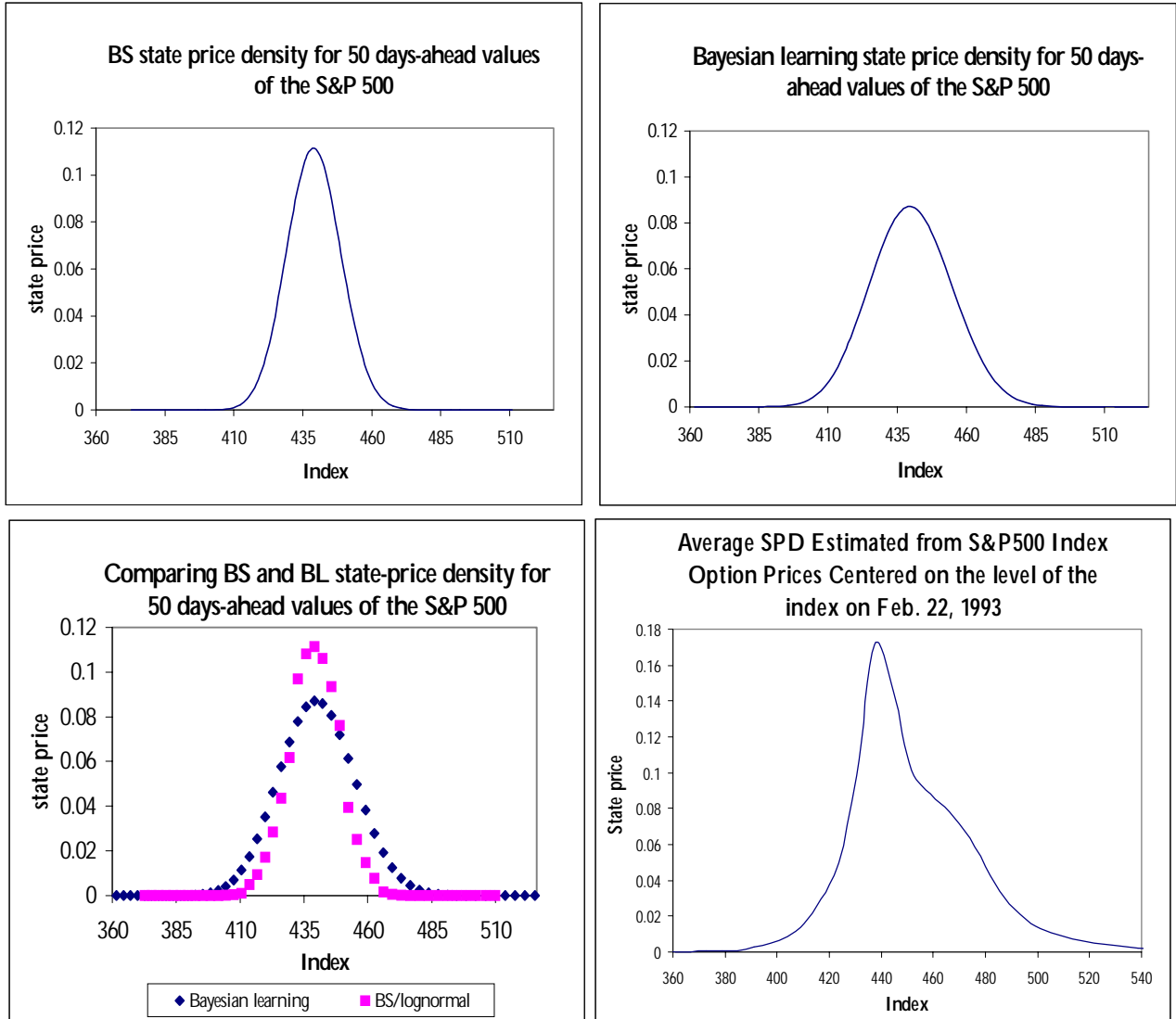
Figure 7  
The Implied Volatility Surface under Bayesian Learning



**Figure 8**

**State-Price Densities under Full Information and Bayesian Learning**

State-price densities for the 50 days ahead ( $\tau=50$ ) 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:  $m = 1$ ,  $g_h = 0.00315$ ,  $g_f = -0.00314$ ,  $\pi = 0.519$ ,  $\rho = 0.02$  (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. For comparison, the fourth panel plots the average empirical SPD estimated in Section 2 when the index takes a value of \$436.38.

