Ch 6. Option Pricing under Non-Constant Volatility

I. Volatility Smile

II. Option Pricing When Volatility is a Function of S and/or t

III. Stochastic Volatility Models

IV. Option Pricing Under GARCH Model

V. Option Pricing Under Stochastic Volatility Process

- It is convincingly believed that the constant volatility assumption of the Black-Scholes model is rejected by many empirical facts. Therefore, this chapter introduces several (stochastic) volatility models and lattice option-pricing approaches to deal with the non-constant volatility.

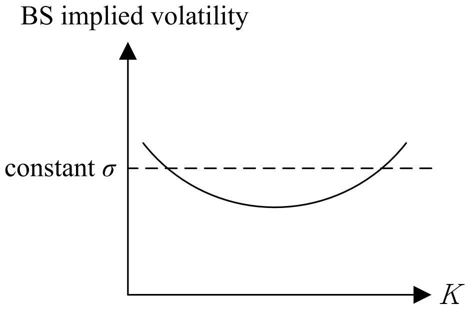
- The first approach is proposed by Nelson and Ramaswamy (1990) to consider the volatility of the underlying asset being a function of the time and the price of the underlying asset. The second approach, developed by Ritchken and Trevor (1999), is to price options when the underlying asset price follows the GARCH model. The third approach is the explicit FDM to price options when the underlying asset price follows stochastic volatility processes.

I. Volatility Smile

- The phenomenon of volatility smile exhibits the variation of the Black-Scholes implied volatility with respect to the strike price. In other words, the constant volatility assumption in the Black-Scholes is not so correct. Next, I will discuss the volatility smile and smirk for currency and equity options, respectively.

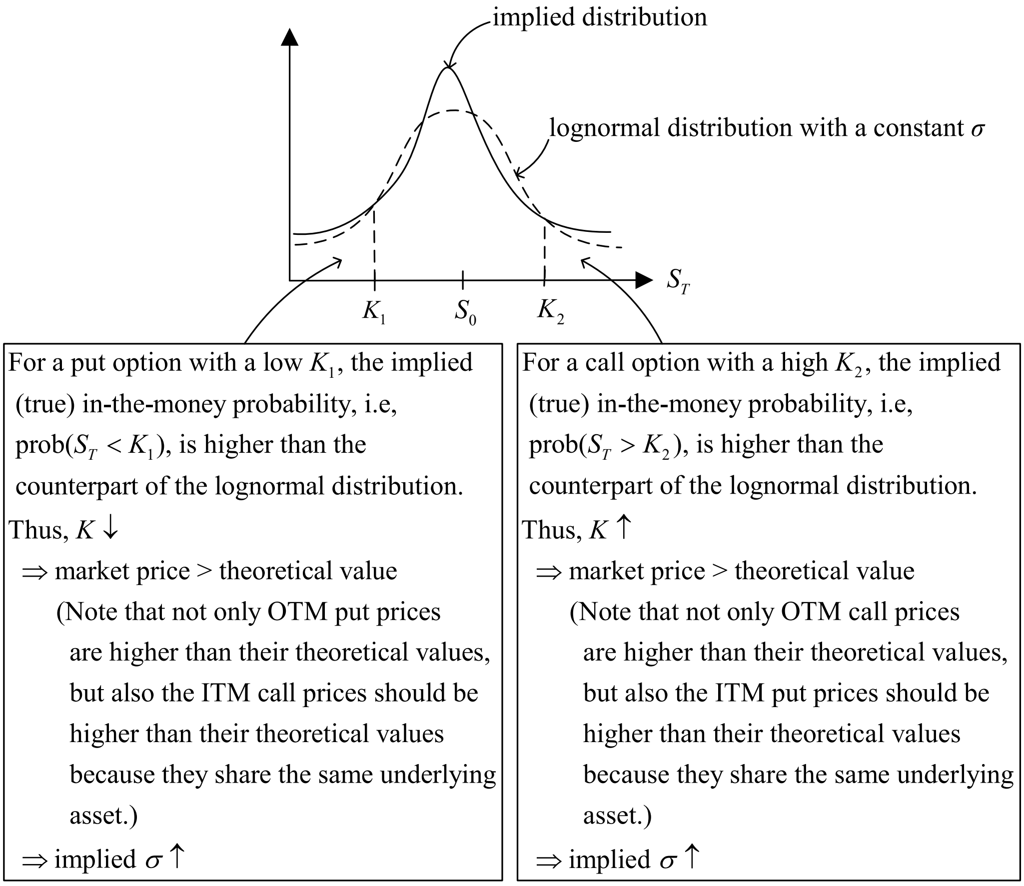
- Volatility smile for currency options

Figure 6-1 Illustration of the Volatility Smile for Currency Options



- It is intuitive that both implied volatilities of call or put options should be the same as long as they share a common underlying asset. Thus, either the implied volatilities of calls or the implied volatilities of puts could be employed to explain the volatility smile phenomenon.

Figure 6-2 Explanation of the Volatility Smile for Currency Options



- Possible reasons for the volatility smile of currency options:

  (i) The volatility of the underlying asset is never to be constant.

  (ii) The existence of price jump causes extreme returns and thus increase the probabilities when S_T is fairly high or low. Furthermore, for longer maturity, the jumps are averaged out and thus the jumps have weaker impacts on the probabilities for S_T being fairly high or low. So, the volatility smile dies out with the increase of maturity, which is consistent with empirical studies.

  (iii) The stochastic volatility or GARCH processes could generate the probability distributions of underlying assets with fatter tails.

- Volatility smirk for equity options (including index options and individual stock options):

Figure 6-3 Illustration of the Volatility Smirk for Equity Options

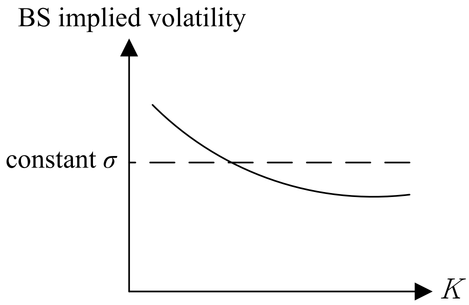
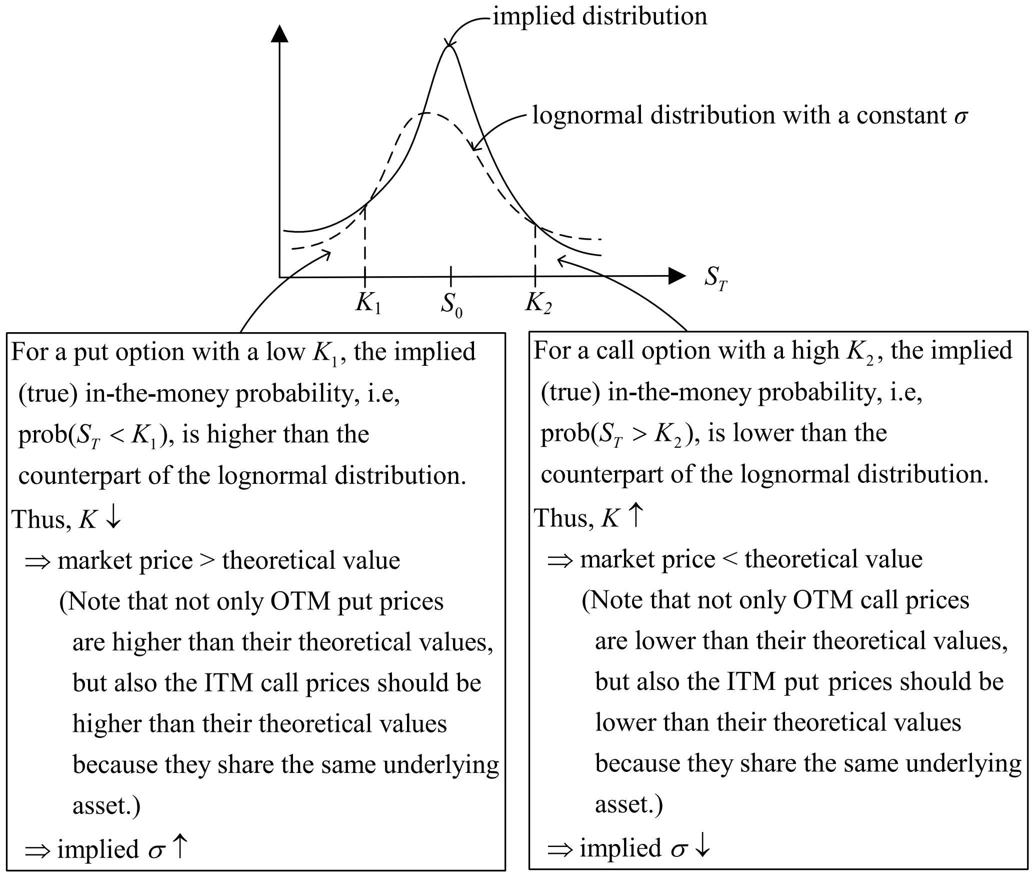


Figure 6-4 Explanation of the Volatility Smirk for Equity Options



- Possible reasons for the volatility smirk of equity options:

  (i) The above volatility smile pattern for equity options is ture after the crash of stock markets in October of 1987. Before October of 1987, the volatility is almost constant for different strike prices. Thus, Mark Rubinstein, a famous academic in the finance field and the inventor of the CRR binomial tree model, terms the reason underlying this volatility pattern as "crashophobia," meaning investors overestimate the probability that the stock markets go download.

