Option Pricing under Regime Switching*

Jin-Chuan Duan[†] Ivilina Popova[‡] Peter Ritchken[§]

October 5, 1999

*The authors thank Stewart Mayhew for helpful conversations and for providing computer programs used to extract some of the empirical data. The paper has benefited from comments of participants at seminars at Yale School of Management, the Fields Institute in Toronto, the 1998 American Finance Association Meetings, the 1998 annual RISK conference in Boston and the IAFE 1999 conference on Computational Intelligence for Financial Engineering in New York City. Duan also acknowledges the support received as Senior Wei Lun Fellow at the Hong Kong University of Science & Technology.

[†]Department of Finance, Hong Kong University of Science & Technology, Clear Water Bay, Kowloon, Hong Kong, Tel: 852-2358 7671, Fax: 852-2358 1749, E-mail:jcduan@ust.hk

[‡]Purdue University, 1310 Krannert Building, West Lafayette, IN, 47907–1310, Phone: (765) 494–9013, Fax: (765) 494–1526, E-mail: ipopova@mgmt.purdue.edu

[§]Case Western Reserve University, WSOM, 10900 Euclid Ave., Cleveland, OH 44106-7235, Phone: (216) 368–3849, Fax: (216) 368–4776, E-mail: phr@pyrite.som.cwru.edu

Abstract

This article develops a family of option pricing models when the underlying stock price dynamic is modeled by a regime switching process in which prices remain in one volatility regime for a random amount of time before switching over into a new regime. Our family includes the regime switching models of Hamilton (1989), in which volatility influences returns. In addition, our models allow for feedback effects from returns to volatilities. Our family also includes GARCH option models as a special limiting case. Our models are more general than GARCH models in that our variance updating schemes do not only depend on levels of volatility and asset innovations, but also allow for a second factor that is orthogonal to asset innovations. The underlying processes in our family capture the asymmetric response of volatility to good and bad news and thus permit negative (or positive) correlation between returns and volatility. We provide the theory for pricing options under such processes, present an analytical solution for the special case where returns provide no feedback to volatility levels, and develop an efficient algorithm for the computation of American option prices for the general case.

This article develops a family of option pricing models obtained when the underlying stock dynamic is modeled by a Markov regime switching process. In such models, the stochastic process remains in one regime for a random amount of time before switching over into a new regime. In our case, the regimes are characterized by different volatility levels. Rather than permitting volatilities to follow a continuous time, continuous state process, as in most stochastic volatility models, our primary focus is on cases where volatilities can take on a finite set of values, and can only switch regimes at finite times. In this regard, our models can be viewed as special cases of the large family of stochastic volatility models, in which the number of distributions for the logarithmic return are constrained to a finite collection. While this may, at first glance, appear unnecessarily restrictive, our family of models includes as special limiting cases, many well-known models, including the family of GARCH option pricing models discussed by Duan (1995). We also can obtain more general limiting models in which variance updating schemes depend not only on levels of variance and on asset innovations, but also on a second factor that is uncorrelated with asset returns. As a result variance levels are not completely determined by the path of prices. This second factor allows for further flexibility in capturing the properties of stock return processes. Our family of models also include the regime switching models of Hamilton (1989) as a special case. His models allow volatility regimes to impact returns, but they do not allow returns to impact future volatilities. Our models do permit this feedback effect. Thus our models allow us to capture the correlation between asset and volatility innovations, or equivalently, the asymmetric volatility response to good and bad news in asset returns.

Our family of models fills the gap of models between the Black Scholes model, which in our framework can be viewed as a single volatility regime model with no feedback effects, and the extended GARCH models, which have infinitely many volatility regimes with feedback effects. We demonstrate that it is possible to establish models with a relatively small number of volatility regimes that produce option prices indistinguishable from models with a continuum of volatility states. We present an algorithm that permits American derivatives to be priced for our most general bi-directional regime switching process. If one is only interested in the rich family of GARCH option models, then our algorithm provides an alternative to the numerical procedure of Ritchken and Trevor (1999) and Duan and Simonato (1999). The addition of the second orthogonal factor causes little complication for the algorithm, and provides meaningful extensions to processes beyond the GARCH family.

Our study is certainly not the first to consider regime switching mechanisms nor option pricing under regime switching. The early regime switching models were primarily designed to capture changes in the underlying economic mechanism that generated the data. Examples include Hamilton (1989) and Gray (1996), Bekaert and Hodrick (1993) and Durland and McCurdy (1994). Recently, attention has been placed on volatility regime switching models, solely for the purpose of better understanding option price behavior. Bollen, Gray and Whaley (1999), for example, show that a very simple regime switching model with independent shifts in the mean and variance dominate a range of GARCH models in the foreign exchange market. Bollen (1998) presents a lattice based algorithm that permits American options to be priced for these regime switching models. The type of regime switching models that are typically considered assume that asset innovations have no feedback effects on volatilities. Further, the assumption that regime shift risk is not priced is made so as to allow option pricing to proceed in the usual risk neutral manner. Our models weaken these restrictions.

The paper proceeds as follows. In section 1 we present the bi-directional regime switching model for asset returns and describe some of its properties. In section 2 we investigate how options can be priced when the underlying follows a regime switching process with feedback. In section 3 we investigate a special case of the model where there are only two volatility regimes and asset innovations have no feedback effects on volatilities. The resulting option model turns out to be a weighted sum of Black and Scholes prices. In section 4 we investigate a second special case that leads to models which include GARCH and stochastic volatility models as limiting cases. In section 5 we provide an efficient numerical scheme for pricing European and American options under our most general bi-directional regime switching process. We illustrate how regime switching models with relatively few volatility states can serve as excellent proxies for GARCH models. We also demonstrate how the second orthogonal factor provides significant flexibility beyond GARCH models in the shapes that return distributions can take on over the lifetime of the option. While there are a huge number of models in the regime switching family, the specific models that we evaluate have up to six unknown parameters, with the simplest model containing just four. In section 6 we investigate the performance of a few specifications of our regime switching models on $S\&P$ 500 stock index option prices. The example provides sufficiently encouraging results to warrant ongoing empirical research in this area.

1 Regime Switching Model With Feedback Effects

We assume that the asset price is governed by a regime switching with feedback dynamic. Let S_t be the asset price at date t, and let σ_{t+1}^2 be the conditional variance of the logarithmic return at date t that holds for the period $[t, t+1]$. Given σ_{t+1}, the dynamics of the price over the next period is assumed to be:

$$\ln \frac{S_{t+1}}{S_t} = r + \lambda\sigma_{t+1} - \frac{1}{2}\sigma_{t+1}^2 + \sigma_{t+1}\varepsilon_{t+1} \tag{1}$$

where ε_{t+1} is a standard normal random variable and λ can be interpreted as the risk premium per one standard deviation. We shall assume that there are K distinct volatility regimes, and in each period there is a chance that the volatility will move into a new regime. The volatility follows a Markov chain, which is fully determined by the $K \times K$ transition matrix between volatility states. The transition probabilities are determined by a threshold model. In particular, the volatility, σ_{t+1}, depends on its previous value, σ_t, on the most recent return innovation, ε_t, and on a variable, ξ_t, that is independent of ε_t. The random process ξ_t can be viewed as a state variable process that impacts the variance but is orthogonal to the asset return innovation process, ε_t.

Let $F(\varepsilon_t, \xi_t)$ be a function, that determines the impact of the most recent return innovation and orthogonal volatility innovation, in the form of a non negative real number. The new volatility state is completely determined by this functional value together with the existing level of volatility. Specifically, corresponding to each volatility level, δ_i, $i = 1, 2, \cdots, K$, is a set of values $\{c_0(\delta_i), c_1(\delta_i), ..., c_K(\delta_i)\}$ such that $c_0(\delta_i) = 0$ and $c_K(\delta_i) = \infty$. At date t, we have, for $i = 1, 2, \cdots, K$:

$$\sigma_{t+1} = \delta_i \quad if \quad c_{i-1}(\sigma_t) \cdot \ \ F(\varepsilon_t, \xi_t) < c_i(\sigma_t) \tag{2}$$

That is, conditional on the current level of volatility, the switch into a new regime, is completely determined by the magnitude of the functional value, $F(\cdot)$. While the updating function, $F(.)$ could be fairly general, to make matters specific, we will assume that it has the structure:

$$F(\varepsilon_t, \xi_t) = q_1(\varepsilon_t - \omega)^+ + q_2(\varepsilon_t - \omega)^- + (1 - q_1 - q_2)|\xi_t| \tag{3}$$

where $(\varepsilon_t - \omega)^+ \equiv \max(\varepsilon_t - \omega, 0)$, $(\varepsilon_t - \omega)^- \equiv \max(\omega - \varepsilon_t, 0)$, $q_1 \geq 0$, $q_2 \geq 0$, and $q_1 + q_2 \cdot \ 1$. This formulation essentially uses the weighted sum of the positive and negative parts of the return innovation and the orthogonal state variable innovation in determining the next volatility state. The return innovation is first subject to a bias adjustment, i.e., ω. This bias adjustment can potentially induce an asymmetric volatility response to a return innovation. A difference in q_1 and q_2 can also induce an asymmetric volatility response to a return innovation. The reason for incorporating two different channels for asymmetric volatility response will become clear later when we discuss limiting properties of the regime switching model. The first channel for asymmetry leads us to the NGARCH model of Engle and Ng (1993) as a limit, whereas the second one yields the GJR-GARCH model of Glosten, *et al.* (1993) model.

When $q_1 = q_2 = 0$, the variance updating scheme is completely determined by the state variable innovation. In this case, variances impact returns but return innovations have no impact on variances. In other words the interaction between returns and volatilities only occurs in one direction. This special case reflects the essence of the regime switching model pioneered by Hamilton (1989) for econometric analysis. Empirical evidence exhibited by asset returns in most markets is, however, at odds with this uni-directional feature of the model. Equity returns are, for example, known to depend on past return innovations with volatility being persistent. Indeed, it is well documented that large absolute returns are more likely to be followed by large absolute returns. This suggests that $q_1 \neq 0$ and $q_2 \neq 0$. In general, the parameters q_1 and q_2 permit past return innovations to feedback into the volatility process.

In addition, returns in many markets are strongly asymmetric. Negative returns are followed by larger increases in volatility than equally large positive returns. This implies that there is a negative correlation between the asset return innovation and its volatility innovation. Following Black's (1976) exploration of this phenomenon, it is now commonly referred to as the leverage effect. As stated earlier, this phenomenon can be captured by our models in two different ways. First, the parameter ω differentiates the impact of negative innovations from positive ones, and hence allows us to capture the leverage effect $(\omega > 0)$. If $\omega = 0$, $q_2 > q_1$ provides an alternative way that makes the impact of a negative return innovation more pronounced in comparison to a positive one of equal magnitude.

Our full model for the asset price dynamic is given by:

$$\ln \frac{S_{t+1}}{S_t} = r + \lambda\sigma_{t+1} - \frac{1}{2}\sigma_{t+1}^2 + \sigma_{t+1}\varepsilon_{t+1} \tag{4}$$

$$\sigma_{t+1} = \delta_i \quad if \quad c_{i-1}(\sigma_t) \cdot \ q_1(\varepsilon_t - \omega)^+ + q_2(\varepsilon_t - \omega)^- + (1 - q_1 - q_2)|\xi_t| < c_i(\sigma_t) \tag{5}$$

for $i = 1, 2, \cdots, K$ and $t \in \{0, 1, \cdots, T-1\}$; $\quad 0 \cdot \ q_1, 0 \cdot \ q_2, q_1 + q_2 \cdot \ 1$, and

$$\begin{bmatrix} \varepsilon_{t+1} \\ \xi_{t+1} \end{bmatrix} |\mathcal{F}_t \overset{P}{\sim} N(\mathbf{0}_{2\times1}, \mathbf{I}_{2\times2}) \tag{6}$$

where P stands for the data generating probability measure and $\mathcal{F}_t$ is the information set up to and including time t, i.e., $\mathcal{F}_t$ is the σ-algebra generated by $\{S_0, \sigma_1, \varepsilon_s, \xi_s : s \in \{1, 2, \cdots, t\}\}$. $\mathbf{0}_{2\times1}$ denotes a two-dimensional column vector of 0 and $\mathbf{I}_{2\times2}$ the 2×2 identity matrix.