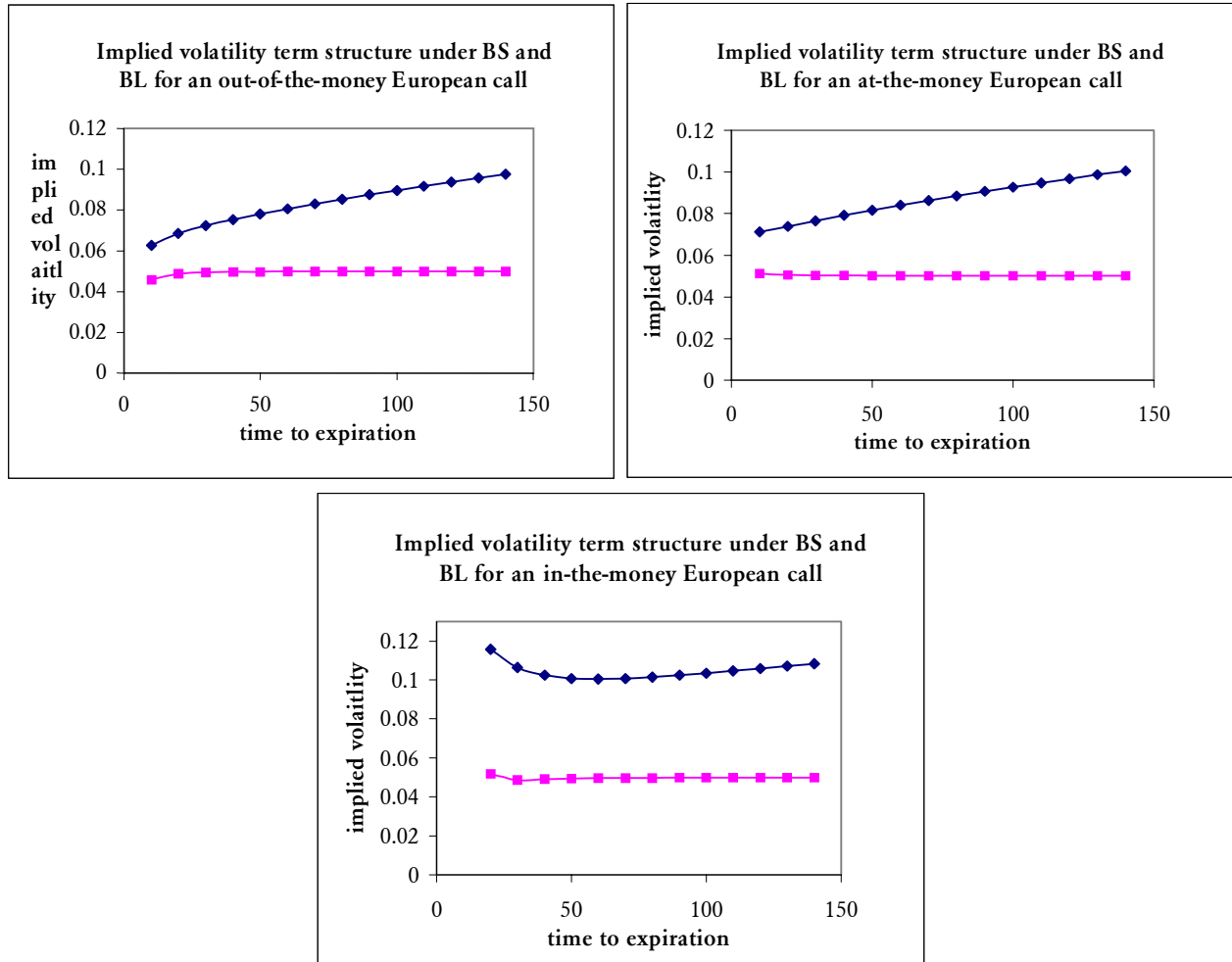




**Figure 9**

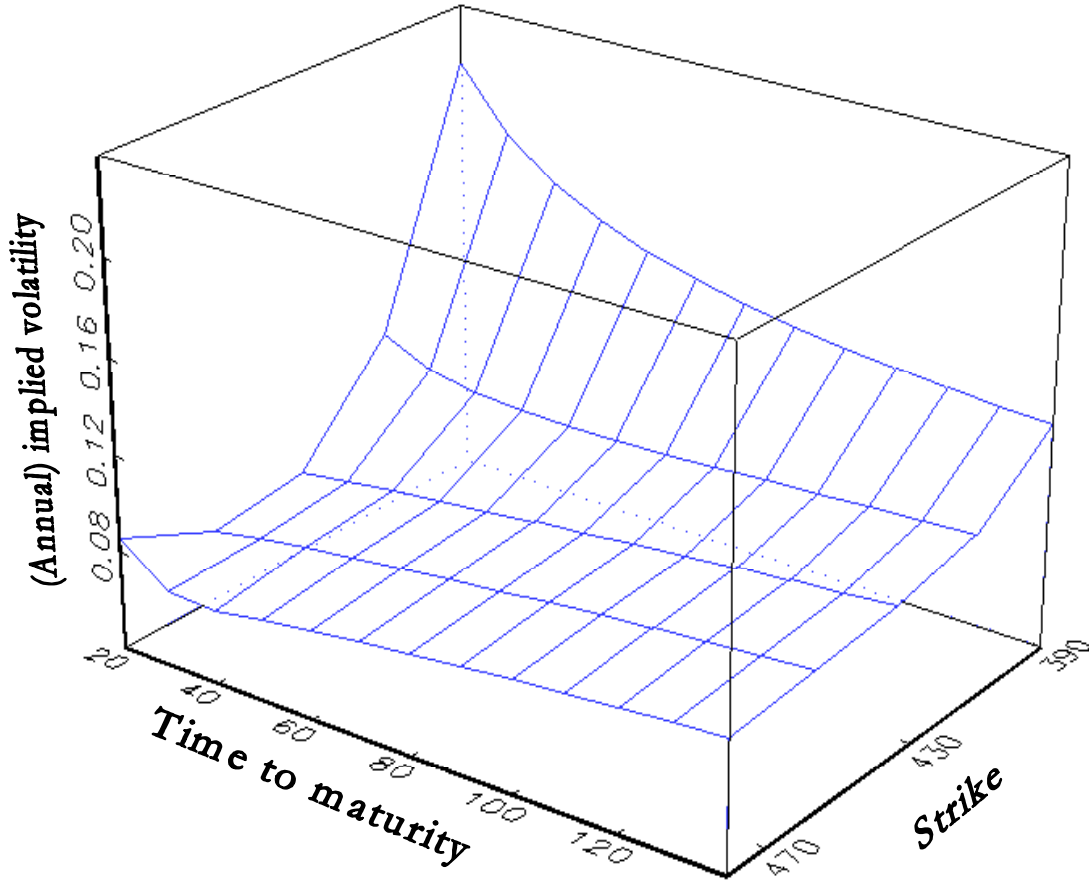
**Implied Volatility Term Structure Under Full Information and Bayesian Learning**