  (ii) Leverage effect (first proposed in Christie (1982)): it is well known that when a firm employs more leverage, the financial risk of the firm increases and thus the volatility of the stock price increases to reflect this risk.

  $S \downarrow$ (leverage ratio $\uparrow$), volatility $\uparrow$, which makes the stock price further decline.

  $S \uparrow$ (leverage ratio $\downarrow$), volatility $\downarrow$, which makes the stock price further rise.

  (iii) The volatility should follow an asymmetric GARCH or stochastic process to reflect the countercyclical variation of the volatility, i.e., when the stock price rises (declines), it volatility decreases (increases).

- The relationship between the implied volatility and the time to maturity can be expressed as a volatility term structure. If both the strike price and the time to maturity are considered, a 3-D implied volatility function can be derived, which is called the volatility surface. By observing the volatility surface, it can be found that the volatility smile is pronounced for shorter times to maturity, but becomes minor for longer times to maturity.

- Option pricing models for non-constant volatilities:

  (i) Volatility is a function of K and/or T:

  $*$ A naive approach in practice is to combine the BS formula and the estimated (or historical) implied volatility surface. First, interpolate with respect to the target (K, T) on the implied volatility surface to obtain a proper implied volatility. Second, calculate the option value by the BS formula with the input of the implied volatility obtained above.

  (ii) Volatility is a function of S and/or t (T):

  Since the current time point t and the time to maturity T always change by the same magnitude but in opposite directions, the volatility to be a function of T implies the volatility to be a function of t.

  $*$ Nelson and Ramaswamy (1990) price options with the volatility as a function of S and t.

  (iii) Volatility movements conform the GARCH model:

  $*$ Ritchken and Trevor (1999) price options with the GARCH model.

  (iv) For stochastic volatility processes:

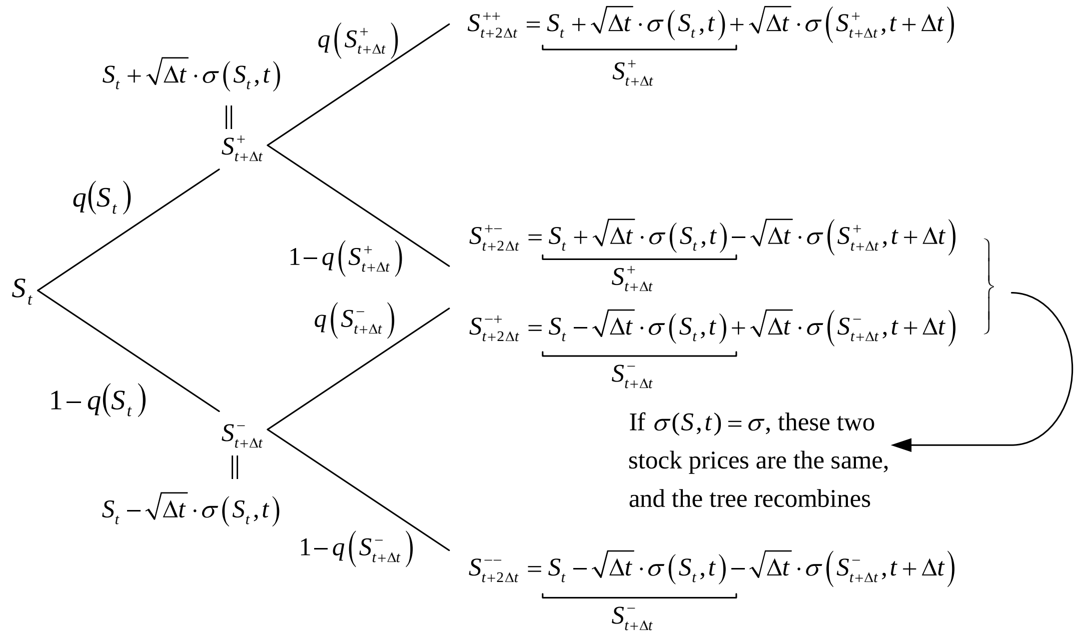
  $*$ I propose a FDM to price options under stochastic volatility.

II. Option Pricing When Volatility is a Function of S and/or t

- The setting of the stock price process and the binomial tree model in Nelson and Ramaswamy (1990), "Simple Binomial Processes as Diffusion Approximations in Financial Models," *Review of Financial Studies* 3, pp. 393–430.

  $\odot$ The underlying asset price is assumed to be an Itô process: $dS = \mu(S, t)dt + \sigma(S, t)dZ_t$. (Note that both the drift and the volatility terms are functions of S and t and could be stochastic. In addition, S can be any kind of underlying asset rather than only the stock price.)

Figure 6-5 Non-recombined issue of the binomial tree model when the volatility is not a constant



where $q(S_t)=\frac{E[S_{t+\Delta t}]-S^-_{t+\Delta t}}{S^+_{t+\Delta t}-S^-_{t+\Delta t}}=\frac{S_t+\mu(S_t,t)\Delta t-S^-_{t+\Delta t}}{S^+_{t+\Delta t}-S^-_{t+\Delta t}}$.

(It is obvious that the non-constant $\sigma(S,t)$ is the reason to make the binomial tree non-recombined. As for $\mu(S,t)$, it only affects the upward and downward branching probabilities for each node.)

- Main idea of Nelson and Ramaswamy (1990):

  ⊙ Suppose X is a function of S and t, and X is twice differentiable with respect to S and once differentiable with respect to t. According to the Itô's Lemma:

$$dX(S,t)=(\mu(S,t)\frac{\partial X(S,t)}{\partial S}+\frac{1}{2}\sigma^2(S,t)\frac{\partial^2 X(S,t)}{\partial S^2}+\frac{\partial X(S,t)}{\partial t})dt+(\sigma(S,t)\frac{\partial X(S,t)}{\partial S})dZ_t.$$

  Deliberately set the volatility term to be 1:

  $\Rightarrow \frac{\partial X(S,t)}{\partial S}\sigma(S_t,t)=1$

  $\Rightarrow X(S,t)=\int_S \frac{1}{\sigma(K,t)}dK$ (a mapping between X and S)

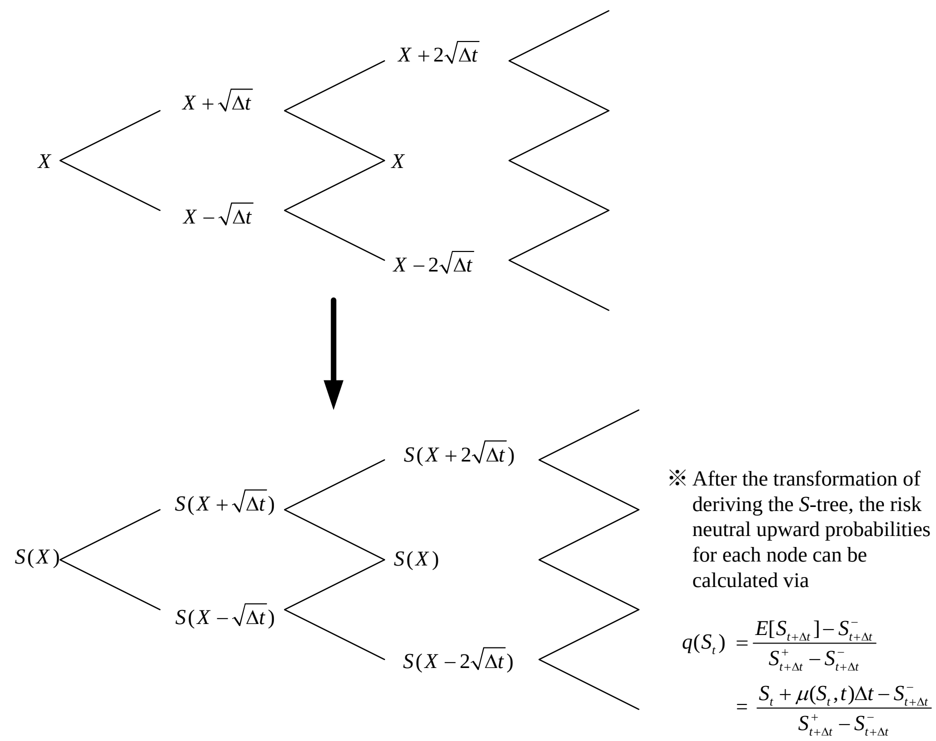
$\odot$ Three-step process to build a recombined tree for S:

(i) Build the binomial tree for X first. Since the volatility term of X is a constant or even equal to 1, the binomial tree for X recombines. (See Figure 6-6)

(ii) For each node, transform the value of X to the corresponding value of S such that $S(X,t) = \{S : X(S,t) = X\}$.