To better understand the regime switching with feedback model, consider the case $K = 2$ where the volatility follows a two-state process. For this case, we can compute the 2×2 transition matrix for the volatility conditional on ε_t. Specifically, we have

$$\begin{aligned}
& \Pr{}^P\{\sigma_{t+1} = \delta_1 | \sigma_t = \delta_1, \varepsilon_t\} \\
= \ & \Pr{}^P\{0 \cdot \ q_1(\varepsilon_t - \omega)^+ + q_2(\varepsilon_t - \omega)^- + (1 - q_1 - q_2)|\xi_t| < c_1(\delta_1) | \sigma_t = \delta_1, \varepsilon_t\} \\
= \ & N\left(\frac{c_1(\delta_1) - q_1(\varepsilon_t - \omega)^+ - q_2(\varepsilon_t - \omega)^-}{1 - q_1 - q_2}\right) - N\left(\frac{-c_1(\delta_1) + q_1(\varepsilon_t - \omega)^+ + q_2(\varepsilon_t - \omega)^-}{1 - q_1 - q_2}\right) \\
& \Pr{}^P\{\sigma_{t+1} = \delta_1 | \sigma_t = \delta_2, \varepsilon_t\} \\
= \ & \Pr{}^P\{0 \cdot \ q_1(\varepsilon_t - \omega)^+ + q_2(\varepsilon_t - \omega)^- + (1 - q_1 - q_2)|\xi_t| < c_1(\delta_2) | \sigma_t = \delta_2, \varepsilon_t\} \\
= \ & N\left(\frac{c_1(\delta_2) - q_1(\varepsilon_t - \omega)^+ - q_2(\varepsilon_t - \omega)^-}{1 - q_1 - q_2}\right) - N\left(\frac{-c_1(\delta_2) + q_1(\varepsilon_t - \omega)^+ + q_2(\varepsilon_t - \omega)^-}{1 - q_1 - q_2}\right)
\end{aligned}$$

When $q_1 + q_2 = 1$, the values are taken as the limit of $(q_1 + q_2)$ approaching one. The transition probability matrix from t to $t + 1$, conditional on ε_t, can be expressed as

$$\begin{bmatrix} \Pr{}^P\{\sigma_{t+1} = \delta_1 | \sigma_t = \delta_1, \varepsilon_t\}, & 1 - \Pr{}^P\{\sigma_{t+1} = \delta_1 | \sigma_t = \delta_1, \varepsilon_t\} \\ \Pr{}^P\{\sigma_{t+1} = \delta_1 | \sigma_t = \delta_2, \varepsilon_t\}, & 1 - \Pr{}^P\{\sigma_{t+1} = \delta_1 | \sigma_t = \delta_2, \varepsilon_t\} \end{bmatrix}$$

At first glance, from the perspective of examining volatility alone, the system appears overparameterized because a two-state Markov transition probability matrix has only two degrees of freedom yet here there are five parameters in the transition matrix, i.e., q_1, q_2, ω, $c_1(\delta_1)$ and $c_1(\delta_2)$. However, from the standpoint of characterizing the overall system of both returns and volatility, the model is not overparameterized. In fact, for the case $q_1 > 0$ and/or $q_2 > 0$ we have generalized the standard two-state Markov variance switching system by building in a feedback (potentially asymmetrical) mechanism so that return innovations can impact volatilities. In contrast to the standard regime switching model, the transition probability matrix is time-varying and its entries depend on the return just realized.

2 Option Pricing Under Regime Switching With Feedback

For option pricing in this framework, Duan's (1995) local risk-neutralization principle can be extended to incorporate the effects of the orthogonal volatility factor. In particular, the proposition below provides conditions under which option contracts can be priced using an equivalent measure as if the economic agents were neutral to risk locally.

Proposition 1 *If the dynamics of the underlying asset and variance are given as in equations (4) and (5), and if the representative agent is an expected utility maximizer (time-separable and additive preferences) with any of the following three conditions holding:*

(i) *the utility function has constant relative risk aversion and changes in the logarithmic aggregate consumption have conditional normal distributions with constant mean and variance under measure P;*

(ii) *the utility function is of constant absolute risk aversion and changes in the aggregate consumption have conditional normal distributions with constant mean and variance under measure P;*

(iii) *the utility function is linear,*

then the asset price and variance dynamics under the equilibrium price measure Q become:

$$\ln \frac{S_{t+1}}{S_t} = r - \frac{1}{2}\sigma_{t+1}^2 + \sigma_{t+1}\varepsilon_{t+1}^* \tag{7}$$

$$\sigma_{t+1} = \delta_i \quad \text{if } c_{i-1}(\sigma_t) \cdot \quad q_1(\varepsilon_t^* - \omega - \lambda)^+ + q_2(\varepsilon_t^* - \omega - \lambda)^- + (1 - q_1 - q_2)|\xi_t^* - v_t| < c_i(\sigma_t) \tag{8}$$

for $i = 1, 2, \cdots, K$ and $t \in \{0, 1, \cdots, T-1\}$,

where

$$\begin{bmatrix} \varepsilon_{t+1}^* \\ \xi_{t+1}^* \end{bmatrix} \equiv \begin{bmatrix} \varepsilon_{t+1} + \lambda \\ \xi_{t+1} + v_{t+1} \end{bmatrix} |\mathcal{F}_t \overset{Q}{\sim} N(\mathbf{0}_{2\times 1}, \mathbf{I}_{2\times 2}) \text{ for some } \mathcal{F}_t\text{-measurable } v_{t+1} \tag{9}$$

Proof: See Appendix.

The parameter λ is a unit risk premium for the asset that can be identified by performing a statistical analysis of return data under measure P. This parameter can also be viewed as a result of locally risk-neutralizing ε_t. The predictable stochastic process v_t, on the other hand, arises from locally risk-neutralizing ξ_t. This conditional parameter can be interpreted as the orthogonal volatility risk premium (being stochastic in general), because it pertains to the volatility risk arising from a source that is entirely independent of asset returns. For the remainder of this paper, we assume that v_t is a constant, i.e., $v_t = v$. This assumption amounts to imposing a constant correlation over time between the marginal rate of substitution (logarithmic) and the orthogonal volatility innovation. If $q_1 + q_2 = 1$, the orthogonal volatility risk premium becomes irrelevant, and the option pricing system is completely characterized by the parameters identifiable using the data generating system under measure P. As stated in Duan (1995), the constant mean and variance assumption for the aggregate consumption in the proposition is merely used to ensure a constant equilibrium interest rate. With the locally risk-neutralized system in place, we can price contingent claims, European or American style, using a lattice scheme described later.

For pricing purposes, the transition matrix under Q has the same form as that under the data generating probability measure, except all innovation terms in the volatility dynamic undergo a mean shift due to local risk-neutralization. The option pricing system under the two-state bi-directional regime switching model has 9 parameters: δ_1, δ_2, $c_1(\delta_1)$, $c_1(\delta_2)$, q_1, q_2, ω, λ and v. For the special case of the uni-directional model, the whole system contains 5 parameters: δ_1, δ_2, $c_1(\delta_1)$, $c_1(\delta_2)$, and v. If we use the pricing system solely on option data without referring back to the underlying asset return, there will be one parameter that cannot be determined by the system. Specifically, the three parameters: $c_1(\delta_1)$, $c_1(\delta_2)$ and v are used to determine the 2×2 transition probability matrix under the risk-neutralized pricing measure. Since the transition matrix has only two degrees of freedom, two parameters can be pinned down when only option data are used to imply out the model parameter values. For convenience, one may choose to set $v = 0$ when deal with option data only without impeding the uni-directional model's performance. If one uses historical price data in conjunction with option data, then $c_1(\delta_1)$ and $c_1(\delta_2)$ will be identified under the data generating measure and the value for parameter v can then be implied out by option prices.

The parameter indeterminacy also occurs for the general two-state regime switching model if one is restricted to using the option data in implying out the model parameters. First, only the sum of ω and λ can be determined. Second, v cannot be separated from $c_1(\delta_1)$ and $c_1(\delta_2)$ for the same reason as described earlier under the uni-directional model. From an application point of view, one may conveniently set $\omega = 0$ and $v = 0$ without in any way impeding the performance of the model for option pricing.

So far very little structure has been imposed on the dynamic of the variance updating mechanism. For example, no structure has been imposed on the functions $\{c_i(\sigma_t) | i = 1, 2, ..., n-1\}$. If we want to reduce the number of parameters in the option pricing model, then additional structure needs to be imposed on the model. In section 4 we will consider a model with additional structure that significantly reduces the number of parameters.

3 The Two-State uni-directional Regime Switching Model

When the feedback mechanism of the regime switching model is switched off, i.e., $q_1 = q_2 = 0$, a closed-form solution for European options can be derived. For the illustration purpose, we focus here on the standard two-state variance switching model in the spirit of Hamilton (1989) where the asset price dynamic under the data generating measure P is

$$\ln \frac{S_{t+1}}{S_t} = r + \lambda\sigma_{t+1} - \frac{1}{2}\sigma_{t+1}^2 + \sigma_{t+1}\varepsilon_{t+1} \tag{10}$$

$$\sigma_{t+1} = \delta_i \quad \text{if} \quad c_{i-1}(\sigma_t) \cdot \quad |\xi_t| < c_i(\sigma_t) \text{ for } i = 1, 2. \tag{11}$$

where

$$\begin{bmatrix} \varepsilon_{t+1} \\ \xi_{t+1} \end{bmatrix} |\mathcal{F}_t \overset{P}{\sim} N(\mathbf{0}_{2\times 1}, \mathbf{I}_{2\times 2}) \tag{12}$$

Note that in this two-state case, $c_0(\sigma_t) = 0$ and $c_2(\sigma_t) = \infty$. The volatility transition probability matrix is

$$\begin{bmatrix} N(c_1(\delta_1)) - N(-c_1(\delta_1)), & 1 - N(c_1(\delta_1)) + N(-c_1(\delta_1)) \\ N(c_1(\delta_2)) - N(-c_1(\delta_2)), & 1 - N(c_1(\delta_2)) + N(-c_1(\delta_2)) \end{bmatrix}$$

Local risk-neutralization leads to a different dynamics under measure Q. By Proposition 1 and the maintained assumption of constant v, we have

$$\ln \frac{S_{t+1}}{S_t} = r - \frac{1}{2}\sigma_{t+1}^2 + \sigma_{t+1}\varepsilon_{t+1}^* \tag{13}$$

$$\sigma_{t+1} = \delta_i \quad \text{if} \quad c_{i-1}(\sigma_t) \cdot \quad |\xi_t^* - v| < c_i(\sigma_t) \text{ for } i = 1, 2. \tag{14}$$

where

$$\begin{bmatrix} \varepsilon_{t+1}^* \\ \xi_{t+1}^* \end{bmatrix} |\mathcal{F}_t \overset{Q}{\sim} N(\mathbf{0}_{2\times 1}, \mathbf{I}_{2\times 2}) \tag{15}$$

and the transition probability matrix becomes

$$\begin{bmatrix} N(v + c_1(\delta_1)) - N(v - c_1(\delta_1)), & 1 - N(v + c_1(\delta_1)) + N(v - c_1(\delta_1)) \\ N(v + c_1(\delta_2)) - N(v - c_1(\delta_2)), & 1 - N(v + c_1(\delta_2)) + N(v - c_1(\delta_2)) \end{bmatrix}$$

Since there is no feedback from returns, we are able to derive an analytical pricing formula for European options. Let $C_0^k(X, n)$ be the time 0 price of an European call option with strike X that matures after n periods given that the initial volatility is δ_k (for $k = 1, 2$).

In the Appendix we show that:

$$C_0^k(X, n) = \sum_{j=0}^{n} \gamma_{nj}^k C_{j0}^k \tag{16}$$

where

$$C_{j0}^k = S_0 N(d_{1j}) - X e^{-r_f n} N(d_{2j}) \tag{17}$$

and

$$
\begin{aligned}
d_{1j} &= \frac{\ln S_0/X + nr + \theta_j^2/2}{\theta_j} \\
d_{2j} &= d_{1j} - \theta_j \\
\theta_j^2 &= j\delta_1^2 + (n-j)\delta_2^2 \quad \text{for } j = 0, \ldots, n
\end{aligned}
$$

In the above expression, γ_{nj}^k represents the probability (under measure Q) that in n periods the number of visits (from state k) to state 1 and state 2 are j and $n-j$, respectively. The formula for γ_{nj}^k is given in the Appendix.

The form of our analytical pricing formula suggests that the option price is a weighted average of the Black-Scholes formula values corresponding to different cumulative volatilities with the weights determined by the probabilities of different cumulative volatilities. This form resembles in spirit to the stochastic volatility option pricing result of Hull and White (1987). The driving force behind our uni-directional regime switching model and that of Hull and White (1987) is the same, which is due to the lack of return feedback.

For the purpose of pricing options, the uni-directional regime switching model can be reparameterized by the two volatility levels and the two diagonal transition probabilities, say p_{11} and p_{22}. Table 1 illustrates the typical sensitivity of option prices to these two transition probability variables controlling how volatilities are clustered.

[Insert Table 1 Here]

In this example, the initial volatility is always fixed at the low variance state. If the transition probability $p_{11} = 1$, the volatility remains in the low variance state and the option prices must always equal the Black Scholes price using the low volatility. This result is reflected in the last row of the table in which all entries are equal and the option value is the lowest. For the case of $p_{22} = 1$, the volatility stays in the low variance state for a random amount of time before being absorbed into a high variance state. The smaller is the value of p_{11}, the higher is the transition probability to the high variance state. Since the high variance state is absorbing, a lower p_{11} imply a higher option value, which is exactly the case shown in the last column of the table.

4 GARCH and Stochastic Volatility Option Pricing Models

As discussed earlier, it makes sense to consider more structured regime switching models that lead to fewer parameters. So far we have not imposed any special form on the functions $\{c_i(\sigma_t); i = 1, \cdots, K-1\}$. In this section we devise a specific regime switching model that leads to option models with between 3 and 8 parameters. The interesting feature of these models is that they can yield GARCH and stochastic volatility option pricing models as special limiting cases. To accomplish this objective, we only need to show the GARCH option pricing model as a special case of the regime switching with feedback model. The result for the stochastic volatility option pricing model follows directly from the result established in Duan (1996&1997) in which the GARCH option pricing model is shown to converge to the stochastic volatility option pricing model as the time step becomes infinitesimal.

Assume K volatility states, denoted by a strictly increasing sequence, $\{\delta_i(K), i = 1, \cdots, K\}$ satisfy the following partition condition.

Partition Condition.

1. $\delta_1(K) \to 0$ and, $\delta_K(K) \to \infty$ as $K \to \infty$, and
2. $\sup_{i\in\{1,2,\cdots K-1\}}[\delta_{i+1}^2(K) - \delta_i^2(K)] \to 0$ as $K \to \infty$.

The first partition condition requires the minimum (maximum) volatility to approach zero (infinity) when the number of volatility states goes to infinity. The second partition condition, on the other hand, ensures that two adjacent volatilities become arbitrarily close uniformly when the number of volatilities goes to infinity. In other words, the partition condition ensures that the rate of getting closer for any two adjacent volatilities is faster than the rate for the maximum volatility to go to infinity. We will provide later in this section a specific construction that satisfies the partition condition.