We assume  $m=1$  while the other 'deep' parameters ( $g_b$ ,  $g_s$ ,  $\pi$ ,  $\rho$ , and  $\gamma$ ) 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 BL case, 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 assumed to be \$436.38, the closing price on Feb. 22, 1993. The first panel sets  $K=455$ , the second  $K=435$ , and the third  $K=420$ .



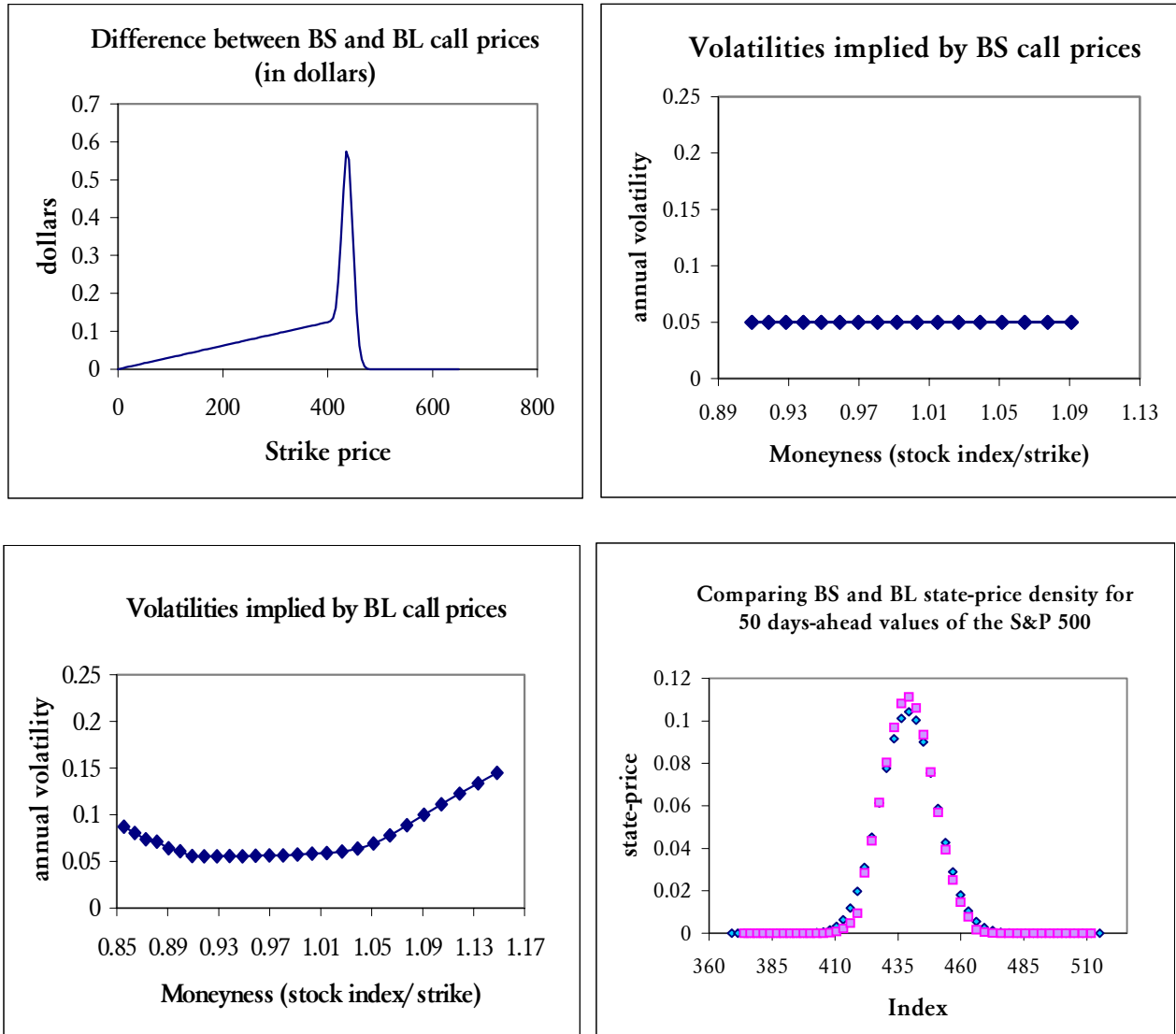
**Figure 10**  
**Implied Volatility Surface under Bayesian Learning**