(iii) The resulting binomial tree for S retain recombined (see Figure 6-6), and the upward and downward probability for each node can be derived via $q(S_t) = \frac{E[S_{t+\Delta t}]-S^-_{t+\Delta t}}{S^+_{t+\Delta t}-S^-_{t+\Delta t}} = \frac{S_t+\mu(S_t,t)\Delta t-S^-_{t+\Delta t}}{S^+_{t+\Delta t}-S^-_{t+\Delta t}}$.

Figure 6-6



- Other implementation issues:

(i) $\sigma(S,t)$ cannot be zero, otherwise the value of X will approach infinity.

(Since S is a stochastic process, $\sigma(S,t)$ should not be zero. However, when the value of S is very small, sometimes $\sigma(S,t)$ is very close to 0, e.g., the CIR interest rate process.)

(ii) Δt should be small enough to ensure that $0 < q < 1$ and thus guarantee the binomial tree of S to model the process $S(t)$ properly.

(iii) There is no constraint for the negative value of X. However, for the underlying asset price S, it should be nonnegative.

(If you need to avoid negative S or sometimes the transformation function $S(X,t)$ only accepts positive values of X, e.g., $S(X,t)=\ln(X)$, the simplest modification you can try is to let the volatility term of X to be a number far smaller than 1 and thus restrict the variation range of X. More specifically, instead of setting $\frac{\partial X(S,t)}{\partial S}\sigma(S_t,t)=1$, you can consider, for example, $\frac{\partial X(S,t)}{\partial S}\sigma(S_t,t)=0.01$.)

- Several examples for Nelson and Ramaswamy (1990):

⊙ Example 1: Constant elasticity of variance (CEV) stock price process:

$dS=\mu S dt+\sigma S^{\gamma}dZ$, where $0<\gamma\leq 1$.

|| The feature of CEV process:

$$\left\{\begin{array}{l} S>1 \Rightarrow S^{\gamma}<S \text{ (higher } S\text{, lower volatility than the lognormal distribution)} \\ S<1 \Rightarrow S^{\gamma}>S \text{ (extermely lower } S\text{, higher volatility than the lognormal distribution)} \end{array}\right. .$$

$$\Rightarrow X(S,t)=\sigma^{-1}\int_S K^{-\gamma}dK=\frac{S^{1-\gamma}}{\sigma(1-\gamma)}$$

|| If $\gamma=1$, $dS=\mu S dt+\sigma S dZ$, $X(S,t)=\frac{1}{\sigma}\ln S$.

$$\Rightarrow S(X,t)=\left\{\begin{array}{ll} [\sigma(1-\gamma)X]^{\frac{1}{1-\gamma}} & \text{if } X>0 \\ \\ 0 & \text{o/w} \end{array}\right.$$

$$\Rightarrow \left\{\begin{array}{l} S_t \equiv S(X,t) \\ S^+_{t+\Delta t} \equiv S(X+J^+\sqrt{\Delta t},t+\Delta t) \\ S^-_{t+\Delta t} \equiv S(X-J^-\sqrt{\Delta t},t+\Delta t) \\ q* = \frac{S_t+\mu S_t\Delta t-S^-_{t+\Delta t}}{S^+_{t+\Delta t}-S^-_{t+\Delta t}} \\ q=\left\{\begin{array}{ll} q* & \text{if } 0\leq q*\leq 1 \\ 0 & \text{if } q*<0 \\ 1 & \text{if } q*>1 \end{array}\right. \end{array}\right. .$$

$*$ J^+ and J^- are introduced by Nelson and Ramaswamy (1990) to allow multiple jumps such that in most cases, the $E[S_{t+\Delta t}]$ is inbetween two following branches and thus the upward probability q^* is in $[0,1]$. The rules to decide J^+ and J^- are as follows.

$$J^+=\left\{\begin{array}{ll} \text{the smallest, odd, positive, integer } j \text{ s.t.} & \\ \quad S(X+j\sqrt{\Delta t},t+\Delta t)-S(X,t)\geq \mu(S,t)\Delta t & \text{if } X<X_L \\ \\ 1 & \text{if } X\geq X_L \end{array}\right. .$$

To prevent the explosive growth of the number of nodes on the binomial tree, the upward multiple jumps are allowed to reach the nodes already on the binomial tree only when $X < X_L$. In addition, the increment of the underlying asset price of the upward branch should be higher than the expected growth of the underlying asset price. (For $X \geq X_L$, if the upward multiple jumps are allowed, new nodes should be generated, which could result in the unexpected grow of the binomial tree.)

$$J^{-} = \begin{cases} \text{the smallest, odd, positive, integer } j \text{ s.t.} \\ \quad S(X,t) - S(X - j\sqrt{\Delta t}, t+\Delta t) \leq \mu(S,t)\Delta t \\ \text{or} \\ \quad S(X - j\sqrt{\Delta t}, t+\Delta t) = 0 \text{ (} S \text{ should be nonnegative)} \end{cases}.$$

The decrement of the underlying asset price of the downward branch should be smaller than the expected growth (may be negative) of the underlying asset price. However, the smallest underlying asset price which can be reached is 0 becuase S is nonnegative.

∗ Given fixed μ, σ, and γ, it is possible to identify a small enough Δt to avoid the negative branching probability issue in this simple model.

∗ I suggest to find a small Δt first. If it is unattainable, the multiple jumps approach suggested by Nelson and Ramaswamy (1990) is then considered.

⊙ Example 2: $dS = \mu(S,t)dt + \sigma(S,t)dZ$, where $\mu(S,t) = \mu S$ and $\sigma(S,t) = \sigma S$.

$$\Rightarrow \sigma \cdot S \cdot \frac{\partial X}{\partial S} = 1 \Rightarrow \partial X = \frac{1}{\sigma} \cdot \frac{\partial S}{S}$$
$$\Rightarrow X(S) = \frac{1}{\sigma} \cdot \ln S \Rightarrow S = e^{\sigma X}.$$

Therefore, for $dX(S) = (\frac{\mu}{\sigma} - \frac{1}{2}\sigma)dt + 1 \cdot dZ$, X-tree can be built first. Next, S-tree can be derived through the transformation $S = e^{\sigma X}$. In fact, the resulting S-tree is exactly identical to the CRR binomial tree.

⊙ Example 3: $dS = \mu(S,t)dt + \sigma(t)SdZ$, where $\sigma(t)$ is a stepwise function of t defined as follows.

Suppose the whole period is partitioned into 3 subperiods, and

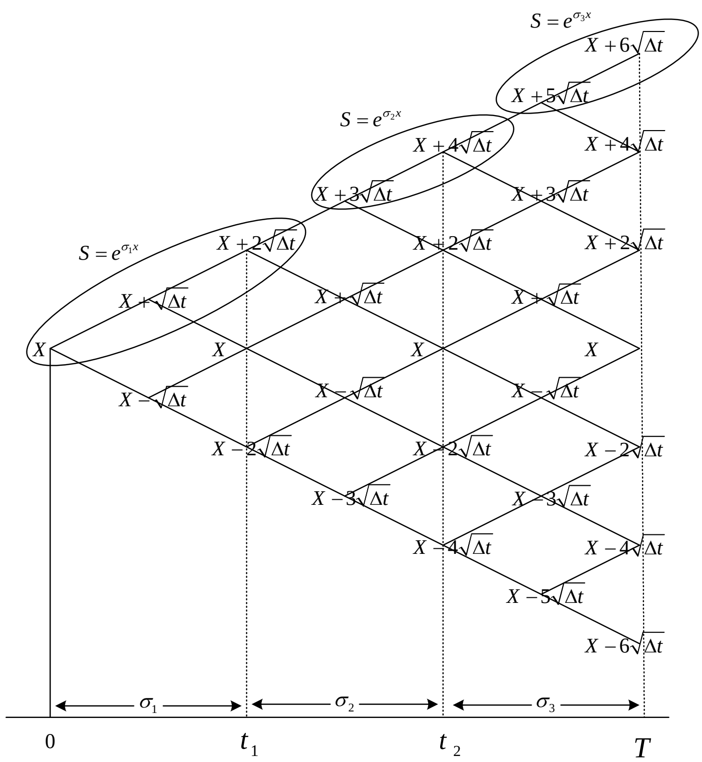
$$\sigma(t) = \begin{cases} \sigma_1 & 0 \leq t \leq t_1 \\ \sigma_2 & t_1 < t \leq t_2 \\ \sigma_3 & t_2 < t \leq T \end{cases}.$$

(The step function of σ_t is useful to discribe the different stages of a growing firm, i.e., seed, growing, and matured stages of a firm.)

$$\text{Nelson and Ramaswamy (1990)} \Rightarrow X(S,t) = \begin{cases} \frac{1}{\sigma_1}\ln S \Rightarrow S = e^{\sigma_1 X} & 0 \le t \le t_1 \\ \frac{1}{\sigma_2}\ln S \Rightarrow S = e^{\sigma_2 X} & t_1 < t \le t_2 \\ \frac{1}{\sigma_3}\ln S \Rightarrow S = e^{\sigma_3 X} & t_2 < t \le T \end{cases} .$$

(It is worth noting that for the 3 subperiods, the volatility terms of the X process are all equal to 1. Therefore, the X-tree is identical to the one in Figure 6-7 regardless of different subperiods. For different subperiods, only the transformation functions listed above are not the same.)