The bi-directional regime switching model converges to a limiting model that contains some standard GARCH(1,1) models as special cases.

Proposition 2 *Consider the bi-directional regime switching model based on K volatility states:*

$$\ln \frac{S_{t+1}^{(K)}}{S_t^{(K)}} = r + \lambda\sigma_{t+1}^{(K)} - \frac{1}{2}\sigma_{t+1}^{(K)2} + \sigma_{t+1}^{(K)}\varepsilon_{t+1} \tag{18}$$

$$\sigma_{t+1}^{(K)} = \delta_i(K) \quad if \quad c_{i-1}(\sigma_t^{(K)}) \leq q_1(\varepsilon_t - \omega)^+ + q_2(\varepsilon_t - \omega)^- + (1 - q_1 - q_2)|\xi_t| < c_i(\sigma_t^{(K)}) \tag{19}$$

for $i = 1, 2, \cdots, K$ and $t \in \{0, 1, \cdots, T-1\}$; $0 \leq q_1, 0 \leq q_2, q_1 + q_2 \leq 1$, and

$$\begin{bmatrix} \varepsilon_{t+1} \\ \xi_{t+1} \end{bmatrix} |\mathcal{F}_t \overset{P}{\sim} N(\mathbf{0}_{2\times 1}, \mathbf{I}_{2\times 2}) \tag{20}$$

The threshold values are set up as:

$$c_i(\sigma_t^{(K)}) = \sqrt{\max(\frac{\frac{1}{2}[\delta_i^2(K) + \delta_{i+1}^2(K)] - \beta_0}{\beta_2\sigma_t^{(K)2}} - \frac{\beta_1}{\beta_2}, 0)} \; for \quad i = 1, \cdots, K-1 \tag{21}$$

with $c_0(\sigma_t^{(K)}) = 0$ *and* $c_K(\sigma_t^{(K)}) = \infty$*. If the volatility states satisfy the partition condition,* $S_0^{(K)} = S_0$ *and* $\sigma_1^{(K)} = \sigma_1$*, then, as* $K \to \infty$*, the bi-directional regime switching model converges almost surely in* P *over* $[0, T]$ *to the following system:*

$$\ln \frac{S_{t+1}}{S_t} = r + \lambda\sigma_{t+1} - \frac{1}{2}\sigma_{t+1}^2 + \sigma_{t+1}\varepsilon_{t+1} \quad (22)$$

$$\sigma_{t+1}^2 = \beta_0 + \beta_1\sigma_t^2 + \beta_2\sigma_t^2[q_1(\varepsilon_t - \omega)^+ + q_2(\varepsilon_t - \omega)^- + (1 - q_1 - q_2)|\xi_t|]^2. \quad (23)$$

Proof: See Appendix

If we set $q_1 + q_2 = 1$, i.e., the effect due to the orthogonal volatility innovation is removed, the limiting volatility dynamic is simplified to

$$\begin{aligned}
\sigma_{t+1}^2 &= \beta_0 + \beta_1\sigma_t^2 + \beta_2\sigma_t^2[q_1(\varepsilon_t - \omega)^+ + q_2(\varepsilon_t - \omega)^-]^2 \\
&= \beta_0 + \beta_1\sigma_t^2 + \beta_2 q_1^2\sigma_t^2\left[(\varepsilon_t - \omega)^+\right]^2 + \beta_2 q_2^2\sigma_t^2\left[(\varepsilon_t - \omega)^-\right]^2 \qquad \text{(because } (\varepsilon_t - \omega)^+(\varepsilon_t - \omega)^- = 0\text{)} \\
&= \beta_0 + \beta_1\sigma_t^2 + \beta_2 q_1^2\sigma_t^2\left[(\varepsilon_t - \omega) + (\varepsilon_t - \omega)^-\right]^2 + \beta_2 q_2^2\sigma_t^2\left[(\varepsilon_t - \omega)^-\right]^2 \\
&= \beta_0 + \beta_1\sigma_t^2 + \beta_2 q_1^2\sigma_t^2\left\{(\varepsilon_t - \omega)^2 - \left[(\varepsilon_t - \omega)^-\right]^2\right\} + \beta_2 q_2^2\sigma_t^2\left[(\varepsilon_t - \omega)^-\right]^2 \\
&= \beta_0 + \beta_1\sigma_t^2 + \beta_2 q_1^2\sigma_t^2(\varepsilon_t - \omega)^2 + \beta_2(q_2^2 - q_1^2)\sigma_t^2\left[(\varepsilon_t - \omega)^-\right]^2. \qquad (24)
\end{aligned}$$

The limiting model becomes the NGARCH(1,1) model of Engle and Ng (1993) if $q_1 = q_2$. If $\omega = 0$, the limiting model reduces to the GJR-GARCH(1,1) specification of Glosten, *et al.* (1993).

By the same argument, the volatility dynamic under the equilibrium price measure Q has the following limit:

$$\sigma_{t+1}^2 = \beta_0 + \beta_1\sigma_t^2 + \beta_2\sigma_t^2[q_1(\varepsilon_t^* - \omega - \lambda)^+ + q_2(\varepsilon_t^* - \omega - \lambda)^- + (1 - q_1 - q_2)|\xi_t^* - v|]^2. \quad (25)$$

Again, if $q_1 + q_2 = 1$, we have

$$\sigma_{t+1}^2 = \beta_0 + \beta_1\sigma_t^2 + \beta_2 q_1^2\sigma_t^2(\varepsilon_t^* - \omega - \lambda)^2 + \beta_2(q_2^2 - q_1^2)\sigma_t^2\left[(\varepsilon_t^* - \omega - \lambda)^-\right]^2, \quad (26)$$

which yields Duan's (1995) GARCH option pricing model for the NGARCH or GJR-GARCH specification depending on how the parameters are specialized.

To gain a better understanding of the properties of this limiting model and its implications for option pricing, we plot the risk-neutral density function under various parameter scenarios. Different combinations of parameter values will inevitably affect the level of volatility and make the comparison less meaningful. It is therefore preferable to control the level of volatility and focus on the impact of a parameter change on skewness and kurtosis of the risk-neutral density. In order to do so, we need to derive an explicit expression for the stationary volatility based on the system in equation (25). The following proposition provides such a result.

Proposition 3 *If the conditional volatility under measure Q has the following dynamic:*

$$\sigma_{t+1}^2 = \beta_0 + \beta_1 \sigma_t^2 + \beta_2 \sigma_t^2 [q_1 (\varepsilon_t^* - \omega - \lambda)^+ + q_2 (\varepsilon_t^* - \omega - \lambda)^- + (1 - q_1 - q_2) |\xi_t^* - v|]^2, \qquad (27)$$

and $\rho < 1$, then for $t \geq 1$,

$$E^Q \left(\sigma_t^2 | \mathcal{F}_0\right) = \sigma_1^2 \rho^{t-1} + \frac{\beta_0 (1 - \rho^{t-1})}{1 - \rho} \qquad (28)$$

and

$$E^Q \left(\sigma_t^2\right) = \frac{\beta_0}{1 - \rho}, \qquad (29)$$

where

$$\rho = \beta_1 + \beta_2 \left\{ \begin{array}{c} q_1^2 E^Q \left(\left[(\varepsilon_t^* - \omega - \lambda)^+ \right]^2 | \mathcal{F}_{t-1} \right) + q_2^2 E^Q \left(\left[(\varepsilon_t^* - \omega - \lambda)^- \right]^2 | \mathcal{F}_{t-1} \right) + \\ (1 - q_1 - q_2)^2 E^Q \left(|\xi_t^* - v|^2 | \mathcal{F}_{t-1} \right) + \\ 2 q_1 (1 - q_1 - q_2) E^Q \left((\varepsilon_t^* - \omega - \lambda)^+ | \mathcal{F}_{t-1} \right) E^Q \left(|\xi_t^* - v| | \mathcal{F}_{t-1} \right) + \\ 2 q_2 (1 - q_1 - q_2) E^Q \left((\varepsilon_t^* - \omega - \lambda)^- | \mathcal{F}_{t-1} \right) E^Q \left(|\xi_t^* - v| | \mathcal{F}_{t-1} \right) \end{array} \right\}$$

$$\begin{aligned} E^Q \left(\left[(\varepsilon_t^* - \omega - \lambda)^+ \right]^2 | \mathcal{F}_{t-1} \right) &= \left[1 + (\omega + \lambda)^2 \right] \left[1 - N(\omega + \lambda) \right] - (\omega + \lambda) N'(\omega + \lambda) \\ E^Q \left(\left[(\varepsilon_t^* - \omega - \lambda)^- \right]^2 | \mathcal{F}_{t-1} \right) &= \left[1 + (\omega + \lambda)^2 \right] N(\omega + \lambda) + (\omega + \lambda) N'(\omega + \lambda) \\ E^Q \left(|\xi_t^* - v|^2 | \mathcal{F}_{t-1} \right) &= 1 + v^2 \\ E^Q \left((\varepsilon_t^* - \omega - \lambda)^+ | \mathcal{F}_{t-1} \right) &= N'(\omega + \lambda) - (\omega + \lambda) \left[1 - N(\omega + \lambda) \right] \\ E^Q \left((\varepsilon_t^* - \omega - \lambda)^- | \mathcal{F}_{t-1} \right) &= N'(\omega + \lambda) + (\omega + \lambda) N(\omega + \lambda) \\ E^Q \left(|\xi_t^* - v| | \mathcal{F}_{t-1} \right) &= 2 N'(v) + 2 v N(v) - v \end{aligned}$$

Remark: $N(\cdot)$ and $N'(\cdot)$ are the standard normal distribution and density functions, respectively. σ_1 is measurable with respect to $\mathcal{F}_0$.

Proof: See Appendix

It is evident from the above proposition, if β_0 and ρ are fixed, then the stationary volatility of the system under measure Q in equation (29) is fixed. Suppose that the values for q_1 and q_2 are changed. One can always make a compensatory adjustment to β_2 so that ρ remains unchanged. Similarly, if the value of $\omega + \lambda$ is changed, β_2 can be adjusted to maintain the level of ρ. Making this type of change allows us to examine the effect of the orthogonal volatility innovation (ξ_t^*) and the asset risk premium (λ) on the risk-neutral density without altering the magnitude of the volatility. Figure 1 illustrates several risk-neutral density functions obtained for the stock price over a three-month horizon. The parameters that are fixed in this analysis are $\beta_0 = 0.00001, \beta_1 = 0.8, \omega = 0$ and $v = 0$. Five different scenarios are considered: (1) $q_1 = q_2 = \frac{1}{2}, \lambda = 0$, (2) $q_1 = q_2 = \frac{1}{2}, \lambda = 1$, (3) $q_1 = \frac{1}{3}, q_2 = \frac{2}{3}, \lambda = 0$, (4) $q_1 = 0, q_2 = 0, \lambda = 0$, and (5) $q_1 = \frac{2}{3}, q_2 = \frac{1}{3}, \lambda = 0$. In all five cases, the stationary risk-neutral standard deviation (annualized) equals 20% because ρ is maintained at 0.90875.

[Insert Figure 1 Here]

The figure illustrates that the regime switching model is capable of generating return distributions that are skewed and have fat tails relative to the normal.

Proposition 2 allows a great deal of freedom in choosing $\{\delta_i(K); i = 1, \cdots, K\}$. In the actual construction of a regime switching model, one hopes to rapidly approach the limiting GARCH model under either measure P or Q by employing some natural designs. Intuitively, an efficient design will have the conditional volatility centered around the stationary volatility of the target system. Indeed, our design relies on the stationary volatility. For option pricing, we are mainly interested in the system under measure Q, and therefore the stationary volatility of interest is the one corresponding to measure Q. We propose the following simple discretization scheme: for any time horizon $T < \infty$, and $K > 1$,

$$\delta_i^2(K) = L(K) + (i-1)\frac{U(K) - L(K)}{K-1} \tag{30}$$

where

$$\begin{aligned} L(K) &= \max(c_{Q,T} - b_{Q,T}\sqrt{K-1}, 0) \\ U(K) &= c_{Q,T} + b_{Q,T}\sqrt{K-1} \\ c_{Q,T} &= E^Q(\sigma_T^2|\mathcal{F}_0) \\ b_{Q,T} &= \frac{c_{Q,T}}{l}, \text{ and } l \text{ is some positive integer.} \end{aligned}$$

Note that the formula for $E^Q(\sigma_T^2|\mathcal{F}_0)$ has been derived in Proposition 3. The choice of l depends on how many volatility states and the maximum volatility are intended in the approximation. For example, $K = 101$ and $l = 5$ give rise to $U(101) = 3c_{Q,T}$ and $L(101) = 0$. That is, the conditional volatility is allowed to move between zero and three times of its stationary level. If a bigger volatility range is desired, l can be decreased without changing K. For example, $l = 4$ gives rise to $U(101) = 3.5c_{Q,T}$ and $L(101) = 0$. It is straightforward to verify that the partition condition is satisfied regardless of the choice of l. In short, we have a specific bi-directional regime switching model design that approximates the limiting process under measure Q over the time horizon $[0, T]$.

5 Pricing American Options in the bi-directional Model

We now develop lattice based models for pricing American contracts under the bi-directional regime switching model. Consider an option that expires after n periods. We shall assume that there are K regimes corresponding to K distinct variances. Without loss of generality we order the K volatilities with $\delta_1 < \delta_2 < .... < \delta_K$. The threshold values, $\{c_0(\delta_i), c_1(\delta_i), \cdots, c_K(\delta_i)\}$ for each volatility level are assumed to be given.