We plot the annualized implied volatilities as a function of both the strike price and time to maturity when the economy is on a Bayesian learning path. We assume  $n_t=42$  and  $N_t=80$ , implying an initial belief  $\hat{\pi} = 0.525$ . The current price of the S&P 500 is assumed to be \$436.38, the closing price on Feb. 22, 1993.



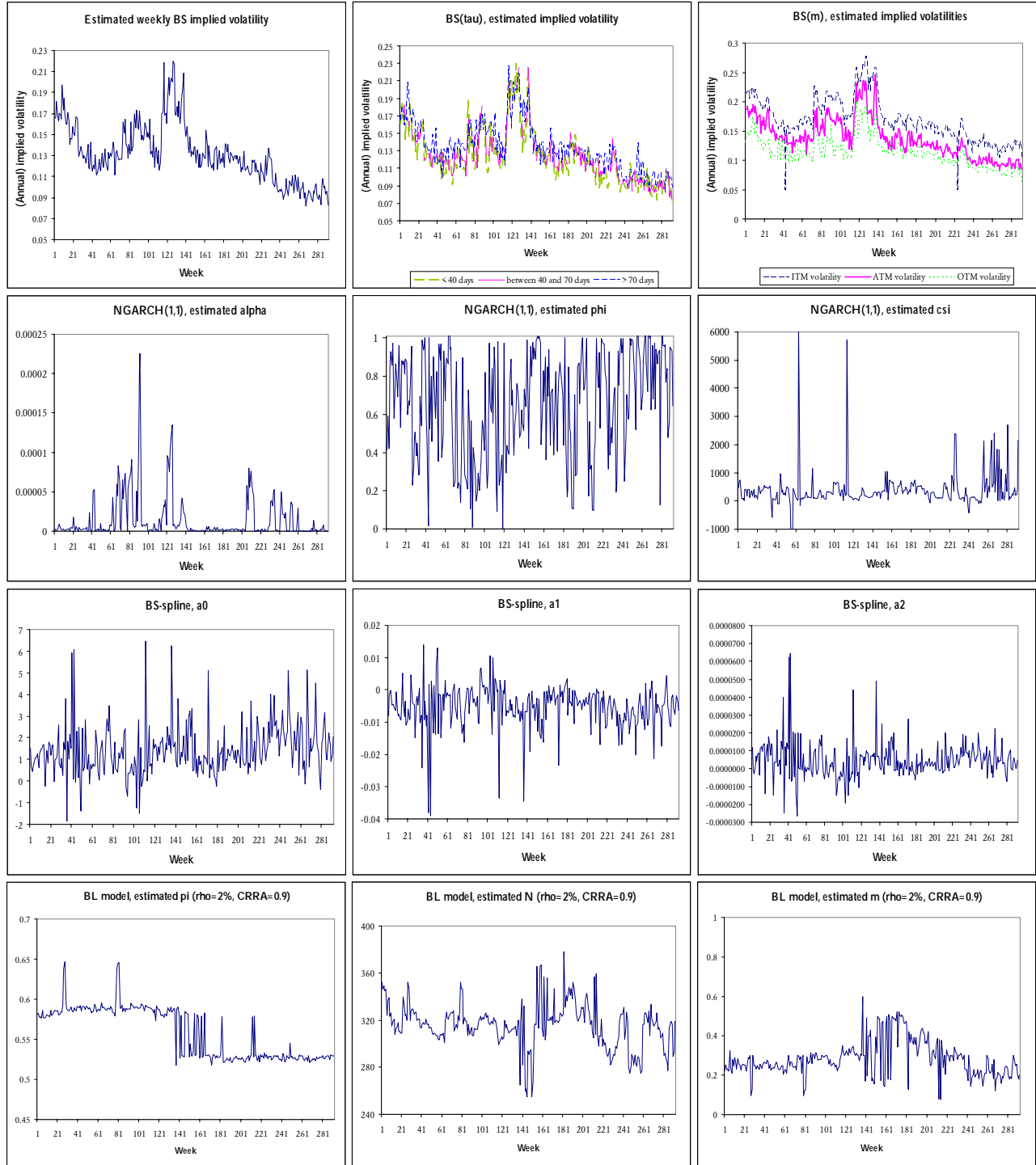
**Figure 11**  
**Weak Learning Effects**

We plot the 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 ( $\tau=50$ ). 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:  $m=1$ ,  $g_h=0.00315$ ,  $g_l=-0.00314$ ,  $\pi=0.519$ ,  $\rho=0.02$  (annual),  $\gamma=0.999$ ,  $n_t=182$ , and  $N_t=350$ , implying an initial unbiased belief  $\hat{\pi}_t=0.52 \cong \pi$ . The current price of the S&P 500 is assumed to be \$436.38 dollars, the closing price on Feb. 22, 1993.