Figure 6-7



III. Stochastic Volatility (SV) Models

- Affine (jump-diffusion) SV models (e.g., Guo, Hung, and So (2009):

$$\frac{dS}{S} = \mu dt + \sqrt{V} dZ_S + (e^x - 1)dq_S - E[(e^x - 1)dq_S],$$
$$dV = \kappa(\theta - V)dt + \sigma_V \sqrt{V} dZ_V + y dq_V,$$

where $\mathrm{corr}(dZ_S, dZ_V) = \rho_{SV}$.

(i) Correlated jumps

q_S and q_V are correlated Poisson counters with intensity $\lambda_{x,y}$,

x is the percentage jump in S and $x \sim ND(\mu_0 + \mu_{x,y} y, \sigma^2_{x,y})$,

y is the level jump in V and $y \sim Exp(\theta_y)$,

$E[(e^x - 1)dq_S] = \lambda_{x,y} \frac{e^{\mu_0 + 0.5\sigma^2_{x,y}}}{1 - \theta_y \mu_{x,y}} - 1.$

(ii) Independent jumps

q_S and q_V are independent Poisson counters with intensity λ_x and λ_y, respectively,

x is the percentage jump in S and $x \sim ND(\ln(1 + \mu_x) - 0.5\sigma^2_x, \sigma^2_x)$,

y is the level jump in V and $y \sim Exp(\theta_y)$,

$E[(e^x - 1)dq_S] = \lambda_x E[(e^x - 1)]dt = \lambda_x \mu_x dt.$

$*$ The introduction of $-E[(e^x - 1)dq_S]$ in the process of dS/S is to maintain the growth rate of S to be μ.

$*$ Only affine SV (jump-diffusion) models are tractable to derive analytical option pricing formula (see Duffie, Pan, and Singleton (2000) or Guo, Hung and So (2009)).

- Non-affine SV models:

$$\frac{dS}{S} = \mu dt + \sqrt{V} dZ_S,$$

(i) Damped CEV and CEV SV models (e.g., Hung, Ko, and Wang (2023))

$$dV = \kappa(\theta - V)dt + (\sigma_1 V^{0.5} + \sigma_2 V^{\gamma}) D(V) dZ_V,$$

where $D(V) = -e^{8V^4}$.

$*$ If $\sigma_1 = 0$ and $D(V) = 1$, the damped CEV SV model reduces to the CEV SV model (non-affine).

$*$ If $\sigma_2 = 0$ and $D(V) = 1$, the damped CEV SV model reduces to the Heston SV model (affine).

$*$ When V is high, $D(V)$ approaches zero and the volatility term of the damped CEV SV model approaches zero as well. Without the intervene of the volatility term, the downward mean-reverting force from the linear drift term is effectively pronounced.

$*$ Moreover, for the pure CEV SV model, if the exponent term γ is greater than 1, it could result in an explosive scenario for some large values of V. In contrast, the DCEV SV model eliminates this possibly explosive behavior with the help of a small $D(V)$ when V is high and thus relaxes the constraint on estimating γ.

$*$ In contrast, when V is low, it is almost impossible to encounter an explosive scenario even though γ is greater than 1. Therefore, the damped CEV SV model reduces to a CEV SV model without the constraint on γ.

(ii) Non-linear drift model (e.g., Christoffersen, Jacobs, and Mimouni (2010), Chourdakis and Dotsis (2011), Mijatovic and Schneider (2014))

$$dV = (\alpha_0 + \alpha_1 V + \alpha_2 V^2 + \alpha_3 V^{-1})dt + \sigma_V V^\gamma dZ_V.$$

$*$ The non-linear form of the drift term, $\alpha_0 + \alpha_1 V + \alpha_2 V^2 + \alpha_3 V^{-1}$, is first proposed in Ait-Sahalia (1996) for modeling the stochastic interest rate process. This form of non-linear drift term is claimed to better capture the mean-reverting behavior of stochastic variance.

$*$ If $\alpha_0 = \alpha_3 = 0$ and $\gamma = 3/2$, the above model reduces to the 3/2N mode, which is demonstrated in Christoffersen, Jacobs, and Mimouni (2010) to be an SV model with superior fitting performance.

(iii) Log variance model (e.g., Durham (2013))

$$dV = [\kappa(\theta - \ln V) + 0.5\sigma_V^2]V dt + \sigma_V \sqrt{V} dZ_V.$$

$*$ If $Y = \ln V$, then $dY = \kappa(\theta - Y)dt + \sigma_V dZ_V$.

$*$ The major advantage of the log variance model is to eliminate the possibility for V to be negative, even with the presence of the discretization error.

$*$ Masoliver and Perello (2006) and Bormetti, Cazzola, and Delpini (2010) argue that it is better to approximate the probability distribution of stochastic variance with a lognormal distribution.

IV. Option Pricing under GARCH Model

- Ritchken and Trevor (1999),"Pricing Option under Generalized GARCH and Stochastic Volatility," *Journal of Finance* 54, pp. 377–402.

⊙ Main idea:

(i) A trinomial tree framework for the log stock price is considered.

(ii) Unlike the standard trinomial tree framework introduced in Chapter 4, a grid structure of log stock price is constructed first and thus the upward and downward tick changes are fixed initially.

(iii) The most important originality of this paper is that the change of each branching log stock price implies an innovation of the standard normally distributed variable in the Wiener process. Next, the innovation is employed to update the conditional variance level reaching that branch node.

(iv) Instead of recording all possible values of variances on each node, only several selected representative variances are recorded on each node. During backward induction, the linear interpolation method is employed to find the corresponding option values for the missing variances.

⊙ General settings:

(i) Nonlinear asymmetric GARCH (NGARCH)

$$\ln(\tfrac{S_{t+1}}{S_t}) = r + \lambda\sqrt{h_t} - \tfrac{1}{2}h_t + \sqrt{h_t}\upsilon_{t+1},$$
$$h_{t+1} = \beta_0 + \beta_1 h_t + \beta_2 h_t(\upsilon_{t+1} - C)^2.$$

(The above two equations represent the NGARCH process under the physical measure, in which h_t represents the conditional variance process, λ is the market price of risk of S, β_0, β_1, β_2, and C are constants, and υ_{t+1} follows the standard normal distribution)

Under the risk-neutral measure, the corresponding NGARCH process is as follows.

$$\ln(\tfrac{S_{t+1}}{S_t}) = (r - \tfrac{1}{2}h_t) + \sqrt{h_t}\varepsilon_{t+1},$$
$$h_{t+1} = \beta_0 + \beta_1 h_t + \beta_2 h_t(\varepsilon_{t+1} - C^*)^2.$$

where $C^* = C + \lambda$, and ε_{t+1} follows the standard normal distribution.

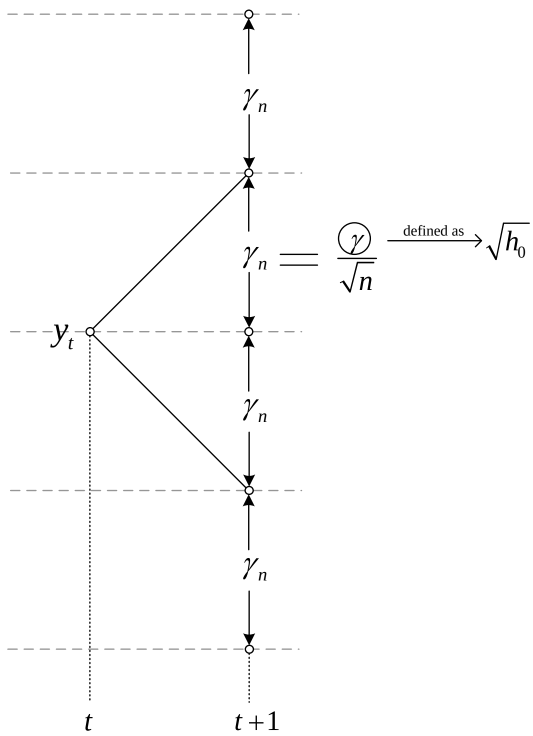
Change variable to y_t by defining $y_t = \ln(S_t)$
$\Rightarrow E_t[y_{t+1}] = y_t + r - \frac{1}{2}h_t$ and $\text{var}(y_{t+1}) = h_t$

(ii) For each Δt, which is assumed to be 1 day in Ritchken and Trevor (1999),

(i) $2n+1$ branches are employed to span the normal distribution for y_{t+1}, e.g., 3 branches are used if $n = 1$. Furthermore, the vertical spacing parameter between node is defined as $\gamma_n \equiv \frac{\gamma}{\sqrt{n}} \equiv \frac{\sqrt{h_0}}{\sqrt{n}}$. The illustration of the case of $n = 1$ is shown in Figure 6-8.