Viewed from date t, and conditional on the volatility over the next time increment being δ_k, the logarithmic return at date $t+1$ is normal under measure Q with mean, $(r - \delta_k^2/2)$ and variance δ_k^2. Let (y_t, σ_{t+1}) represent the logarithm of the stock price at date t and the volatility for the period $[t, t+1]$ respectively.

We now establish a discrete Markov chain approximation for the dynamics of $\{(y_t, \sigma_{t+1}) \mid t = 0, 1, 2, \cdots, n\}$ that converges to the bi-directional regime switching model. We begin by fixing the topology of points for the logarithm of the stock price over time. Let $y_{t+1}^{a(m)}$ be a discrete random variable approximating y_{t+1}, for $t = 0, 1, 2, \cdots, n-1$ with $y_0^{a(m)} = y_0 = \ln S_0$. The approximating process captures the conditional transitions by a variable that can take on one of $2m+1$ points, with m–up jumps, m–down jumps and a horizontal move. In particular, given $(y_t^{a(m)}, \sigma_{t+1} = \delta_k)$, the conditional normal distribution is approximated by the random variable, that takes on the following values:

$$(y_{t+1}^{a(m)} \mid \sigma_{t+1} = \delta_k) = y_t^{a(m)} + j(\eta_k \gamma)$$

where $j = 0, \pm 1, \pm 2, ..., \pm m$, and η_k is the smallest positive integer that ensures that the conditional first two moments of the approximating distribution match the moments of the true distribution. The conditional probability distribution for $(y_{t+1}^{a(m)} | \sigma_{t+1} = \delta_k)$ is defined by:

$$\Pr\{y_{t+1}^{a(m)} = y_t^{a(m)} + j(\eta_k \gamma) \mid \delta_k\} = P(j;k) \qquad j = 0, \pm 1, \pm 2, ..., \pm m$$

where

$$P(j;k) = \sum_{k_1, k_2, k_3} \begin{pmatrix} n \\ k_1 \; k_2 \; k_3 \end{pmatrix} p_u^{k_1} \, p_m^{k_2} \, p_d^{k_3}$$

with $j = k_1 - k_3$, $\; k_1 + k_2 + k_3 = m$, and $k_1, \; k_2, \; k_3 \geq 0$. Further

$$\begin{aligned} p_u &= \frac{\delta_k^2}{2\eta_k^2\gamma^2} + \frac{(r - \delta_k^2/2)\sqrt{\Delta t/m)}}{2\eta_k\gamma} \\ p_m &= 1 - \frac{\delta_k^2}{\eta_k^2\gamma^2} \\ p_d &= \frac{\delta_k^2}{2\eta_k^2\gamma^2} - \frac{(r - \delta_k^2/2)\sqrt{\Delta t/m}}{2\eta_k\gamma} \end{aligned}$$

It is straightforward to verify that this approximation ensures that the approximating conditional distribution of the discrete random variable $(y_{t+1}^{a(m)} \mid \sigma_{t+1} = \delta_k)$ has mean $r - \delta_k^2/2$ and variance δ_k^2, and that the discrete random variable $(y_{t+1}^{a(m)} \mid \sigma_t = \delta_k)$, converges in distribution to a normal distribution as $m \to \infty$.

To make matters specific, let (i, j) represent time period i where the stock price is at "level j", with node $(0, 0)$ representing date 0 when the logarithm of the stock price is y_0, and the levels of logarithmic prices in future periods are separated by γ. The lowest level that the approximating logarithmic stock price can take in period i is $y_{min}(i)$ and the maximum is $y_{max}(i)$ where

$$y_{min}^{a(m)}(i) \quad = \quad y^{a(m)}(0,0) - im\gamma\eta_k$$

$$y^{a(m)}_{max}(i) \quad = \quad y^{a(m)}(0,0) + im\gamma\eta_k$$

and

$$y^{a(m)}(i,j) = y_0 + j\gamma \text{ for } j = 0 \pm 1 \pm 2 \pm im\eta_k$$

Of course, knowledge of node (i,j) is not sufficient to uniquely characterize the option price. At each node, we shall carry a vector of K prices. The k^{th} entry of this vector at this node represents the price of an option given that the stock enters this time period with the state variables set at $(y^{a(m)}(i,j), \delta_k)$. Let $C^{a(m)}(i,j,k)$ represent the option price at node (i,j) when the variance for the next period is δ_k.

Option prices can now readily be computed using standard backward recursion procedures. The exact procedure used is a modification of the methods used by Ritchken and Trevor (1999) who price options under GARCH processes. Unlike their model, our model does not require a forward scan to identify the extreme volatility levels at each stock price node in the lattice. We begin in the last period, n, and note that the payout of the claim is fully determined by the stock price alone. Hence, each K-vector, at each terminal node, will consist of K equal entries. We now can apply backward recursion to establish the option price at date 0. In general, consider a node in period i, ($0 \cdot \quad i < n$) say at node (i,j), and assume $C^{a(m)}(i,j,k)$ is to be computed. Given the volatility for the time increment $[i, i+1]$ is δ_k, we first define the normalized innovations as

$$\varepsilon(\ell;k) = \frac{\ell\gamma\eta_k - (r - \delta_k^2/2)\Delta t}{\delta_k\sqrt{\Delta t}}$$

As $m \to \infty$, the normalized residual converges to a drawing from a standard normal distribution.

Given node (i,j), the volatility level, k, and the innovation term, ℓ, the jump magnitude is $\ell\gamma\eta_k$, and the new node is $(i+1, j+\ell\eta_k)$. The stock price is determined, but the volatility regime is still not identified, unless $q_1 + q_2 = 1$.

The conditional probability distribution for the new regime is partially influenced by the asset innovation, $\varepsilon(\ell;k)$ and the orthogonal innovation, ξ. Let $\pi_{k,p}(\ell)$ represent the probability of switching to level p from level k as a function of the asset jump innovation, determined by ℓ, where $\ell = -m, ..., m$.

If the volatility switches to regime p, then the exact option price at the successor node can be recovered. The expected call price at time $i+1$ viewed from node (i,j) when the volatility is δ_k is:

$$E^Q\{C^{a(m)}_{i+1} \mid y^{a(m)}(i,j), \delta_k\} = \sum_{p=1}^{K}\sum_{\ell=-m}^{m} C^{a(m)}(i+1, j+\eta_k\ell, p)P(\ell;k)\pi_{k,p}(\ell)$$

Finally, the option value in regime k, at this node is computed as:

$$C^{a(m)}(i,j,k) = e^{-r_f\Delta t}E^Q\{C^{a(m)}_{i+1} \mid y^{a(m)}(i,j), \delta_k\}$$

The above value is the unexercised value of the claim at the node. For an American call option, this value has to be compared to the intrinsic value which is given by:

$$C^{a(m)}_{stop}(i,j,k) = Max(e^{y^{a(m)}(i,j)} - X, 0)$$

The final option price, conditional on the volatility being δ_k, obtained using backward recursion, is given by $C^{a(m)}(0,0,k)$.

Table 2 demonstrates the typical convergence rate of the algorithm for the case where analytical solutions are available. Specifically, the table shows the convergence rate of the algorithm when the uni-directional model is used. As m increases, the discrete approximation to the normal distribution over each trading period improves. As can be seen, the algorithm produces good prices for small values of m. For this case, and for the bi-directional case, $m = 5$ appears more than satisfactory.

[Insert Table 2 Here]

Table 3 shows the rate of convergence of the regime switching model prices to NGARCH option prices. The contract priced is a 50 day European call option with the stock price at 100. All prices are computed using the lattice approximation described in Section 5. The parameters used are $\beta_0 = 0.000006575, \beta_1 = 0.9, \omega = 0.0$. The initial and stationary volatility are at 20%. Different regime levels are chosen according to (30). Particularly, we used $l = 10$, which results in volatility levels between 0.1% and 40% for the 100 regime model and between 0% and 90% for the 1000 regime model. The reported prices correspond to a volatility level artificially fixed at 20%. The last two rows show the 95% confidence intervals for the limiting NGARCH option prices. Note that the prices start to converge around 400 regimes.

[Insert Table 3 Here]

When 400 regimes are used, the regime switching model produces option prices that are consistently in the GARCH confidence intervals. Since the algorithm we use requires substantially less computations than that of Ritchken and Trevor (1999) and is much simpler, it offers a good alternative for pricing American options under GARCH.

6 Empirical Analysis of the Regime Switching Models

In this section we investigate the pricing of the $S\&P500$ stock index options (European calls) and identify the benefits, if any, in advancing beyond the Black-Scholes model to the uni-directional and bi-directional regime switching models. We address the in-sample biases that result from using these models, and the out-of-sample performance when these models are used to predict prices in the future.

The $S\&P500$ index options are European call options that exist with maturities in the next two calendar months, and also for the time periods corresponding to the expiration dates of the futures. Our price data on the options covering the five year period from January 1991 to December 1995 was obtained from the Berkeley Option Database. We collected data on Wednesdays and excluded contracts with maturities fewer than six days. We only used options with bid/ask price quotes during the last half hour of trading. For these contracts we also captured the reported concurrent stock index level associated with each option trade.

In order to price the options we need to adjust the index level according to the dividends paid out over the time to expiration. We follow Harvey and Whaley (1992), Jackwerth and Rubinstein (1996) and Bakshi, Cao and Chen (1997), and use the actual cash dividend payments made during the life of the option to proxy for the expected dividend payments. The present value of all the dividends is then subtracted from the reported index levels to obtain the contemporaneous adjusted index levels. This procedure assumes that the reported index level is not stale and reflects the actual price of the basket of stocks representing the index. Since intraday data and not the end of the day option prices are used, the problem with the index level being stale is not severe.[1] Since we used the actual contemporaneous index level associated with each option trade that was reported in the data base, the actual adjusted index level would vary slightly among the individual contracts depending on their time of trade. We normalize all option and strike prices so that the adjusted index price is exactly $1. Finally, we used the T-Bill term structure to extract the appropriate discount rates.

We are interested in evaluating the relative performance of four models. The first model, BS, is the standard Black Scholes model; the second model, Ad-Hoc, is the ad hoc procedure used by Dumas, Fleming and Whaley (1998), that merely smooths Black Scholes implied volatilities across strike prices and expiry dates. Dumas, Fleming and Whaley found that this model performed as well as more sophisticated models where volatility was allowed to be a deterministic function of asset price and time. This model provides a useful practical benchmark for evaluating model performance of European contracts.

The ad-hoc BS model works as follows. Given a collection of n option prices observed at a date, compute their implied Black-Scholes volatilities. Then, run a linear regression model, that is quadratic in maturity and strike. Specifically, we have:

$$IV_i = \beta_0 + \beta_1 T_i + \beta_2 T_i^2 + \beta_3 X_i + \beta_4 X_i^2 + \beta_5 X_i T_i + \epsilon_i$$

where IV_i is the implied volatility of the i^{th} contract that has strike price X_i (divided by index value) and maturity T_i, $i = 1, 2...n$, and ϵ_i represents the random error term, with mean 0 and constant variance. The regression model is then used to estimate the implied volatility for each contract. The resulting n implied volatilities are used in the Black Scholes equation to generate theoretical prices, referred to as Ad-Hoc prices.

[1]There are other methods for establishing the adjusted index level. The first is to compute the mid points of call and put options with the same strikes and then to use put-call parity to imply out the value of the underlying index. Of course, this method has its own problems, since with non negligible bid ask spreads, put call parity only holds as an inequality. An alternative approach is to use the stock index futures price to back out the implied dividend adjusted index level. This leads to one stock index adjusted value that is used for all option contracts. For a discussion of these approaches see Jackwerth and Rubinstein (1996).

We also included two specific regime switching models; a bi-directional regime switching model, BRS, obtained using the locally risk-neutralized version of equations (22) and (23); that is, equations (7) and (25). For the bi-directional model, we specifically set $q_1 = q_2 = q$. For its uni-directional special case, we impose $q_1 = q_2 = 0$, which has volatility feedback switched off.

We use all our option contracts at each date, to imply out the parameters of the five models. The theoretical prices for the regime switching models are actually based on a continuum of regimes, and were computed using large sample Monte Carlo simulations with control variates. Specifically, the prices were obtained using equations (7) and (25). The empirical martingale simulation method of Duan and Simonato (1998) was implemented, and 25,000 sample paths were used for each contract. Common random numbers were used in each pricing calculation. The criteria adopted was to minimize the sum of squared errors between the actual and theoretical prices. This criterion has been employed in several studies.[2] For the BS model, there is only one unknown volatility parameter, so one would expect this model to produce the worst in-sample performance. In contrast, the *ad hoc* procedure incorporates significant information on all prices, and one would expect its in-sample performance to be far superior. The bi-directional regime switching model has six unknowns (β_0, β_1, β_2 ω, q and the initial volatility, σ_1), whereas the uni-directional model only has four unknowns (β_0, β_1, β_2 and the initial volatility σ_1).

We perform this fitting operation every 5 weeks over the time period from 1991 to 1995, yielding 52 data sets. After fitting the parameters, we also investigate the one week, two week, and four week out of sample performance.

Table 4 shows the number of contracts analyzed in each moneyness and maturity category in the in-sample fitting periods and in the out-of-sample periods.

[Insert Table 4 Here]

As expected, the models with the most parameters perform the best in fitting data in-sample. In particular, the Black Scholes and uni-directional models produced larger errors on average than the Ad-Hoc and bi-directional models. For each of the four models, Table 5 shows a box and whiskers plot of the in-sample pricing residuals categorized by moneyness and maturity.