## Figure 12 Estimated Parameter Plots

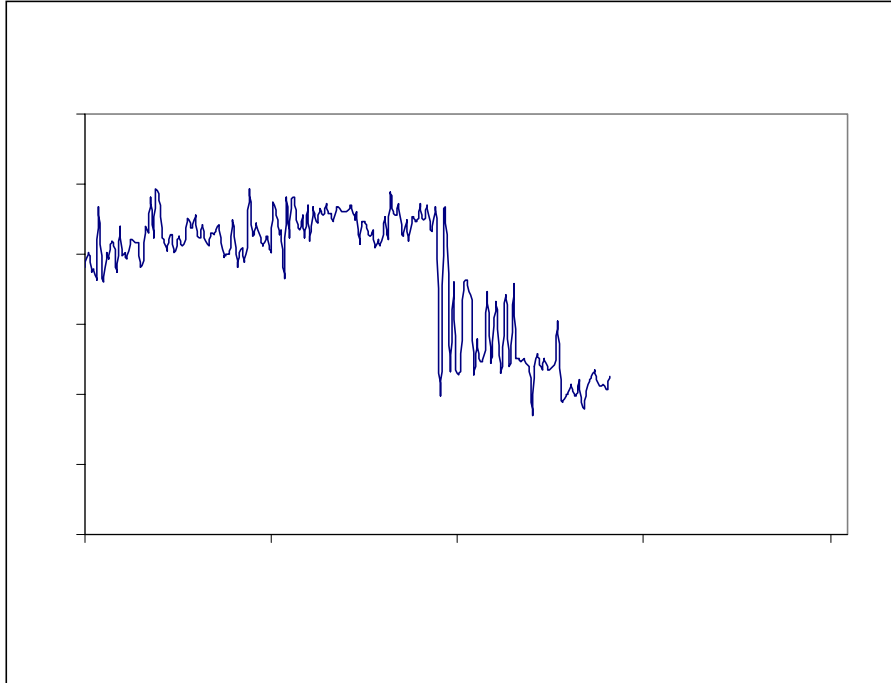
The graphs plot weekly parameter estimates obtained by fitting to the cross section of S&P 500 index option prices each day Black-Scholes (BS), BS with three maturity parameters (BS( $\tau$ )), BS with three moneyness parameters (BS(m)), a BS-spline model (BS-spline) with five parameters, Heston and Nandi,<sup>s</sup> (2000) NGARCH(1,1) model, and a Bayesian learning model (BL) obtained by setting  $\rho = 0.02$  and  $\gamma = 0.9$ . The sample consists of the weekly data for the period June 1988 fi December 1993. In the case of the ,ad hoc,,strawman of Dumas et al. (1993), only the estimates of the coefficients  $\alpha_0$ ,  $\alpha_1$ , and  $\alpha_2$  are reported; for the NGARCH (1,1) only the estimates of  $\alpha$ ,  $\xi$ , and  $\phi$  are plotted.



**Figure 13**

**Implied Risk-free Rate and Expectations on Annual Growth under Bayesian Learning**

The graphs plot weekly values of the implied risk-free interest rate on a one year zero coupon bond as well as implied expectations on the annual rate of growth of fundamentals for an economy on a Bayesian learning path. For each week,  $\pi$ ,  $N$ , and  $m$  are set equal to corresponding estimates obtained from S&P 500 index option prices under the BL model. The estimation sets  $\rho = 0.02$  and  $\gamma = 0.9$ . The sample consists of the weekly data for the period June 1988 to December 1993. In the case of the risk-free rate, the observed values over the sample period are also reported.