(ii) The value of h_{t+1} is updated at the end of each Δt.

Figure 6-8



(iii) η "multiple-sized" jumps are allowed such that the probabilities of the upward, middle, and downward branches are guaranteed to be in $[0, 1]$. The rule to decide η for any variance h_t is as follows.

$$\eta - 1 < \frac{\sqrt{h_t}}{\gamma} \leq \eta, \text{ where } \gamma \equiv \sqrt{h_0}.$$

(iv) For y_{t+1}, $2n+1$ branches span the corresponding normal distribution, i.e.,

$$y_{t+1} = y_t + \theta\eta\gamma_n \ , \ \theta = 0, \pm 1, \pm 2, ..., \pm n.$$

Therefore, each possible value of y_{t+1} implies a realized value of ε_{t+1} through

$$\varepsilon_{t+1} = \frac{\theta\eta\gamma_n - (r - \frac{h_t}{2})}{\sqrt{h_t}},$$

and thus leads to an update of h_{t+1} as follows.

$$h_{t+1} = \beta_0 + \beta_1 h_t + \beta_2 h_t[\varepsilon_{t+1} - C^*]^2.$$

(v) How to decide the probability for each branch?

(1) Partition Δt (1 day in Ritchken and Trevor (1999)) into n intervals, and the trinomial tree model is employed to model the movements of y for each interval.

(2) Therefore, if $n = 2$, there are $2n+1 = 5$ branches for the period of Δt. Note that although n intervals are considered in the period of Δt, we do not consider the intermediate value of y during Δt, and instead we use directly the $2n+1$ branches for each Δt.

(3) For each of the outgoing $2n+1$ branches of y_t, its probability equals the sum of conditional probabilities of n-interval paths starting from y_t and reaching that node. The details to decide $P(\theta) \equiv \text{Prob}(y_{t+1} = y_t + \theta\eta\gamma_n)$, for $\theta = 0, \pm 1, \pm 2, ..., \pm n$, are as follows.

$$\begin{aligned} p(\theta) &= \sum_{j_u, j_m, j_d} \left(\begin{array}{c} n \\ j_u j_m j_d \end{array} \right) p_u^{j_u} p_m^{j_m} p_d^{j_d} \\ \text{s.t. } & n = j_u + j_m + j_d \\ & \theta = j_u - j_d \end{aligned}$$

where

$$\begin{aligned} p_u &= \frac{h_t}{2\eta^2\gamma^2} + \frac{(r - \frac{h_t}{2})\sqrt{\frac{1}{n}}}{2\eta\gamma}, \\ p_m &= 1 - \frac{h_t}{\eta^2\gamma^2}, \\ p_d &= \frac{h_t}{2\eta^2\gamma^2} - \frac{(r - \frac{h_t}{2})\sqrt{\frac{1}{n}}}{2\eta\gamma}. \end{aligned}$$

(Note that there are typos in the formulas for p_u, p_m, and p_d in Ritchken and Trevor (1999). To correct their formulas for p_u, p_m, and p_d, you need to replace γ_n with γ in their formulas, that yields exactly the same formulas as above.)

$*$ Illustration of the n-subperiod paths to reach $y_{t+1} = y_t + \theta\eta\gamma_n$ for different θ.

If $n = 1$, it is not necessary to make partitions for Δt and the Ritchken and Trevor's model reduces to a trinomial tree model since for each Δt, $2n + 1 = 3$ branches are considered.

For $\theta = 1$, $(j_u, j_m, j_d) = (1, 0, 0)$.
For $\theta = 0$, $(j_u, j_m, j_d) = (0, 1, 0)$.
For $\theta = -1$, $(j_u, j_m, j_d) = (0, 0, 1)$.

If $n = 2$, the Ritchken and Trevor's model becomes a pentanomial tree model since for each Δt, $2n + 1 = 5$ branches are considered.

For $\theta = 2$, $(j_u, j_m, j_d) = (2, 0, 0)$.
For $\theta = 1$, $(j_u, j_m, j_d) = (1, 1, 0)$.
For $\theta = 0$, $(j_u, j_m, j_d) = (0, 2, 0)$ and $(j_u, j_m, j_d) = (1, 0, 1)$.
For $\theta = -1$, $(j_u, j_m, j_d) = (0, 1, 1)$.
For $\theta = -2$, $(j_u, j_m, j_d) = (0, 0, 2)$.

(vi) Forward induction process to build the stock price tree and derive the possible variances reaching each node.

⊙ Note that there could be so many conditional variance levels for each node because different paths reaching that node generates different variances.

⊙ Since the total number of paths is at least 3^N, where $T/N = \Delta t$, in the case of $n = 1$, we can infer that the number of paths grows exponentially.

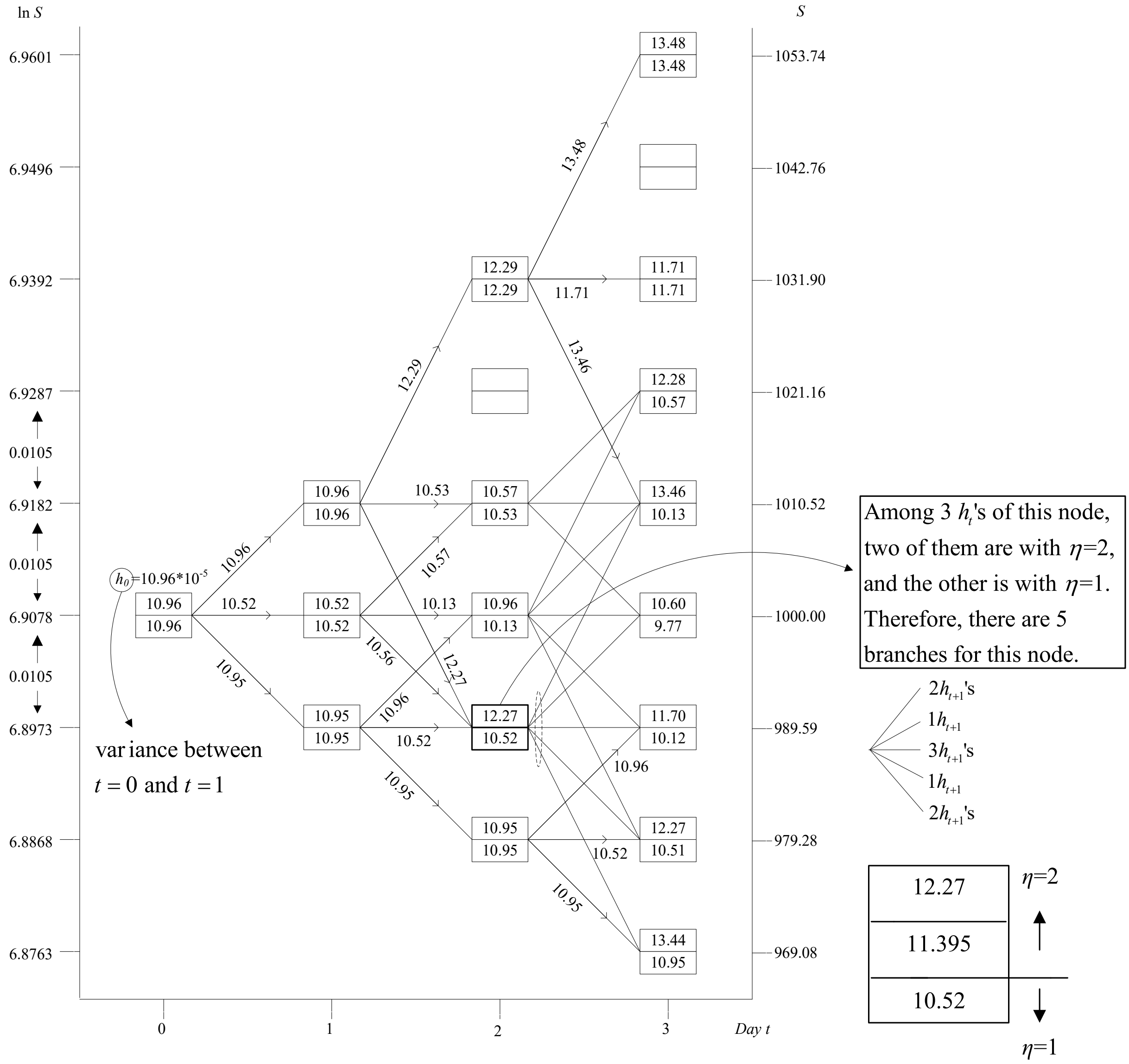
⊙ It is infeasible to record all conditional variances reaching each node due to the availability of memory space in a PC and the concern of the efficiency problem.