[Insert Table 5 Here]

The first column of the table clearly reveals the volatility smile effect associated with the Black-Scholes model. With just one free parameter, the BS model is not capable of fitting prices of options. With four free parameters, the uni-directional model does not remove much of the bias and significant patterns in the residuals exist over strike prices, for all maturity buckets. The Ad-Hoc model, with its 6 parameters,

[2] See for example, Bakshi, Cao and Chen (1997) and Dumas, Fleming and Whaley (1998).

seems to remove most of this bias, especially for the mid-term contracts. The box and whiskers plots also reveal that the bi-directional model, like the Ad-Hoc model, removes much of the strike price biases. The results indicate that the incorporation of feedback effects can result in a substantial improvement in the fitting of a regime switching model.

Our results here echo the conclusions of Dumas, Fleming and Whaley (1998), namely that in-sample fits typically improve as the number of free parameters increase. The test of any model, in terms of pricing capability, is better revealed in out-of-sample comparisons, however.

Towards this goal we look at the relative performance for the four models one week out-of-sample. For the Black Scholes model, conditional on the dividend adjusted S&P 500 one week later, we reestimate the single volatility, that minimizes the sum of squared errors, and use that value to reprice all option contracts and to establish pricing residuals.[3]

For the uni-directional and bi-directional models we use the parameters estimated in the previous week for β_0, β_1 and β_2 (and q and ω). To estimate the theoretical option prices, the values of both state variables, namely the adjusted index level, and the local volatility, are required. The adjusted index value is of course known, and the local volatility is implied out from the data. To make a fair comparison with the Ad-Hoc model, we keep all the regression coefficients fixed except for the intercept which is reestimated from the data. In this way, in the out-of-sample periods, all models have one degree of freedom.

The box and whiskers plots for the out-of-sample fits one week after the in-sample fitting has taken place are summarized in Table 6a.

[Insert Table 6a Here]

The plots of the pricing residuals reveal that the properties in the in sample fitting period continue to hold in the out-of-sample periods. The uni-directional model prices, like the Black and Scholes prices, are not capable of explaining market prices. The Ad-Hoc model produces pricing residuals with patterns similar to the in-sample fitting period. In particular, there are few biases, and the mid term contracts have low standard errors relative to the unadjusted Black Scholes model. This indicates that a quadratic function of strike and maturity is somewhat effective in removing biases one week out of sample. The bi-directional model continues to perform well relative to the uni-directional model. The box and whiskers plots indicate that the bi-directional model is about the same as the Ad-Hoc model.

Table 6b repeats the above analysis two weeks after the in sample fitting was done. As the time horizon increases, the standard error of the pricing residuals for each maturity and strike category expand.

[3] In other words, for the BS model there is no out-of-sample tests. We perform this so that each of our models in the out-of-sample periods has one degree of freedom.

However, the overall patterns in the residuals remain.[4]

[Insert Table 6b Here]

The above results provide graphical evidence that incorporating feedback effects in a regime switching model leads to a substantial removal of biases that exist when using a uni-directional model. Further, the results indicate that, for European options, the Ad-Hoc model provides a substantial improvement over the unadjusted Black Scholes model. While the Ad-Hoc procedure might be competitive with the bi-directional model, the former model is very restrictive in that it can only be applied to European options. The big advantage of the bi-directional model is its reliance on a parametric structure to describe the evolution of asset prices. Calibrating against the data set is thus not a pure curve fitting exercise. Once the parameters are estimated, the model can be used to price American claims and other exotic options.

Since our main objective is to measure the benefits of the feedback effects, we computed the root mean square errors of the one week out-of-sample pricing residuals (by moneyness and maturity) for the two regime switching models. In 14 out of the 15 cases, the standard errors are smaller for the bi-directional model. Table 7 provides a summary statistic for comparing the two regime switching models in the out-of-sample period. Specifically, for each maturity-moneyness bucket, the square root of the ratio of the out-of-sample sum of squared errors for the bi-directional model relative to the uni-directional model was computed. The logarithm of this statistic serves as a measure of relative performance for the models. A negative statistic indicates that the bi-directional model is better in average than the uni-directional model. The table illustrates that in 14 out of the 15 categories the bi-directional model performed better than the uni-directional model. Only for long term in the money options, did the uni-directional model perform better, and in this case the improvement was very marginal.

[Insert Table 7 Here]

Table 8 shows the mean value of each of the in-sample parameter estimates, together with their standard deviations for all the models over the 52 data sets.

[Insert Table 8 Here]

As can be seen some of the estimates fluctuate quite a bit. This either reflects the fact that the objective function is fairly flat over a wide region or it may indicate that the model is not well specified. For the bi-directional model in all 52 optimizations the parameter that controls the feedback effect, namely q was found to be significant. Its mean value was 0.40, Our analysis suggests that the bi-directional model adds significantly to the uni-directional model.

[4]The same patterns continue to hold four weeks later. These results are available from the authors.

In summary, our empirical study has confirmed the importance of modeling the feedback effect. A bi-directional model is an improvement over a uni-directional model. Further, the bi-directional model is capable of explaining most of the volatility smile. While the Ad-Hoc model also can explain a significant fraction of the Black-Scholes bias, it is difficult to adapt that model in pricing American claims or any exotic instrument.

7 Conclusion

This article has developed a family of option pricing models that permit volatilities to follow a regime switching process. If the transition matrix that characterizes the regime switching process is independent of the return innovations, then the model reduces to the uni-directional regime switching model pioneered by Hamilton (1989). We can, however, obtain a richer structure of models by permitting the transition probabilities to be influenced by the return innovations. This leads to the bi-directional model for asset returns. In this paper, we considered a family of such models and identified an equilibrium pricing measure which permits option prices to be determined. A pricing algorithm was presented that permits American options to be priced. For the special case of a uni-directional process, a simple analytical solution exists for the pricing of European call options.

Option prices can be generated under a very rich family of bi-directional regime switching processes. Unfortunately, this comes at a high cost of requiring many parameters to estimate. Fortunately, it is possible to parameterize the regime switching mechanism in such a way so as to ensure that GARCH option models, as developed by Duan (1995), appear as special limiting solutions. Actually, we showed how generalized GARCH option pricing models could be established. These models are based on GARCH processes with the additional feature that volatility updates are also influenced by a second factor that is orthogonal to the return innovation. The number of parameters for these models do not depend on the number of regimes. For option pricing purposes, the number of unknown parameters range from 3 to 7.

Reducing the number of regimes to a small collection is advantageous since computations grow more than linearly with the number of regimes. Numerical simulations reveal that a relatively small number of regimes might suffice to capture the complex dynamics of volatilities. Moreover, the extended GARCH models allows the modeler to produce a very rich array of return distributions beyond GARCH models.

We have provided empirical evidence that suggests that bi-directional regime switching models provide significant improvements beyond the Black Scholes and the uni-directional models that do not permit feedback effects. We have also seen that the bi-directional model performs comparably to the Ad-Hoc model. While the Ad-Hoc model is capable of removing much of the bias, the use of this heuristic method is at best limited to European contracts. Our models permit American options and exotics to be priced.

It remains for future empirical studies to evaluate the contribution of models within our family of bi-directional regime switching models. At worst, the algorithms here can be used to implement GARCH option pricing, for which there appears to be considerable empirical support. At best, perhaps there are some bi-directional regime switching models with $0 < q < 1$ that only require a few regimes, that outperform GARCH models, are computationally more efficient, and contain no additional parameters. Identifying these models and closely investigating pricing and hedging effectiveness issues, along the lines of Bakshi, Cao and Chen (1997) is worthwhile. Indeed, an attractive feature of the models presented is that since the dynamics under the data generating measure are prescribed, the consequences of discrete time hedging can be assessed. Finally, it remains for future studies to combine cross sectional information from option prices with the time series behavior of asset prices in a study that leads to the extraction of all the parameters that describe the pricing dynamics. Such a study should lead to a more thorough reconciliation of the time series estimates of volatility formation versus the implied estimates of volatility that are extracted from option prices alone.

Appendix

Proof of Proposition 1

By Theorem 1 of Duan (1995), the local risk-neutral valuation relationship holds under any of the three conditions given in the statement of the theorem. This means that (1) $\ln \frac{S_{t+1}}{S_t}$ is, conditional on $\mathcal{F}_t$, a Q-normal random variable, (2) $Var^Q(\ln \frac{S_{t+1}}{S_t}|\mathcal{F}_t) = Var^P(\ln \frac{S_{t+1}}{S_t}|\mathcal{F}_t)$, and (3) $E^Q(\ln \frac{S_{t+1}}{S_t}|\mathcal{F}_t) = e^r$. These three properties together imply that $\ln \frac{S_{t+1}}{S_t} = r - \frac{1}{2}\sigma_t^2 + \sigma_t \varepsilon^*_{t+1}$, which is equation (7). It remains to show that the conditional variance dynamics can be written as in equation (8) and $E^Q(\varepsilon^*_{t+1}\xi^*_{t+1}|\mathcal{F}_t) = 0$.

By the proof of Theorem 1 of Duan (1995; p. 27), the logarithmic marginal rate of substitution, denoted by m_{t+1}, for the period from t to $t+1$, is normally distributed with a constant mean and variance. This result can be used to compute the following quantity:

$$\begin{aligned}
E^Q(e^{c\xi_{t+1}}|\mathcal{F}_t) &= e^r E^P(e^{c\xi_{t+1}+m_{t+1}}|\mathcal{F}_t) \\
&= e^{\alpha_{t+1}} E^P(e^{(c-v_{t+1})\xi_{t+1}+U_{t+1}}|\mathcal{F}_t) \\
&= \exp\left(\alpha_{t+1} + \frac{1}{2}E^P(U^2_{t+1}|\mathcal{F}_t)\right) E^P(e^{(c-v_{t+1})\xi_{t+1}}|\mathcal{F}_t) \\
&= \exp\left(\alpha_{t+1} + \frac{1}{2}E^P(U^2_{t+1}|\mathcal{F}_t) + \frac{c^2}{2} + \frac{v^2_{t+1}}{2} - cv_{t+1}\right)
\end{aligned}$$

The first equality is due to the definition of marginal rate of substitution.. The second equality results from linearly projecting m_{t+1} onto ξ_{t+1}. Note that α_{t+1} and $v_{t+1}$are $\mathcal{F}_t$-measurable projection coefficients. The third equality is due to the independence of ξ_{t+1} and U_{t+1} and the moment generating function for normal random variables. The last equality is again due to the moment generating function for normal random variables. Let $c = 0$ and take advantage of the fact that $E^Q(1|\mathcal{F}_t) = 1$ to obtain

$$E^Q(e^{c\xi_{t+1}}|\mathcal{F}_t) = \exp(\frac{c^2}{2} - cv_{t+1})$$

Thus, ξ_{t+1} is distributed normally, conditional on $\mathcal{F}_t$ and with respect to measure Q, with mean $-v_{t+1}$ and variance 1. Define $\xi^*_{t+1} \equiv \xi_{t+1} + v_{t+1}$. The conditional volatility dynamic in equation (8) is thus established.

To prove $E^Q(\varepsilon^*_{t+1}\xi^*_{t+1}|\mathcal{F}_t) = 0$, consider

$$\begin{aligned}
E^Q(e^{c(\varepsilon_{t+1}+\xi_{t+1})}|\mathcal{F}_t) &= e^r E^P(e^{c(\varepsilon_{t+1}+\xi_{t+1})+m_{t+1}}|\mathcal{F}_t) \\
&= e^{a_{t+1}} E^P(e^{(c-b_{t+1})(\varepsilon_{t+1}+\xi_{t+1})+U_{t+1}}|\mathcal{F}_t) \\
&= \exp\left(a_{t+1} + \frac{1}{2}E^P(U^2_{t+1}|\mathcal{F}_t)\right) E^P(e^{(c-b_{t+1})(\varepsilon_{t+1}+\xi_{t+1})}|\mathcal{F}_t) \\
&= \exp\left(a_{t+1} + \frac{1}{2}E^P(U^2_{t+1}|\mathcal{F}_t) + c^2 + b^2_{t+1} - 2cb_{t+1}\right).
\end{aligned}$$

The above derivation is similar to the one used earlier to prove equation (8). Again, a_{t+1} and $b_{t+1}$are $\mathcal{F}_t$-measurable projection coefficients. The only difference is that $(\varepsilon_{t+1} + \xi_{t+1})$ has a variance equal to 2 (with respect to measure P). Again, let $c = 0$ and recognize that $E^Q(1|\mathcal{F}_t) = 1$ to yield

$$E^Q(e^{c(\varepsilon_{t+1}+\xi_{t+1})}|\mathcal{F}_t) = \exp(c^2 - 2cb_{t+1})$$

Thus, $(\varepsilon_{t+1} + \xi_{t+1})$ is a normal random variable with a variance equal to 2 under measure Q, which implies that $(\varepsilon^*_{t+1} + \xi^*_{t+1})$ also has a variance of 2 under measure Q. Since ε^*_{t+1} and ξ^*_{t+1} are standard normal random variables individually, it must be that ε^*_{t+1} and ξ^*_{t+1} are Q -independent. The proof is thus complete.

Derivation of the Two-State uni-directional Regime Switching Option Model

The total logarithmic asset return over the n periods, conditional on the state at time 0, is a mixture of $(n+1)$ normal variables determined by the number of times the volatility visits a state, say state 1. The reason that the distribution is a mixture of normals is because of the uni-directional model does not permit feedbacks. Use $\phi_j(y; \eta_j, \theta_j^2)$ to denote the normal density function with mean η_j and variance θ_j^2. Let N_n^k represents the number of visits to state 1 in n trials, given that at time 0, the state is k. Let the density function (under measure Q) of the total logarithmic return after n periods (starting from the volatility state k) be denoted by $f^n(y|\sigma_0 = \delta_k)$. It can be written as

$$f^n(y|\sigma_0 = \delta_k) = \sum_{j=0}^{n} \gamma_{nj}^k \phi_j(y; \eta_j, \theta_j^2) \quad \text{for } k = 1, 2$$

where

$$\begin{aligned}
\eta_j &= nr - \frac{1}{2}\theta_j^2 \\
\theta_j^2 &= j\delta_1^2 + (n-j)\delta_2^2 \\
\gamma_{nj}^k &= \Pr{}^Q(N_n^k = j) \quad \text{for } j = 0, 1, \cdots, n \text{ and } k = 1, 2
\end{aligned}$$

Since the European call option value is the discounted expected value of $\max(S_n - X, 0)$, it can be written as in equation (16), which is a weighted average of the Black-Scholes formula values corresponding to different cumulative volatilities.