### Figure 14 Learning Dynamics on an Expanding Window

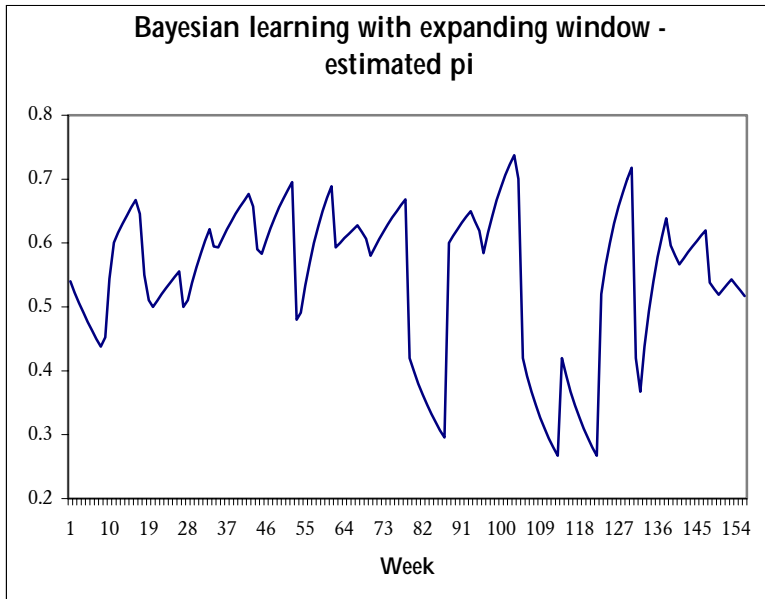
We plot the time series for the probability of a high dividend growth state ( $\hat{\pi}$ ) implied by observed S&P 500 index option prices during the period June 1988 to June 1991. We also report the precision of these beliefs (the number of previously observed states).  $\{\hat{\pi}_t, \hat{n}_t\}_{t=1}$  are estimated by solving the program:

$$\min_{\{\hat{\pi}_t, \hat{n}_t\}_{t=1}} \sum_{t=1}^T [(\tau_t, \pi_t) - (\tau_{t-1}, \pi_{t-1})]^2$$

$$\cdot \frac{\pi_{t-1}}{\hat{n}_t + 1} \leq \pi_{t+1} \leq \frac{\pi_{t-1} + 1}{\hat{n}_t + 1}, 0 \leq \pi_t \leq 1 \quad t \geq 1$$

$$\hat{n}_t > 0, \quad \hat{n}_1 \in \mathbb{N}$$

i.e. by minimizing the (cross-sectional) sum of the squared pricing errors implied by the model under BL on blocks of 2 months length. The exercise assumes  $\rho = 0.02$  (annually), CRRA-power preferences with constant relative risk aversion  $\gamma = 0.9$ , and a volatility of fundamental news  $\sigma = 5\%$  (annualized).



**Table 1**  
**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 7 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  $\sigma\sqrt{\tau}$ . The price of the contingent claims are calculated according to the formula:

$$Q^- = \sum_{j=0}^v \lambda \left( \frac{1}{\sigma\sqrt{\tau}} \left[ \ln \left( \frac{S^j_{t+v}}{S_t} \right) - \mu \right] \right) I_{\left\{ \frac{1}{\sigma\sqrt{\tau}} \left[ \ln \left( \frac{S^j_{t+v}}{S_t} \right) - \mu \right] < X \right\}} \text{ and}$$

$$Q^+ = \sum_{j=0}^v \lambda \left( \frac{1}{\sigma\sqrt{\tau}} \left[ \ln \left( \frac{S^j_{t+v}}{S_t} \right) - \mu \right] \right) I_{\left\{ \frac{1}{\sigma\sqrt{\tau}} \left[ \ln \left( \frac{S^j_{t+v}}{S_t} \right) - \mu \right] > X \right\}}$$

where the  $\lambda_j$ s are the state-prices, I is the standard indicator function,

$$\mu = \sum_{j=0}^v \ln \left( \frac{S^j_{t+v}}{S_t} \right) \text{Pr ob} \{ S^j_{t+v} = u^j d^{v-j} S_t \}, \quad \sigma^2 = \sum_{j=0}^v \left[ \ln \left( \frac{S^j_{t+v}}{S_t} \right) - \mu \right]^2 \text{Pr ob} \{ S^j_{t+v} = u^j d^{v-j} S_t \}$$

and  $\text{Pr ob} \{ S^j_{t+v} = u^j d^{v-j} S_t \}$  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 June 1988 to December 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<	1.42E-05	0.00069	3.81E-13	1.71E-10
-6<	2.01E-05	0.00079	7.35E-10	7.37E-10
-5<	2.08E-05	0.00089	3.33E-07	1.83E-09
$Q^-$ -4<	6.17E-05	0.00128	4.34E-05	3.59E-09
-3<	0.00176	0.00188	0.00186	6.97E-09
-2<	0.02720	0.00446	0.02932	2.06E-08
-1<	0.17622	0.06631	0.18695	4.78E-05
>1	0.13626	0.05123	0.12875	0.04078
>2	0.01852	0.00831	0.01629	0.00239
>3	0.00110	0.00262	0.00080	0.00022
$Q^+$ >4	7.94E-05	0.00136	1.37E-05	5.22E-05
>5	4.83E-05	0.00061	7.23E-08	1.20E-05
>6	4.02E-05	0.00043	9.71E-11	4.44E-06
>7	3.20E-05	0.00028	2.46E-14	2.00E-06