⊙ The solution proposed by Ritchken and Trevor (1999):
For each node, record only the maximum and minimum conditional variances among all conditional variances generated by the paths reaching that node. In addition, M interpolated representative conditional variances are equally-spaced placed from the maximum to minimum conditional variances. The table of representative conditional variances are constructed as follows.

$$h(i, j, k) = \frac{M - k}{M - 1} h_{\max}(i, j) + \frac{k - 1}{M - 1} h_{\min}(i, j), \text{ for } k = 1, ..., M,$$

where $h_{\max}(i, j)$ and $h_{\min}(i, j)$ denote the maximum and minimum conditional variances reaching node(i, j).
$*$ Based on these M interpolated representative conditional variances, the conditional variances h_t are updated in the next period of Δt.
$*$ When M approaches infinity, the error caused by the above approximation can be ignored.

Figure 6-9 Lattice Model of Ritchken and Trevor (1999) Over Three Days.



This figures shows the first three days of the first phase of the lattice for an NGARCH model with parameters $r = 0$, $\lambda = 0$, $\beta_0 = 6.575 \times 10^{-6}$, $\beta_1 = 0.90$, $\beta_2 = 0.04$, and $C = 0$. The grid of values for the logarithmic price of the underlying, $y = \ln S$ is determined by taking intervals of size $\gamma = \sqrt{h_0} = 0.0105$ around the log of the initial price $S_0 = 1000$. In this example, n=1, giving three possible paths from each node for a given conditional variance. Each node is represented by a box containing two numbers. The top (bottom) number is the maximum (minimum) conditional variance (multiplied by 10^5) of all paths reaching that node. In this example $M = 3$, so for each node, one additional representative conditional variance is inserted between the maximum and minimum conditional variances. Each of the three variances is examined individually to determine whether the successor nodes are one or more units of γ higher and lower the examined node. The formulas to update the conditional variance are as follows.

$$h_{t+1} = \beta_0 + \beta_1 h_t + \beta_2 h_t(\varepsilon_{t+1} - C^*)^2 \text{ given } h_0 = 0.0001096,$$

$$\text{where } \varepsilon_{t+1} = \frac{j\eta\gamma_n - (r - \frac{h_t}{2})}{\sqrt{h_t}} \text{ and } C^* = C + \lambda.$$

⊙ Backward induction process for option pricing:

Step 1. Decide the payoff of each conditional variance of each terminal node. Since the conditional variance is independent of the option value, the payoffs of different conditional variances on each node are the same. See Figure 6-10.

Step 2. For every conditional variance $h(i,j,k)$ on node(i,j) for $i = N-1, N-2, \ldots, 0$,

(i) Find the evolutions of the conditional variance on the next time step to be

$$h^{next}(\theta) = \beta_0 + \beta_1 h(i,j,k) + \beta_2 h(i,j,k) \left(\frac{\theta\eta\gamma_n - (r - h(i,j,k)/2)}{\sqrt{h(i,j,k)}} - c^* \right)^2,$$

for $\theta = 0, \pm 1, \pm 2, \ldots, \pm n$.

(ii) Suppose that $h^{next}(\theta)$ is inside the range $[h(i+1, j+\theta\eta, k_\theta), h(i+1, j+\theta\eta, k_\theta - 1)]$. By the linear interpolation method, the option value C_θ for the conditional variance $h^{next}(\theta)$ can be approximated as $C_\theta = w_\theta C(i+1, j+\theta\eta, k_\theta) + (1-w_\theta)C(i+1, j+\theta\eta, k_\theta - 1)$, where $w_\theta = (h(i+1, j+\theta\eta, k_\theta - 1) - h^{next}(\theta))/(h(i+1, j+\theta\eta, k_\theta - 1) - h(i+1, j+\theta\eta, k_\theta))$.

(iii) The continuation value for each $h(i,j,k)$ is

$$C(i,j,k) = e^{-r} \sum_{\theta=-n}^{n} p(\theta) C_\theta.$$

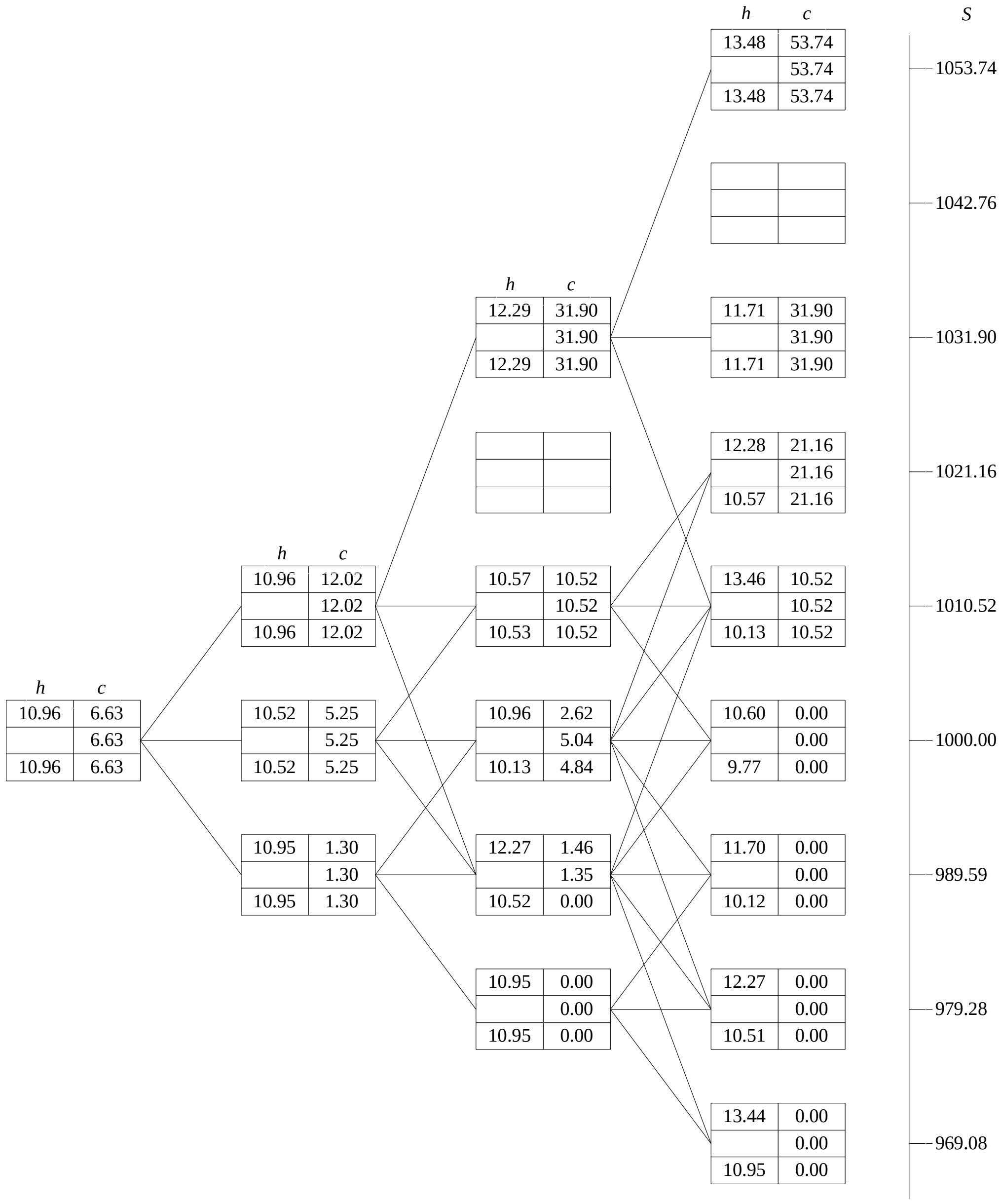
If the feature of early exercise is taken into account, taking vanilla call options as examples, the option value corresponding to $h(i,j,k)$ becomes

$$\max(C(i,j,k), e^{y(i,j)} - K).$$

Step 3. Repeat Step 2 for all $h(i,j,k)$'s backward over the lattice model, the value of $C(0,0,1)$ will be the GARCH option price derived by Ritchken and Trevor (1999).

∗ See Figure 6-10 for a numerical example of the above backward induction process.

Figure 6-10 Lattice for Pricing Three-Period At-The-Money Call Option.



This figures shows the valuation of a three-period at-the-money European call option. Each node is represented by a box containing five numbers. The top (bottom) number in the first column is the maximum (minimum) variance (multiplied by 10^5) of all paths reaching that node. As for the second column, the option values corresponding to different conditional variance levels are reported.