To complete the derivation, we derive the expressions for γ_{nj}^k. In particular, we shall develop the equations for γ_{nj}^1 in detail. For $m = 1, 2, \cdots, n$,

$$\begin{aligned}
\gamma_{11}^1 &= p_{11} \\
\gamma_{m0}^1 &= p_{12}p_{22}^{m-1} \text{ for } m = 1, 2, \ldots, n \\
\gamma_{m1}^1 &= p_{11}p_{12}p_{22}^{m-2} + (m-2)p_{12}^2 p_{21} p_{22}^{m-3} + p_{12}p_{21}p_{22}^{m-2} \text{ for } m = 2, 3, \ldots, n
\end{aligned}$$

To compute the remaining probabilities, we first compute the first passage probabilities to state 1. Let $F^1(k)$ be the probability that the first visit to state 1 occurs after k periods, given that the initial state is state 1. Clearly:

$$\begin{aligned} F^1(1) &= p_{11} \\ F^1(m) &= p_{12}p_{22}^{m-2}p_{21} \text{ for } m = 2,3,\ldots,n \end{aligned}$$

Then, for $k = 2, 3, ...m$, we have

$$\gamma_{mk}^1 = \sum_{j=1}^{m-k+1} F^1(j)\gamma_{m-j,k-1} \text{ for } k = 2,3,\ldots,m$$

Similar expressions hold for the probabilities, γ_{mk}^2 for $m = 1, 2.., n$ and $k = 1, 2.., m$.

Proof of Proposition 2

According to the volatility updating scheme of this bi-directional regime switching model, $\sigma_{t+1}^{(K)} = \delta_i(K)$ (for some $1 < i < K$) if

$$\begin{aligned} \sqrt{\max(\frac{\frac{1}{2}[\delta_{i-1}^2(K)+\delta_i^2(K)]-\beta_0}{\beta_2\sigma_t^{(K)2}} - \frac{\beta_1}{\beta_2},0)} &\cdot \quad q_1(\varepsilon_t-\omega)^+ + q_2(\varepsilon_t-\omega)^- + (1-q_1-q_2)|\xi_t| \\ &< \sqrt{\max(\frac{\frac{1}{2}[\delta_i^2(K)+\delta_{i+1}^2(K)]-\beta_0}{\beta_2\sigma_t^{(K)2}} - \frac{\beta_1}{\beta_2},0)}. \end{aligned}$$

This in turn implies that

$$\begin{aligned} \max(\frac{\frac{1}{2}[\delta_{i-1}^2(K)+\delta_i^2(K)]-\beta_0}{\beta_2\sigma_t^{(K)2}} - \frac{\beta_1}{\beta_2},0) &\cdot \quad \left[q_1(\varepsilon_t-\omega)^+ + q_2(\varepsilon_t-\omega)^- + (1-q_1-q_2)|\xi_t|\right]^2 \\ &< \max(\frac{\frac{1}{2}[\delta_i^2(K)+\delta_{i+1}^2(K)]-\beta_0}{\beta_2\sigma_t^{(K)2}} - \frac{\beta_1}{\beta_2},0). \end{aligned}$$

Without loss of generality, we can drop the "max" function for $1 < i < K$. This is true because of two facts. First, if the upper bound equals zero for all $i < K$, then $\sigma_{t+1}^{(K)} = \delta_K(K)$. Second, if the lower bound equals zero for all $i > 1$, then $\sigma_{t+1}^{(K)} = \delta_1(K)$. Thus,

$$\begin{aligned} \frac{1}{2}[\delta_{i-1}^2(K)+\delta_i^2(K)] &\cdot \quad \beta_0 + \beta_1\sigma_t^{(K)2} + \beta_2\sigma_t^{(K)2}\left[q_1(\varepsilon_t-\omega)^+ + q_2(\varepsilon_t-\omega)^- + (1-q_1-q_2)|\xi_t|\right]^2 \\ &< \frac{1}{2}[\delta_i^2(K)+\delta_{i+1}^2(K)]. \end{aligned}$$

or

$$\begin{aligned} \frac{1}{2}[\delta_{i-1}^2(K)-\delta_i^2(K)] &\cdot \quad \beta_0 + \beta_1\sigma_t^{(K)2} + \beta_2\sigma_t^{(K)2}\left[q_1(\varepsilon_t-\omega)^+ + q_2(\varepsilon_t-\omega)^- + (1-q_1-q_2)|\xi_t|\right]^2 - \delta_i^2(K) \\ &< \frac{1}{2}[\delta_{i+1}^2(K)-\delta_i^2(K)]. \end{aligned}$$

By item (ii) of the partition condition, the lower (upper) bound approaches zero uniformly from below (above) as K tends to infinity. Thus, if such an i $(1 < i < K)$ exists,

$$\sigma_{t+1}^{(K)2} = \beta_0 + \beta_1\sigma_t^{(K)2} + \beta_2\sigma_t^{(K)2}\left[q_1(\varepsilon_t-\omega)^+ + q_2(\varepsilon_t-\omega)^- + (1-q_1-q_2)|\xi_t|\right]^2 + h(K)$$

where $h(K) \to 0$ as $K \to \infty$. Note that the existence of such an i is equivalent to

$$\begin{aligned}\frac{1}{2}[\delta_1^2(K)+\delta_2^2(K)] \quad &\cdot \quad \beta_0+\beta_1\sigma_t^{(K)2}+\beta_2\sigma_t^{(K)2}\left[q_1(\varepsilon_t-\omega)^+ + q_2(\varepsilon_t-\omega)^- + (1-q_1-q_2)|\xi_t|\right]^2 \\ &< \quad \frac{1}{2}[\delta_{K-1}^2(K)+\delta_K^2(K)]\end{aligned}$$

The variance updating scheme calls for

$$\sigma_{t+1}^{(K)2} = \begin{cases} \delta_1^2(K) & \text{if } \beta_0+\beta_1\sigma_t^{(K)2}+\beta_2\sigma_t^{(K)2}\left[q_1(\varepsilon_t-\omega)^+ + q_2(\varepsilon_t-\omega)^- + (1-q_1-q_2)|\xi_t|\right]^2 \\ & \qquad < \frac{1}{2}[\delta_1^2(K)+\delta_2^2(K)] \\ \delta_K^2(K) & \text{if } \beta_0+\beta_1\sigma_t^{(K)2}+\beta_2\sigma_t^{(K)2}\left[q_1(\varepsilon_t-\omega)^+ + q_2(\varepsilon_t-\omega)^- + (1-q_1-q_2)|\xi_t|\right]^2 \\ & \qquad \geq \frac{1}{2}[\delta_{K-1}^2(K)+\delta_K^2(K)] \end{cases}$$

Item (i) of the partition condition ensures that $\delta_1^2(K) \to 0$ and $\delta_K^2(K) \to \infty$ as $K \to \infty$. Therefore, for a fixed $\sigma_t^{(K)}$,

$$\sigma_{t+1}^{(K)2} \to \beta_0+\beta_1\sigma_t^{(K)2}+\beta_2\sigma_t^{(K)2}\left[q_1(\varepsilon_t-\omega)^+ + q_2(\varepsilon_t-\omega)^- + (1-q_1-q_2)|\xi_t|\right]^2 \text{ almost surely in } P$$

because $\mathrm{Prob}^P\{\beta_0+\beta_1\sigma_t^{(K)2}+\beta_2\sigma_t^{(K)2}\left[q_1(\varepsilon_t-\omega)^+ + q_2(\varepsilon_t-\omega)^- + (1-q_1-q_2)|\xi_t|\right]^2 < 0\} = 0$ and $\mathrm{Prob}^P\{\beta_0+\beta_1\sigma_t^{(K)2}+\beta_2\sigma_t^{(K)2}\left[q_1(\varepsilon_t-\omega)^+ + q_2(\varepsilon_t-\omega)^- + (1-q_1-q_2)|\xi_t|\right]^2 = \infty\} = 0$. It is also clear that $S_{t+1}^{(K)} \to S_{t+1}$ almost surely in P as $K \to \infty$ if $S_t^{(K)} \to S_t$ and $\sigma_{t+1}^{(K)2} \to \sigma_{t+1}^2$ almost surely in P. Starting from $t=0$ and knowing that $S_0^{(K)} = S_0$ and $\sigma_1^{(K)2} = \sigma_1^2$, we can conclude that $S_1^{(K)} \to S_1$ almost surely in P as $K \to \infty$. Then, $\sigma_2^{(K)2} \to \sigma_2^2$ and $S_2^{(K)} \to S_2$ almost surely in P as $K \to \infty$. Repeating the same argument all the way to time T. Since T is finite, the almost sure convergence holds true over the horizon $[0, T]$.

Proof of Proposition 3

First note that

$$\rho = E^Q\left\{\beta_1+\beta_2[q_1(\varepsilon_t^*-\omega-\lambda)^+ + q_2(\varepsilon_t^*-\omega-\lambda)^- + (1-q_1-q_2)|\xi_t^*-v|]^2|\mathcal{F}_{t-1}\right\}.$$

Its various components can be computed using the following function:

$$\begin{aligned}\Phi(c;a) &= \frac{1}{\sqrt{2\pi}}\int_a^\infty \exp(cz-\frac{z^2}{2})dz \\ &= \exp(\frac{c^2}{2})[1-N(a-c)].\end{aligned}$$

Specifically,

$$\begin{aligned}E^Q\left\{(\varepsilon_t^*-\omega-\lambda)^+|\mathcal{F}_{t-1}\right\} &= \Phi'(0;\omega+\lambda)-(\omega+\lambda)[1-N(\omega+\lambda)] \\ &= N'(\omega+\lambda)-(\omega+\lambda)[1-N(\omega+\lambda)] \\ E^Q\left\{(\varepsilon_t^*-\omega-\lambda)^-|\mathcal{F}_{t-1}\right\} &= E^Q\left\{(\varepsilon_t^*-\omega-\lambda)^+|\mathcal{F}_{t-1}\right\} - E^Q\left\{(\varepsilon_t^*-\omega-\lambda)|\mathcal{F}_{t-1}\right\}\end{aligned}$$

$$
\begin{aligned}
&= N'(\omega+\lambda) - (\omega+\lambda)[1 - N(\omega+\lambda)] + (\omega+\lambda) \\
&= N'(\omega+\lambda) + (\omega+\lambda)N(\omega+\lambda) \\
E^Q\left\{|\xi_t^* - v||\mathcal{F}_{t-1}\right\} &= E^Q\left\{(\xi_t^* - v)^+|\mathcal{F}_{t-1}\right\} + E^Q\left\{(\xi_t^* - v)^-|\mathcal{F}_{t-1}\right\} \\
&= N'(v) - v[1 - N(v)] + N'(v) + vN(v) \\
&= 2N'(v) + 2vN(v) - v \\
E^Q\left\{\left[(\varepsilon_t^* - \omega - \lambda)^+\right]^2|\mathcal{F}_{t-1}\right\} &= \Phi^{''}(0;\omega+\lambda) - 2(\omega+\lambda)\Phi'(0;\omega+\lambda) + \\
&\quad (\omega+\lambda)^2[1 - N(\omega+\lambda)] \\
&= \left[1 + (\omega+\lambda)^2\right][1 - N(\omega+\lambda)] - (\omega+\lambda)N'(\omega+\lambda) \\
E^Q\left\{\left[(\varepsilon_t^* - \omega - \lambda)^-\right]^2|\mathcal{F}_{t-1}\right\} &= E^Q\left\{\left[(\varepsilon_t^* - \omega - \lambda)^+ - (\varepsilon_t^* - \omega - \lambda)\right]^2|\mathcal{F}_{t-1}\right\} \\
&= E^Q\left\{(\varepsilon_t^* - \omega - \lambda)^2|\mathcal{F}_{t-1}\right\} - E^Q\left\{\left[(\varepsilon_t^* - \omega - \lambda)^+\right]^2|\mathcal{F}_{t-1}\right\} \\
&= 1 + (\omega+\lambda)^2 - \left[1 + (\omega+\lambda)^2\right][1 - N(\omega+\lambda)] + \\
&\quad (\omega+\lambda)N'(\omega+\lambda) \\
&= \left[1 + (\omega+\lambda)^2\right]N(\omega+\lambda) + (\omega+\lambda)N'(\omega+\lambda) \\
E^Q\left\{|\xi_t^* - v|^2|\mathcal{F}_{t-1}\right\} &= E^Q(\xi_t^{*2}|\mathcal{F}_{t-1}) - 2vE^Q(\xi_t^*|\mathcal{F}_{t-1}) + v^2 \\
&= 1 + v^2 \\
E^Q\left\{(\varepsilon_t^* - \omega - \lambda)^+(\varepsilon_t^* - \omega - \lambda)^-|\mathcal{F}_{t-1}\right\} &= 0 \\
E^Q\left\{(\varepsilon_t^* - \omega - \lambda)^+|\xi_t^* - v||\mathcal{F}_{t-1}\right\} &= E^Q\left\{(\varepsilon_t^* - \omega - \lambda)^+|\mathcal{F}_{t-1}\right\}E^Q\left\{|\xi_t^* - v||\mathcal{F}_{t-1}\right\} \\
E^Q\left\{(\varepsilon_t^* - \omega - \lambda)^-|\xi_t^* - v||\mathcal{F}_{t-1}\right\} &= E^Q\left\{(\varepsilon_t^* - \omega - \lambda)^-|\mathcal{F}_{t-1}\right\}E^Q\left\{|\xi_t^* - v||\mathcal{F}_{t-1}\right\}
\end{aligned}
$$