Table 2

**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 June 1988 - December 1993. MAVE is the average of the mean absolute dollar evaluation error. AIC is the average of the Akaike Information Criterion. MOE is the average pricing error outside the bid/ask spread. '% 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', '% best-AIC', and '% best-MOE' are defined analogously. BS is the Black-Scholes model, BS( $\tau$ ) 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  $\tau$  (time-to-maturity) to the implied volatility surface and then prices options using the fitted volatility levels. NGARCH is the option pricing GARCH model proposed by Heston and Nandi (2000). BL is a Bayesian learning model (assuming  $\rho = 0.02$  and  $\gamma = 0.9$ ) re-estimated on a weekly basis. All the models are estimated by minimizing the sum of squared pricing errors. 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	MOE	% best-RMSVE	% best-MAVE	% best-AIC	% best-MOE
BS	1.5048	1.2634	0.7829	1.1098	0%	0%	0.34%	0%
BS ( $\tau$ )	1.4508	1.1916	0.8259	1.0401	0%	0%	0.34%	0%
BS(m)	0.7846	0.5756	-0.3847	0.4364	11.30%	14.04%	19.52%	14.73%
BS spline	0.6240	0.4391	-0.7822	0.3015	53.77%	63.70%	52.06%	55.67%
NGARCH	0.6999	0.5396	-0.5105	0.7141	13.36%	8.56%	10.96%	0%
BL	0.7190	0.5612	-0.5361	0.3512	21.58%	13.70%	17.47%	29.59%

Panel B: Results by Moneyness									
Moneyness (%)									
Model	Less than -2% (OTM)			-2% to +2% (ATM)			More than 2% (ITM)		
	RMSVE	MAVE	MOE	RMSVE	MAVE	MOE	RMSVE	MAVE	MOE
BS	2.155	1.352	1.203	1.494	0.990	0.797	1.184	1.452	1.321
BS ( $\tau$ )	1.481	1.271	1.124	1.078	0.833	0.646	1.638	1.426	1.295
BS(m)	0.422	0.348	0.348	1.087	0.929	0.929	0.685	0.523	0.523
BS spline	0.553	0.407	0.278	0.855	0.709	0.518	0.435	0.313	0.204
NGARCH	0.795	0.664	0.829	0.876	0.757	0.979	0.469	0.417	0.562
BL	0.640	0.484	0.288	0.853	0.665	0.394	0.689	0.548	0.348

Panel C: Results by Time-to-Maturity									
Days to Expiration									
Model	Less than 40			40 to 70			More than 70		
	RMSVE	MAVE	MOE	RMSVE	MAVE	MOE	RMSVE	MAVE	MOE
BS	0.477	0.623	0.450	0.871	0.886	0.645	2.078	1.358	0.999
BS ( $\tau$ )	0.482	0.422	0.334	0.749	0.636	0.523	1.299	1.068	0.925
BS(m)	0.536	0.418	0.330	0.648	0.493	0.382	0.767	0.565	0.431
BS spline	0.273	0.211	0.129	0.344	0.260	0.158	0.565	0.404	0.273
NGARCH	0.450	0.431	0.505	0.520	0.504	0.542	0.684	0.825	0.758
BL	0.439	0.359	0.219	0.507	0.405	0.237	0.695	0.577	0.358



Table 3

**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 ( $\tau$ ) 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  $\tau$  (time-to-maturity) to the implied volatility surface and then prices options using the fitted volatility levels. NGARCH is the option pricing GARCH model proposed by Heston and Nandi (2000). 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	$\sigma$	0.1314	0.1283	0.0274
BS( $m$ )	$\sigma_{OTM}$	0.1126	0.1090	0.0241
	$\sigma_{ATM}$	0.1381	0.1343	0.0335
	$\sigma_{ITM}$	0.1681	0.1620	0.0371
BS( $\tau$ )	$\sigma_{short}$	0.1221	0.1185	0.0333
	$\sigma_{medium}$	0.0997	0.1130	0.0532
	$\sigma_{long}$	0.1329	0.1310	0.0279
<i>BS-spline</i>	$\alpha_0$	1.4299	1.2683	1.1750
	$\alpha_1$	-0.0053	-0.0045	0.0066
	$\alpha_2$	4.80E-06	3.80E-06	10.0E-06
	$\alpha_3$	-0.0019	-0.0014	0.0022
	$\alpha_4$	5.29E-06	4.33E-06	5.67E-06
NGARCH	$\omega$	-7.44E-06	4.53E-07	26.06E-06
	$\alpha$	14.16E-06	2.29E-06	28.80E-06
	$\beta$	0.6560	0.7003	0.2757
	$\gamma$	373.4480	261.1294	651.4054
	$\lambda$	0.6698	0.6768	1.2937
Bayesian learning	$\pi_t$	0.5589	0.5776	0.0315
	$N_t$	315.52	317.00	18.96
	$m_t$	0.2839	0.2676	0.0913

Table 4

**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. MOE is the average pricing error outside the bid/ask spread. '% 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' and '% best-MOE' are defined analogously. BS is the Black-Scholes model, BS( $\tau$ ) 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  $\tau$  (time-to-maturity) to the implied volatility surface and then prices options using the fitted volatility levels. NGARCH is the option pricing GARCH model proposed by Heston and Nandi (2000). BL is a Bayesian learning model (assuming either  $\rho = 0.1$  or  $\rho = 0.05$  and  $\gamma = 0.9$ ) re-estimated on a weekly basis. Moneyness is defined as  $100 \times (F/K - 1)$ , where F is the futures price with maturity identical to the option contract.