V. Option Pricing under Stochastic Volatility

- The content in this section belongs to the advanced content.

- This section introduces an explicit FDM to price plain vanilla options given stochastic volatility processes for stock prices.

- General setting for stochastic volatility:

$$\frac{dS}{S} = (r-q)dt + \sqrt{V}dZ_S,$$
$$dV = \mu_V dt + \sigma_V dZ_V,$$

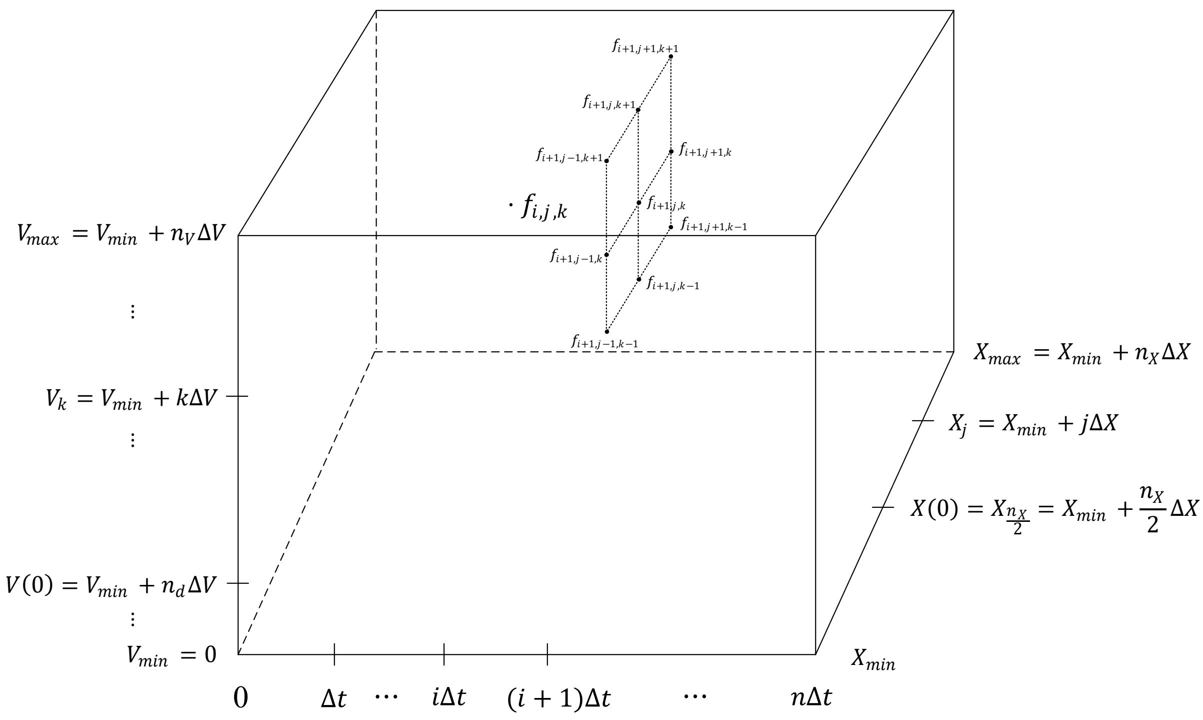
  where $\text{corr}(dZ_S, dZ_V) = \rho_{SV}$. Taking the CEV SV model for example, $\mu_V = \kappa(\theta - V)$ and $\sigma_V = \sigma V^\gamma$. Next, define $X = \ln S$ and $X(0) = \ln S(0)$ and perform Itô's Lemma to yield

$$dX = (r - q - \tfrac{V}{2})dt + \sqrt{V}dZ_S = \mu_X dt + \sigma_X dZ_S.$$

- The PDE for the price of any derivative, $f(X,V)$, which can be expressed as a function of X and V, is derived as follows:

$$\mu_X \frac{\partial f}{\partial X} + \mu_V \frac{\partial f}{\partial V} + \frac{\partial f}{\partial t} + \frac{1}{2}\sigma_X^2 \frac{\partial^2 f}{\partial X^2} + \frac{1}{2}\sigma_V^2 \frac{\partial^2 f}{\partial V^2} + \rho_{SV}\sigma_X\sigma_V \frac{\partial^2 f}{\partial X \partial V} = rf.$$

Figure 6-11 Discretized stock-variance-time space for the FDM.



- Input parameters for the explicit FDM:

  In addition to $S(0)$, K, T, r, q, $V(0)$, other required parameters are summarized as follows.

  1. n (the number of time steps) such that $\Delta t = \frac{T}{n}$,

  2. $V_{min} = 0$, $\bar{V}_{max}$ is suggested to be 1,

  3. n_d (the number of partitioned intervals between $V(0)$ and $V_{min} = 0$), which is used to determine $\Delta V = V(0)/n_d$,

  4. $n_V = \lceil \frac{\bar{V}_{max} - V(0)}{\Delta V} \rceil + n_d$ (the number of partitioned intervals between V_{max} and V_{min}), which is then used to determine $V_{max} = V_{min} + n_V \Delta V$,

  5. α (a multiplicative factor) such that $\Delta X = \sqrt{\alpha V_{max} \Delta t}$,

  (α is suggested to be 3 and cannot be too small for obtaining convergent option values.)

  6. $n_X = 2\lceil \frac{\beta\sqrt{V^* T}}{\Delta X} \rceil$ (the number of partitioned intervals between X_{max} and X_{min}), where $\beta \geq 10$ and V^* is the average of $V(0)$ and the long-term mean of V,

  (Taking the CEV SV model for instance, $V^* = \frac{V(0)+\theta}{2}$.)

  7. $X_{min} = X(0) - \frac{n_X}{2}\Delta X$ and $X_{max} = X(0) + \frac{n_X}{2}\Delta X$.

- Define

  $\mu_X(k) = r - q - \frac{V_k}{2}$,

  $\sigma_X(k) = \sqrt{V_k}$,

  $\mu_V(k) = \kappa(\theta - V_k)$ (taking the CEV SV model for example),

  $\sigma_V(k) = \sigma V_k^{\gamma}$ (taking the CEV SV model for example),

  and the PDE for $f(X, V)$ can be approximated by the explicit FDM as follows.

$$
\begin{aligned}
&\mu_X(k)\frac{f_{i+1,j+1,k} - f_{i+1,j-1,k}}{2\Delta X} + \mu_V(k)\frac{f_{i+1,j,k+1} - f_{i+1,j,k-1}}{2\Delta V} + \frac{f_{i+1,j,k} - f_{i,j,k}}{\Delta t} \\
&+\frac{1}{2}\sigma_X^2(k)\frac{f_{i+1,j+1,k} - 2f_{i+1,j,k} + f_{i+1,j-1,k}}{\Delta X^2} + \frac{1}{2}\sigma_V^2(k)\frac{f_{i+1,j,k+1} - 2f_{i+1,j,k} + f_{i+1,j,k-1}}{\Delta V^2} \\
&+\rho_{SV}\sigma_X(k)\sigma_V(k)\frac{f_{i+1,j+1,k+1} - f_{i+1,j+1,k-1} - f_{i+1,j-1,k+1} + f_{i+1,j-1,k-1}}{4\Delta X \Delta V} \\
&= r f_{i,j,k},
\end{aligned}
$$

  for $i = 0, ..., n-1$, $j = 1, ..., n_X - 1$, and $k = 1, ..., n_V - 1$.

  Therefore, the backward induction method for the normal nodes is expressed as

$$
\begin{aligned}
f_{i,j,k} = {} & A_k f_{i+1,j,k+1} + B_k f_{i+1,j,k} + C_k f_{i+1,j,k-1} + D_k f_{i+1,j-1,k} \\
& + E_k f_{i+1,j+1,k} + F_k(f_{i+1,j+1,k+1} - f_{i+1,j+1,k-1} - f_{i+1,j-1,k+1} + f_{i+1,j-1,k-1}),
\end{aligned}
$$

  where

  $A_k = [\frac{\mu_V(k)}{2\Delta V} + \frac{\sigma_V^2(k)}{2\Delta V^2}]\Delta t/(1 + r\Delta t)$,

$B_k = [1 - (\frac{\sigma_X^2(k)}{\Delta X^2} + \frac{\sigma_V^2(k)}{\Delta V^2})\Delta t]/(1 + r\Delta t),$

$C_k = [-\frac{\mu_V(k)}{2\Delta V} + \frac{\sigma_V^2(k)}{2\Delta V^2}]\Delta t/(1 + r\Delta t),$

$D_k = [-\frac{\mu_X(k)}{2\Delta X} + \frac{\sigma_X^2(k)}{2\Delta X^2}]\Delta t/(1 + r\Delta t),$

$E_k = [\frac{\mu_X(k)}{2\Delta X} + \frac{\sigma_X^2(k)}{2\Delta X^2}]\Delta t/(1 + r\Delta t),$

$F_k = \frac{\rho_{SV}\sigma_X(k)\sigma_V(k)}{4\Delta X\Delta V}\Delta t/(1 + r\Delta t),$

for $i = 0, ..., n-1$, $j = 1, ..., n_X - 1$, and $k = 1, ..., n_V - 1$.