The expected volatility, conditional on $\mathcal{F}_0$, can be derived as:

$$
\begin{aligned}
E^Q(\sigma_{t+1}^2|\mathcal{F}_0) &= \beta_0 + E^Q\left\{\sigma_t^2\left(\begin{array}{c}\beta_1 + \beta_2[q_1(\varepsilon_t^* - \omega - \lambda)^+ + q_2(\varepsilon_t^* - \omega - \lambda)^- + \\ (1 - q_1 - q_2)|\xi_t^* - v|]^2\end{array}\right)|\mathcal{F}_0\right\} \\
&= \beta_0 + E^Q\left\{\sigma_t^2 E^Q\left[\left(\begin{array}{c}\beta_1 + \beta_2[q_1(\varepsilon_t^* - \omega - \lambda)^+ + q_2(\varepsilon_t^* - \omega - \lambda)^- + \\ (1 - q_1 - q_2)|\xi_t^* - v|]^2\end{array}\right)|\mathcal{F}_{t-1}\right]|\mathcal{F}_0\right\} \\
&= \beta_0 + \rho E^Q(\sigma_t^2|\mathcal{F}_0) \\
&= \sigma_1^2\rho^t + \frac{\beta_0(1-\rho^t)}{1-\rho}.
\end{aligned}
$$

Equivalently,

$$
E^Q(\sigma_t^2|\mathcal{F}_0) = \sigma_1^2\rho^{t-1} + \frac{\beta_0(1-\rho^{t-1})}{1-\rho}.
$$

Let t go to infinity, we have

$$
E^Q(\sigma_t^2) = \frac{\beta_0}{1-\rho}.
$$

References

Bakshi, G., C. Cao and Z. Chen (1997):Empirical Performance of Alternative Option Pricing Models, *The Journal of Finance*, 53, 499-547.

Black, F. (1976): Studies in Stock Price Volatility Changes, *Proceedings of the 1976 Meeting of the Business and Economics Statistics Section*, American Statistical Association, 177–181.

Bekaert, G. and R. Hodrick (1993): On Biases in the Measurement of Foreign Exchange Risk Premiums, *Journal of International Money and Finance*, 12, 115-138.

Bollen (1998): Valuing Options in Regime-Switching Models, *Journal of Derivatives*, 6, p.38-49.

Bollen, Gray and R. Whaley (1999): Valuation of Currency Options When Exchange Rates Shift Regimes, *Journal of Econometrics*, Forthcoming.

Duan, J. (1995): The GARCH Option Pricing Model, *Mathematical Finance*, 5, 13–32.

Duan, J. (1996): A Unified Theory of Option Pricing under Stochastic Volatility - from GARCH to Diffusion, unpublished manuscript, Hong Kong University of Science and Technology.

Duan, J. (1997): Augmented GARCH(p,q) Process and its Diffusion Limit, *Journal of Econometrics* 79, 97-127.

Duan, J. and Simonato (1998): Empirical Martingale Simulation for Asset Prices, *Management Science 44*, 1218-1433.

Duan, J. and Simonato (1999): American Option Pricing under GARCH by a Markov Chain Approximation, unpublished manuscript, Hong Kong University of Science and Technology.

Dumas, B., J. Fleming and B. Whaley (1998): Implied Volatility Functions: Empirical Tests, *Journal of Finance, 53*.

Durland, M. and T. McCurdy (1994): Duration Dependent Transitions in a Markov Model of U.S. GNP Growth, *Journal of Business and Economic Statistics*.

Engle, R. and V. Ng (1993): Measuring and Testing the Impact of News on Volatility, *Journal of Finance*, 48, 1749-1778.

Glosten, L., R. Jagannathan and D. Runkle (1993): Relationship between the Expected Value and the Volatility of the Nominal Excess Return on Stocks, *Journal of Finance*, 48, 1779-1801.

Gray, S. F. (1996): Modeling the Conditional Distribution of Interest Rates as a Regime Switching Process, *Journal of Financial Economics*, 42, 27-62.

Hamilton, J. (1989): A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle, *Econometrica*, 57, 357-384.

Harvey, C. and R. Whaley (1992): Dividends and *S&P* 100 Index Option Valuation, *Journal of Futures Markets*, 12, 123-137.

Hull, J. and A. White (1987): The Pricing of Options on Assets with Stochastic Volatility, *Journal of Finance*, 42, 281-300.

Jackwerth and M. Rubinstein (1996): Recovering probability Distributions from Option Prices, *Journal of Finance*, 51, 1611-1631.

Ritchken, P. and R. Trevor (1999): Pricing Options Under Generalized GARCH and Stochastic Volatility Processes, *Journal of Finance*, 54, 377-402.

Table 1
Sensitivity of European call option prices to transition probabilities

p_{11} \ p_{22}	0.0	0.1	0.2	0.4	0.6	0.8	0.9	1.0
0	3.745	3.842	3.924	4.109	4.333	4.613	4.781	4.974
0.1	3.687	3.765	3.848	4.039	4.274	4.572	4.754	4.966
0.2	3.598	3.675	3.760	3.958	4.204	4.523	4.721	4.956
0.4	3.369	3.449	3.537	3.745	4.016	4.384	4.626	4.926
0.6	3.049	3.126	3.213	3.425	3.716	4.145	4.451	4.864
0.8	2.553	2.616	2.688	2.875	3.153	3.625	4.020	4.664
0.9	2.172	2.216	2.267	2.403	2.621	3.033	3.435	4.223
1.0	1.595	1.595	1.595	1.595	1.595	1.595	1.595	1.595

Table 1 shows the sensitivity of a 30-day at the money option to the parameters p_{11} and p_{22} of the transition matrix. The values for the two volatilities are 10% and 50% per year. The interest rate is 10% per year. When $p_{11} = 1$ the variance can never move into the high level state. When $p_{22} = 1$, the variance remains in the low level state for a random amount of time before being absorbed into the high level state.

Table 2
Convergence of European call option prices to their theoretical values

	Maturity (days)				
m	10	30	60	90	180
1	1.867	3.641	5.463	6.943	10.553
2	1.850	3.631	5.457	6.938	10.550
3	1.852	3.631	5.456	6.937	10.549
5	1.849	3.629	5.455	6.936	10.549
10	1.846	3.627	5.454	6.935	10.548
∞	1.845	3.625	5.452	6.934	10.547

Table 2 shows the convergence rate of the algorithm. As m increases, the discrete approximation to the normal distribution over each trading period improves. As an example when $m = 3$, then each normal distribution is approximated by a discrete random variable taking on $2m+1=7$ points. The case parameters for this problem are the same as in Table 1. As the maturity of the contract increases, the value required for m to give very precise results decreases. The above results are typical of convergence rates for at-the-money contracts. Prices of out- the-money contracts converge equally fast.

Table 3
Convergence of the lattice

# Regimes	strike=95	strike=100	strike=105
10	6.060	2.976	1.185
20	6.040	2.950	1.163
50	5.986	2.878	1.105
100	6.071	2.990	1.197
200	6.050	2.961	1.173
300	6.029	2.936	1.154
400	6.037	2.946	1.161
500	6.037	2.945	1.161
1000	6.036	2.944	1.160
95% CI	(6.009,6.044)	(2.918,2.944)	(1.143,1.159)

Table 3 shows the rate of convergence of the regime switching model prices to NGARCH option prices. The contract priced is a 50 day European call option with the stock price at 100. All prices are computed using the lattice approximation described in Section 5. The parameters used are $\beta_0 = 0.000006575$, $\beta_1 = 0.9$, $\omega = 0.0$. The initial and stationary volatility are at 20%. Different regime levels satisfy the partition condition. Particularly, we used $l = 10$, which results in volatility levels between 0.1% and 40% for 100 regime model and between 0% and 90% for 1000 regime model. The reported prices correspond to a volatility level artificially fixed at 20%. The last two rows show the 95% confidence intervals for the limiting NGARCH option prices.

Table 4
Number of contracts in each moneyness and maturity category

Maturity	Moneyness	In-Sample	Week 1	Week 2	Week 4	Total
	<-0.04	205	155	194	201	755
10-45 days	(-0.04,-0.01)	158	144	151	157	610
Short term	(-0.01,0.01)	102	101	99	108	410
contracts	(0.01,0.04)	123	123	121	123	490
	>0.04	32	32	36	33	133
	<-0.04	296	286	305	293	1180
46-90 days	(-0.04,-0.01)	189	196	197	188	770
Mid term	(-0.01,0.01)	131	131	133	117	512
contracts	(0.01,0.04)	191	187	203	185	766
	>0.04	167	175	182	167	691
	<-0.04	266	265	269	265	1065
91-200 days	(-0.04,-0.01)	157	154	147	154	612
Longer term	(-0.01,0.01)	101	101	98	99	399
contracts	(0.01,0.04)	149	137	141	146	573
	>0.04	234	233	229	225	921
Total Contracts		2501	2420	2505	2461	9887

Table 5
In-sample pricing residuals box plots

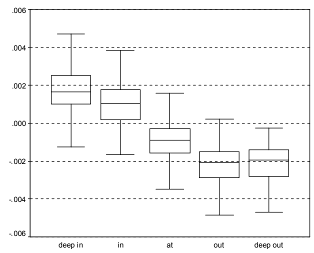
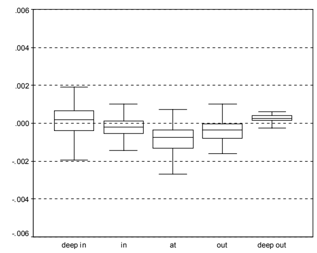
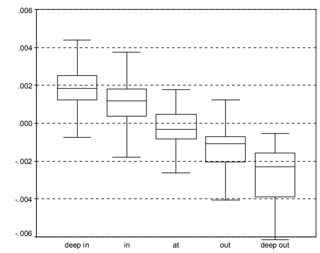
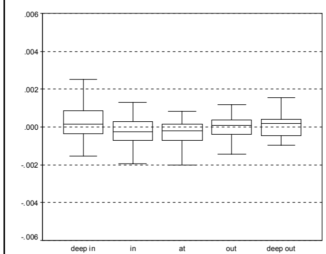
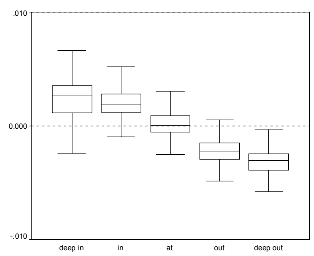
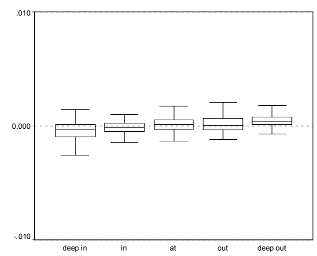
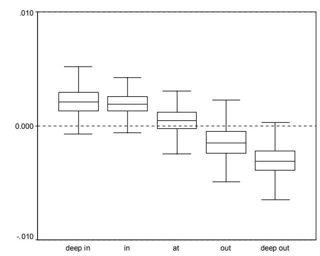
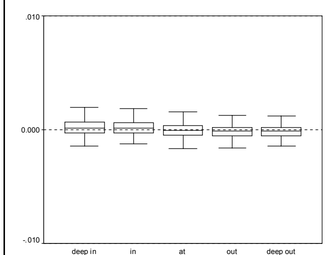
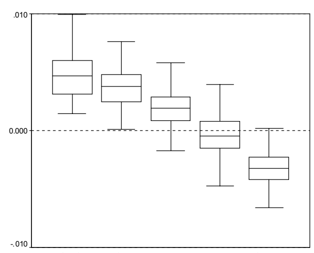
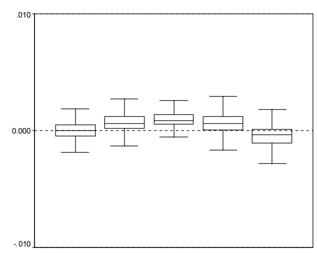
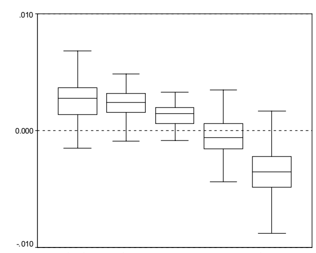
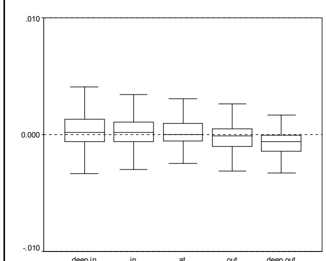
<table>
<tr><th></th><th>BLACK-SCHOLES</th><th>ADHOC</th><th>UNIDIRECTIONAL</th><th>BIDIRECTIONAL</th></tr>
<tr><td>SHORT-TERM RESIDUAS</td>
<td></td>
<td></td>
<td></td>
<td></td></tr>
<tr><td>MID-TERM RESIDUALS</td>
<td></td>
<td></td>
<td></td>
<td></td></tr>
<tr><td>LONGER-TERM RESIDUALS</td>
<td></td>
<td></td>
<td></td>
<td></td></tr>
</table>

Table 6a
Out-of-sample pricing residuals (1 week).