<b>Panel A: Aggregate Results</b>						
Model	RMSPE	MAPE	MOE-P	% best-RMSPE	% best-MAPE	% best-MOE-P
BS	1.6767	1.3630	1.1079	0%	0%	0%
BS ( $\tau$ )	1.8873	1.4886	1.0380	0.34%	0%	0%
BS(m)	1.0713	0.7888	0.4367	15.12%	18.56%	17.18%
BS spline	1.0614	0.7321	0.3009	42.61%	48.45%	58.35%
NGARCH	1.1602	0.9010	0.7546	7.22%	5.84%	0.69%
BL	0.9990	0.7704	0.3532	34.71%	27.15%	23.78%

<b>Panel B: Results by Moneyness</b>									
Model	Moneyness (%)								
	Less than -2% (OTM)			-2% to +2% (ATM)			More than 2% (ITM)		
	RMSPE	MAPE	MOE-P	RMSPE	MAPE	MOE-P	RMSPE	MAPE	MOE-P
BS	1.606	1.452	1.201	1.422	1.162	0.796	1.729	1.472	1.320
BS ( $\tau$ )	1.766	1.522	1.122	1.816	1.402	0.644	1.821	1.559	1.294
BS(m)	0.771	0.619	0.228	1.399	1.169	0.739	0.878	0.680	0.403
BS spline	1.003	0.756	0.277	1.129	0.909	0.516	0.820	0.596	0.203
NGARCH	1.166	1.030	0.891	1.195	1.056	0.863	0.896	0.805	0.680
BL	0.897	0.698	0.398	1.158	0.910	0.512	0.899	0.718	0.423

<b>Panel C: Results by Time-to-Maturity</b>									
Model	Days to Expiration								
	Less than 40			40 to 70			More than 70		
	RMSPE	MAPE	MOE-P	RMSPE	MAPE	MOE-P	RMSPE	MAPE	MOE-P
BS	0.602	0.580	0.451	1.053	0.800	0.646	2.543	1.219	0.997
BS ( $\tau$ )	0.635	0.542	0.334	1.376	1.064	0.523	1.682	1.322	0.923
BS(m)	0.626	0.500	0.329	0.770	0.600	0.382	0.989	0.736	0.432
BS spline	0.416	0.347	0.129	0.504	0.408	0.158	1.000	0.635	0.273
NGARCH	0.659	0.644	0.554	0.778	0.763	0.610	1.027	0.908	0.834
BL	0.553	0.460	0.454	0.659	0.526	0.297	0.925	0.734	0.405

Table 5

### 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 ( $\tau$ ) 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  $\tau$  (time-to-maturity) to the implied volatility surface and then prices options using the fitted volatility levels. NGARCH is the option pricing GARCH model proposed by Heston and Nandi (2000). BL is a Bayesian learning model (assuming either  $\rho = 0.1$  or  $\rho = 0.05$  and  $\gamma = 0.9$ ) re-estimated on a weekly basis. Moneyness is defined as  $100 \times (F/K - 1)$ , where F is the futures price with maturity identical to the option contract.

<b>Panel A: Aggregate Results</b>						
Model	RMSHE	MAHE	% best-RMSHE	% best-MAHE		
BS	0.9399	0.6804	2.06%	5.84%		
BS ( $\tau$ )	0.9414	0.6647	6.53%	8.94%		
BS(m)	0.9327	0.6586	13.75%	12.72%		
BS spline	0.7556	0.5233	28.87%	29.90%		
NGARCH	0.8126	0.6082	21.99%	17.87%		
BL	0.7441	0.5404	26.80%	24.74%		
<b>Panel B: Results by Moneyness</b>						
Moneyness (%)						
Model	Less than -2% (OTM)		-2% to +2% (ATM)		More than 2% (ITM)	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
BS	0.792	0.652	1.177	0.919	0.552	0.461
BS ( $\tau$ )	0.815	0.649	1.148	0.870	0.579	0.476
BS(m)	0.615	0.496	1.184	0.945	0.680	0.513
BS spline	0.664	0.522	0.938	0.718	0.449	0.353
NGARCH	0.713	0.646	0.931	0.774	0.479	0.443
BL	0.716	0.574	0.994	0.756	0.592	0.493
<b>Panel C: Results by Time-to-Maturity</b>						
Days to Expiration						
Model	Less than 40		40 to 70		More than 70	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
BS	0.575	0.479	0.659	0.513	0.903	0.650
BS ( $\tau$ )	0.575	0.381	0.696	0.454	1.037	0.633
BS(m)	0.638	0.510	0.736	0.546	0.953	0.672
BS spline	0.415	0.337	0.482	0.354	0.881	0.657
NGARCH	0.598	0.583	0.615	0.511	0.802	0.600
BL	0.537	0.432	0.600	0.455	0.789	0.560