- The remaining tasks are to deal with the boundary conditions for $t = n\Delta t$, $X_j = X_{max}$, $X_j = X_{min}$, $V_j = V_{max}$, $V_j = V_{min}$, which will be discussed as follows individually.

  $*$ Boundary conditions for node(n, j, k) at $t = n\Delta t = T$:

  $f_{n,j,k}$ equals option payoff corresponding to $S(T) = e^{X_j}$,

  for $j = 0, ..., n_X$ and $k = 0, ..., n_V$.

  (For example, $f_{n,j,k} = \max(e^{X_j} - K, 0)$ for calls.)

  $*$ Boundary conditions for node(i, n_X, k) at $X_j = X_{max}$:

  $$f_{i,n_X,k} = \begin{cases} e^{X_{max}} - K & \text{for calls} \\ 0 & \text{for puts} \end{cases},$$

  for $i = 0, ..., n-1$ and $k = 0, ..., n_V$.

  $*$ Boundary conditions for node$(i, 0, k)$ at $X_j = X_{min}$:

  $$f_{i,0,k} = \begin{cases} 0 & \text{for calls} \\ K - e^{X_{min}} & \text{for puts} \end{cases},$$

  for $i = 0, ..., n-1$ and $k = 0, ..., n_V$.

  $*$ Boundary conditions for node(i, j, n_V) at $V_k = V_{max}$:

  Using

  $$\frac{\partial f}{\partial V} = \frac{3f_{i+1,j,n_V} - 4f_{i+1,j,n_V-1} + f_{i+1,j,n_V-2}}{2\Delta V},$$

  $$\frac{\partial^2 f}{\partial V^2} = \frac{f_{i+1,j,n_V} - 2f_{i+1,j,n_V-1} + f_{i+1,j,n_V-2}}{\Delta V^2},$$

  $$\begin{aligned} \frac{\partial^2 f}{\partial X \partial V} &= \frac{\frac{3f_{i+1,j+1,n_V} - 4f_{i+1,j+1,n_V-1} + f_{i+1,j+1,n_V-2}}{2\Delta V} - \frac{3f_{i+1,j-1,n_V} - 4f_{i+1,j-1,n_V-1} + f_{i+1,j-1,n_V-2}}{2\Delta V}}{2\Delta X} \\ &= \frac{3f_{i+1,j+1,n_V} - 4f_{i+1,j+1,n_V-1} + f_{i+1,j+1,n_V-2} - 3f_{i+1,j-1,n_V} + 4f_{i+1,j-1,n_V-1} - f_{i+1,j-1,n_V-2}}{4\Delta X\Delta V}, \end{aligned}$$

  then one can obtain

  $$\mu_X(n_V)\frac{f_{i+1,j+1,n_V} - f_{i+1,j-1,n_V}}{2\Delta X} + \mu_V(n_V)\frac{3f_{i+1,j,n_V} - 4f_{i+1,j,n_V-1} + f_{i+1,j,n_V-2}}{2\Delta V} + \frac{f_{i+1,j,n_V} - f_{i,j,n_V}}{\Delta t}$$

$$+\frac{1}{2}\sigma_X^2(n_V)\frac{f_{i+1,j+1,n_V}-2f_{i+1,j,n_V}+f_{i+1,j-1,n_V}}{\Delta X^2}+\frac{1}{2}\sigma_V^2(n_V)\frac{f_{i+1,j,n_V}-2f_{i+1,j,n_V-1}+f_{i+1,j,n_V-2}}{\Delta V^2}$$
$$+\rho_{SV}\sigma_X(n_V)\sigma_V(n_V)\frac{3f_{i+1,j+1,n_V}-4f_{i+1,j+1,n_V-1}+f_{i+1,j+1,n_V-2}-3f_{i+1,j-1,n_V}+4f_{i+1,j-1,n_V-1}-f_{i+1,j-1,n_V-2}}{4\Delta X\Delta V}$$
$$=rf_{i,j,n_V},$$

for $i=0,...,n-1$ and $j=1,...,n_X-1$.

Therefore, the backward induction method for these boundary nodes is expressed as

$$f_{i,j,n_V}=A_{n_V}f_{i+1,j,n_V}+B_{n_V}f_{i+1,j,n_V-1}+C_{n_V}f_{i+1,j,n_V-2}+D_{n_V}f_{i+1,j-1,n_V}$$
$$+E_{n_V}f_{i+1,j+1,n_V}+F_{n_V}(-4f_{i+1,j+1,n_V-1}+f_{i+1,j+1,n_V-2}+4f_{i+1,j-1,n_V-1}-f_{i+1,j-1,n_V-2}),$$

where

$$A_{n_V}=[1+(\frac{3\mu_V(n_V)}{2\Delta V}-\frac{\sigma_X^2(n_V)}{\Delta X^2}+\frac{\sigma_V^2(n_V)}{2\Delta V^2})\Delta t]/(1+r\Delta t),$$
$$B_{n_V}=[-\frac{2\mu_V(n_V)}{\Delta V}-\frac{\sigma_V^2(n_V)}{\Delta V^2}]\Delta t/(1+r\Delta t),$$
$$C_{n_V}=[\frac{\mu_V(n_V)}{2\Delta V}+\frac{\sigma_V^2(n_V)}{2\Delta V^2}]\Delta t/(1+r\Delta t),$$
$$D_{n_V}=[-\frac{\mu_X(n_V)}{2\Delta X}+\frac{\sigma_X^2(n_V)}{2\Delta X^2}-\frac{3\rho_{SV}\sigma_X(n_V)\sigma_V(n_V)}{4\Delta X\Delta V}]\Delta t/(1+r\Delta t),$$
$$E_{n_V}=[\frac{\mu_X(n_V)}{2\Delta X}+\frac{\sigma_X^2(n_V)}{2\Delta X^2}+\frac{3\rho_{SV}\sigma_X(n_V)\sigma_V(n_V)}{4\Delta X\Delta V}]\Delta t/(1+r\Delta t),$$
$$F_{n_V}=\frac{\rho_{SV}\sigma_X(n_V)\sigma_V(n_V)}{4\Delta X\Delta V}\Delta t/(1+r\Delta t),$$

for $i=0,...,n-1$ and $j=1,...,n_X-1$.

$*$ Boundary conditions for node$(i,j,0)$ at $V_k=V_{min}=0$:

When $k=0$, the PDE is

$$\mu_X(0)\frac{\partial f}{\partial X}+\mu_V(0)\frac{\partial f}{\partial V}+\frac{\partial f}{\partial t}+\frac{1}{2}\sigma_X^2(0)\frac{\partial^2 f}{\partial X^2}+\frac{1}{2}\sigma_V^2(0)\frac{\partial^2 f}{\partial V^2}+\rho_{SV}\sigma_X(0)\sigma_V(0)\frac{\partial^2 f}{\partial X\partial V}=rf.$$

In most situations, $\sigma_X(0)=\sigma_V(0)=0$, and $\mu_X(0)$ becomes a constant which is independent of V, so the PDE is simplified to

$$\mu_X(0)\frac{\partial f}{\partial X}+\mu_V(0)\frac{\partial f}{\partial V}+\frac{\partial f}{\partial t}=rf.$$

(If either $\sigma_X(0)$ or $\sigma_V(0)$ is not zero, a PDE different from above should be derived.)

The above PDE can be approximated with the implicit FDM as follows.

$$\mu_X(0)\frac{f_{i+1,j+1,0}-f_{i+1,j-1,0}}{2\Delta X}+\mu_V(0)\frac{-3f_{i+1,j,0}+4f_{i+1,j,1}-f_{i+1,j,2}}{2\Delta V}+\frac{f_{i+1,j,0}-f_{i,j,0}}{\Delta t}=rf_{i,j,0}.$$

Therefore, the backward induction method for these boundary nodes is expressed as

$$f_{i,j,0}=A_0f_{i+1,j,2}+B_0f_{i+1,j,1}+C_0f_{i+1,j,0}+D_0f_{i+1,j-1,0}+E_0f_{i+1,j+1,0},$$

where

$$A_0=-\frac{\mu_V(0)}{2\Delta V}\Delta t/(1+r\Delta t),$$
$$B_0=\frac{2\mu_V(0)}{\Delta V}\Delta t/(1+r\Delta t),$$
$$C_0=[1-\frac{3\mu_V(0)}{2\Delta V}]\Delta t/(1+r\Delta t),$$

$$D_0 = -\frac{\mu_X(0)}{2\Delta X}\Delta t/(1+r\Delta t),$$
$$E_0 = \frac{\mu_X(0)}{2\Delta X}\Delta t/(1+r\Delta t),$$

for $i = 0, ..., n-1$ and $j = 1, ..., n_X - 1$.