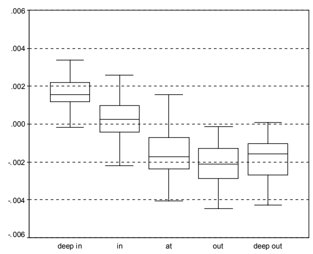
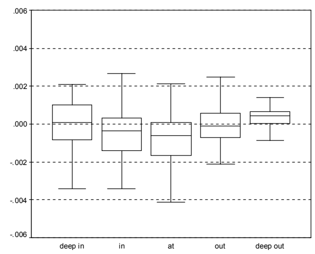
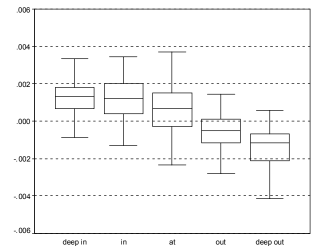
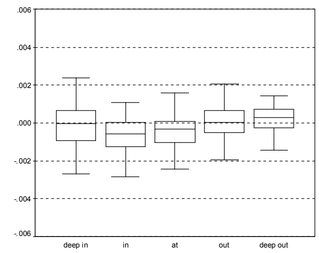
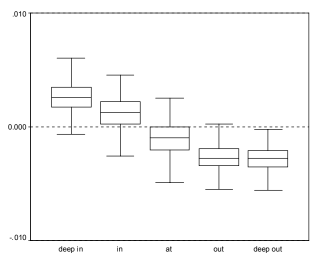
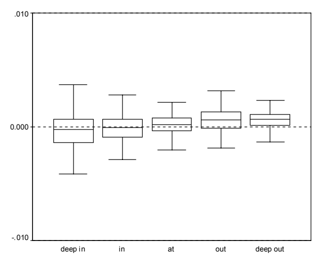
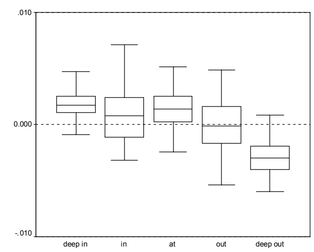
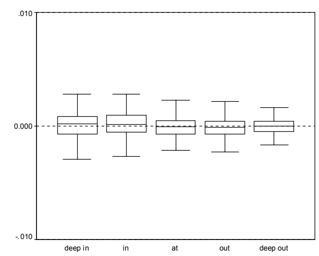
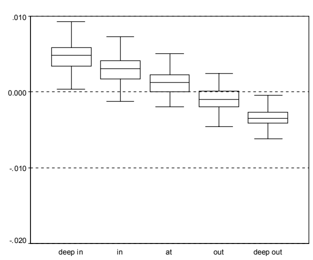
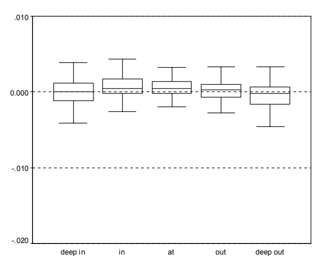
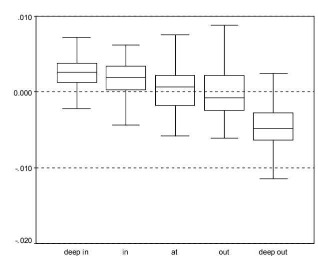
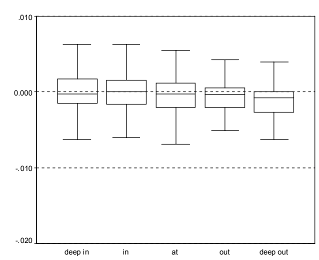
	BLACK-SCHOLES	ADHOC	UNIDIRECTIONAL	BIDIRECTIONAL
SHORT-TERM RESIDUAS				
MID-TERM RESIDUALS				
LONGER-TERM RESIDUALS				

Table 6a compares the relative performance for the four models one week out-of-sample. For the Black Scholes model, conditional on the dividend adjusted S&P 500 one week later, we reestimate the single volatility, that minimizes the sum of squared errors, and use that value to reprice all option contracts and to establish pricing residuals. For the uni-directional and bi-directional models we use the parameters estimated in the previous week. To estimate the theoretical option prices, the values of both state variables, namely the adjusted index level, and the local volatility, are required. The adjusted index value is of course known, and the local volatility is implied out from the data. To make a fair comparison with the Ad-Hoc model, we keep all the regression coefficients fixed except for the intercept, which is reestimated from the data. In this way, in the out-of-sample periods, all models have one degree of freedom.

Table 6b
Out-of-sample pricing residuals (2 weeks).

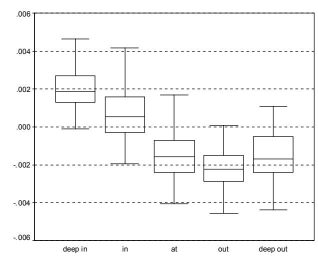
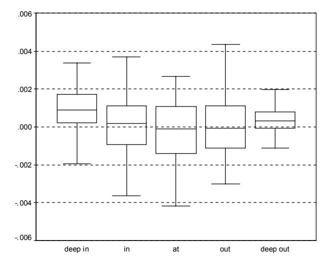
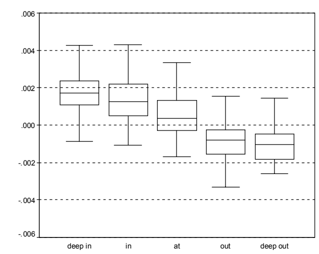
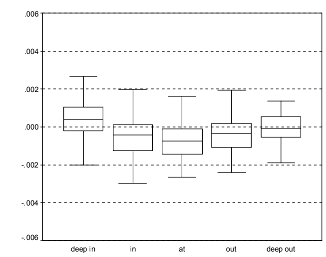
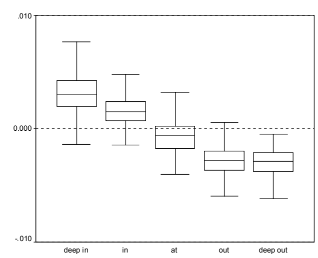
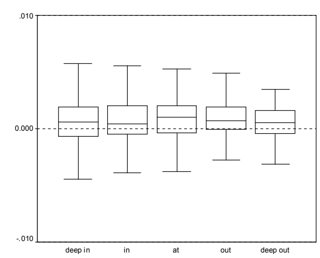
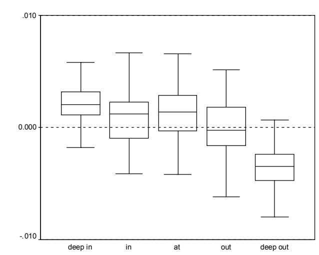
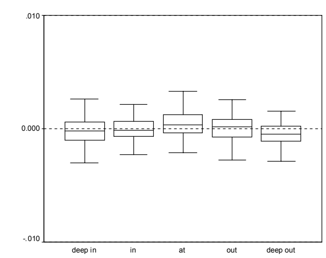
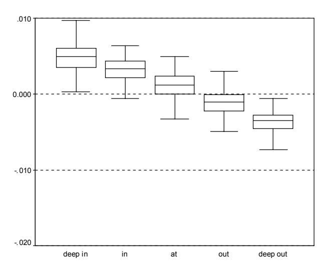
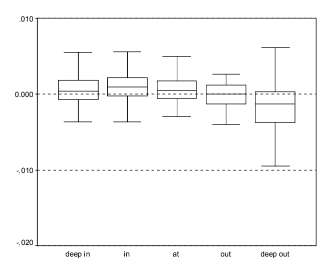
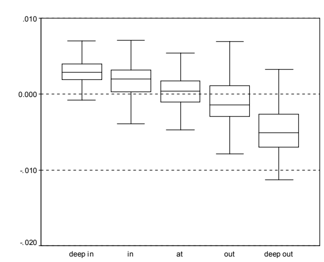
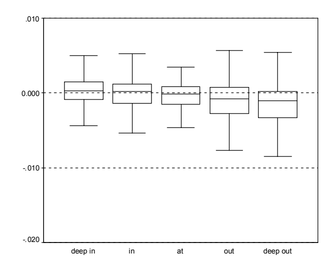
	BLACK-SCHOLES	ADHOC	UNIDIRECTIONAL	BIDIRECTIONAL
SHORT-TERM RESIDUAS				
MID-TERM RESIDUALS				
LONGER-TERM RESIDUALS				

Table 6b repeats the analysis from Table 6a two weeks after the in sample fitting was done.

Table 7
Out-of-sample (1 week) comparison of the Bidirectional and Unidirectional models (two regimes)

	Deep In	In	At	Out	Deep Out
Short	-0.32491	-0.39126	-0.20100	-0.22223	-0.85833
Mid	-0.57837	-0.51692	-0.68832	-0.82002	-1.58278
Long	-0.11244	0.01980	-0.04208	-0.32964	-0.73472

The entries in the cells correspond to $\ln\left[\frac{\sqrt{BI\_SSE}}{\sqrt{UNI\_SSE}}\right]$. Negative number indicates that Bidirectional model is better in average than Unidirectional. Note that only in 1 out of 15 categories Unidirectional performs better than Bidirectional.

Table 8
Parameter estimates

Models	Parameters					
Black-Scholes	σ					
Mean	0.138536					
StDev	0.020994					
AdHoc	β_0	β_1	β_2	β_3	β_4	β_5
Mean	2.984539	-1.526730	0.130279	-0.010840	1.01E-05	0.003308
StDev	1.534938	1.325737	0.435531	0.006444	7.65E-06	0.002413
Unidirectional	β_0	β_1	β_2	ω	σ_1	
Mean	6.74559E-07	0.77467	0.223375	0.834072	0.120833	
StDev	4.14482E-07	0.060883	0.060341	0.521808	0.037021	
Bidirectional	β_0	β_1	β_2	q	ω	σ_1
Mean	3.77E-06	0.403558	0.580733	0.402407	0.926783	0.093300
StDev	4.31E-06	0.241761	0.238937	0.074785	0.121275	0.049935

Figure 1

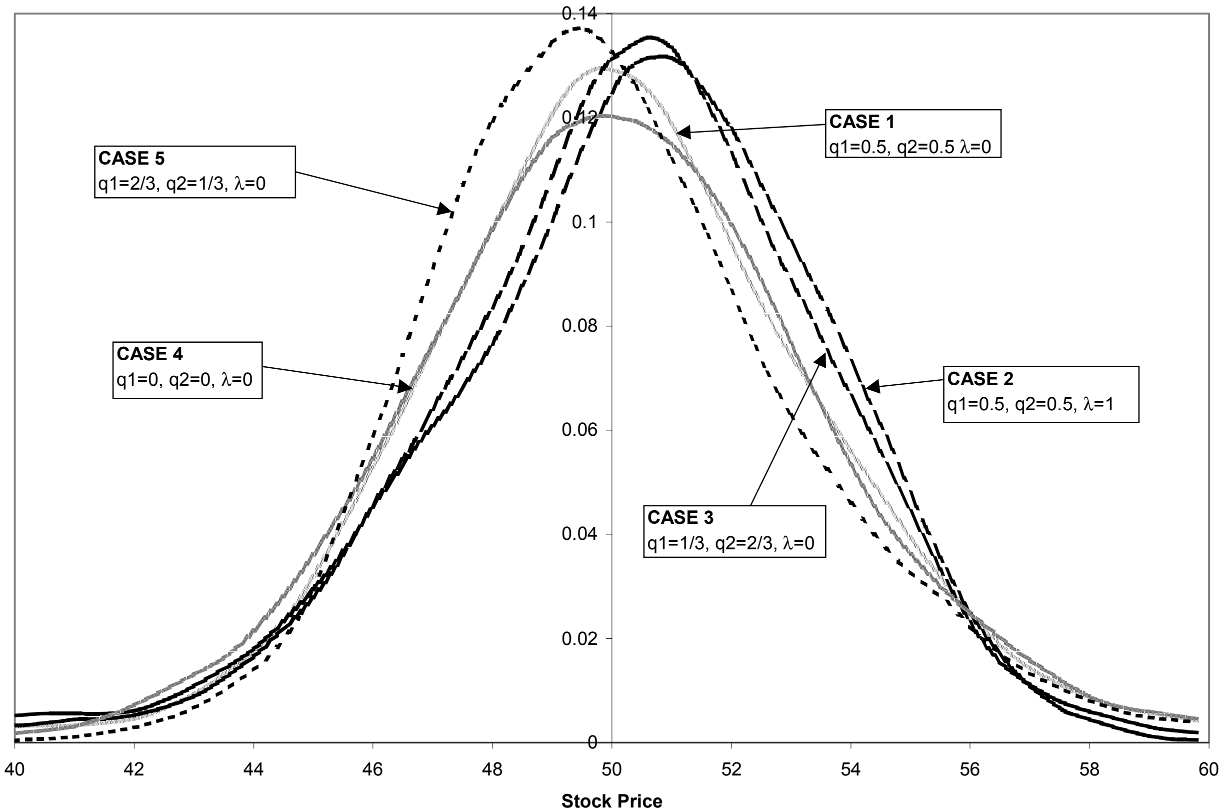


Figure 1 illustrates several risk-neutral density functions obtained for the stock price over a three-month horizon. The parameters that are fixed in this analysis are $\beta_0 = 0.00001$, $\beta_1 = 0.8$, $\omega = 0$, $v = 0$. Five different scenarios are considered: (1) $q_1 = q_2 = 0.5$, λ =0, (2) $q_1 = q_2 = 0.5$, λ =1, (3) $q_1 = 1/3$, $q_2 = 2/3$, λ =0, (4) $q_1 = q_2 = 0$, λ =0, and (5) $q_1 = 2/3$, $q_2 = 1/3$, λ =0. In all five cases, the stationary risk-neutral standard deviation (annualized) equals 20% because ρ is maintained at 0.90875.