Conditional Variance Models for Price Volatility

- Although a stationary model (see text for definition) has constant variance, its conditional variance may vary.
- Take for example an AR(1) process $X_t = aX_{t-1} + \epsilon_t$ with $|a| < 1$.
  - Here, $\epsilon_t$ is a stationary, uncorrelated process with zero mean and constant variance $\sigma^2$.
- The conditional variance,
  \[ \text{Var}[X_t | X_{t-1}, X_{t-2}, \ldots] \]
equals $\sigma^2$, which is smaller than the unconditional variance $\text{Var}[X_t] = \sigma^2/(1-a^2)$.

Conditional Variance Models for Price Volatility (continued)

- Past information thus has no effects on the variance of prediction.
- To address this drawback, consider models for returns $X_t$ consistent with a changing conditional variance:
  \[ X_t - \mu = V_t U_t. \]
  - $U_t$ has zero mean and unit variance for all $t$.
  - $E[X_t] = \mu$ for all $t$.
  - $\text{Var}[X_t | V_t = v_t] = v_t^2$. 

The historian is a prophet in reverse.
— Friedrich von Schlegel (1772–1829)
Conditional Variance Models for Price Volatility (continued)

• The process \{V_t^2\} models the conditional variance.

• Suppose \{U_t\} and \{V_t\} are independent of each other, which means \{U_1, U_2, \ldots, U_n\} and \{V_1, V_2, \ldots, V_n\} are independent for all \(n\).

• Then \{X_t\} is uncorrelated because
  \[\text{Cov}[X_t, X_{t+\tau}] = 0\]  
  for \(\tau > 0\) (see text for proof).

Conditional Variance Models for Price Volatility (continued)

If, furthermore, \{V_t\} is stationary, then \{X_t\} has constant variance because
  \[E[(X_t - \mu)^2] = E[V_t^2 U_t^2] = E[V_t^2] E[U_t^2] = E[V_t^2].\]

This makes \{X_t\} stationary.

ARCH Models

• An influential model in this direction is the autoregressive conditional heteroskedastic (ARCH) model.

• Assume \(U_t\) is independent of \(V_t, U_{t-1}, V_{t-1}, U_{t-2}, \ldots\) for all \(t\).

• Consequently \{\(X_t\)\} is uncorrelated by Eq. (77) on p. 712.

• Assume furthermore that \{\(U_t\)\} is a Gaussian stationary, uncorrelated process.

• Then \(X_t \mid I_{t-1} \sim N(\mu, V_t^2)\).

\(^a\)Engle (1982), co-winner of the 2003 Nobel Prize in Economic Sciences.
ARCH Models (continued)

- The ARCH($p$) process is defined by
  \[ X_t - \mu = \left( a_0 + \sum_{i=1}^{p} a_i (X_{t-i} - \mu)^2 \right)^{1/2} U_t, \]
  where $a_1, \ldots, a_p \geq 0$ and $a_0 > 0$.
- The variance $V_t^2$ thus satisfies
  \[ V_t^2 = a_0 + \sum_{i=1}^{p} a_i (X_{t-i} - \mu)^2. \]
- The volatility at time $t$ as estimated at time $t-1$ depends on the $p$ most recent observations on squared returns.

GARCH Models

- A very popular extension of the ARCH model is the generalized autoregressive conditional heteroskedastic (GARCH) process.
- The simplest GARCH(1,1) process adds $a_2 V_{t-1}^2$ to the ARCH(1) process, resulting in
  \[ V_t^2 = a_0 + a_1 (X_{t-1} - \mu)^2 + a_2 V_{t-1}^2. \]
- The volatility at time $t$ as estimated at time $t-1$ depends on the squared return and the estimated volatility at time $t-1$.  
  
  \(^*\)Bollerslev (1986) and Taylor (1986).

ARCH Models (concluded)

- The ARCH(1) process
  \[ X_t - \mu = (a_0 + a_1 (X_{t-1} - \mu)^2)^{1/2} U_t \]
  is the simplest.
- For it,
  \[ \text{Var}[X_t | X_{t-1} = x_{t-1}] = a_0 + a_1 (x_{t-1} - \mu)^2. \]
- The process \{ $X_t$ \} is stationary with finite variance if and only if $a_1 < 1$, in which case $\text{Var}[X_t] = a_0/(1 - a_1)$.
- The parameters can be estimated by statistical techniques.

GARCH Models (concluded)

- The estimate of volatility averages past squared returns by giving heavier weights to recent squared returns (see text).
- It is usually assumed that $a_1 + a_2 < 1$ and $a_0 > 0$, in which case the unconditional, long-run variance is given by $a_0/(1 - a_1 - a_2)$.
- A popular special case of GARCH(1,1) is the exponentially weighted moving average process, which sets $a_0$ to zero and $a_2$ to $1 - a_1$.
- This model is used in J.P. Morgan’s RiskMetrics™.
GARCH Option Pricing

• Options can be priced when the underlying asset’s return follows a GARCH process.
• Let $S_t$ denote the asset price at date $t$.
• Let $h_t^2$ be the conditional variance of the return over the period $[t, t+1]$ given the information at date $t$.
  - “One day” is merely a convenient term for any elapsed time $\Delta t$.

GARCH Option Pricing (continued)

• Adopt the following risk-neutral process for the price dynamics:
  
  $$\ln \frac{S_{t+1}}{S_t} = r - \frac{h_t^2}{2} + h_t \epsilon_{t+1},$$

  where

  $$h_{t+1}^2 = \beta_0 + \beta_1 h_t^2 + \beta_2 h_t^2 (\epsilon_{t+1} - c)^2,$$

  $$\epsilon_{t+1} \sim N(0, 1) \text{ given information at date } t,$$

  $$r = \text{ daily riskless return},$$

  $$c \geq 0.$$  

  \(a\)Duan (1995).

GARCH Option Pricing (continued)

• The five unknown parameters of the model are $c, h_0, \beta_0, \beta_1$, and $\beta_2$.
• It is postulated that $\beta_0, \beta_1, \beta_2 \geq 0$ to make the conditional variance positive.
• The above process, called the nonlinear asymmetric GARCH model, generalizes the GARCH(1, 1) model (see text).

GARCH Option Pricing (concluded)

• With $y_t \equiv \ln S_t$ denoting the logarithmic price, the model becomes
  
  $$y_{t+1} = y_t + r - \frac{h_t^2}{2} + h_t \epsilon_{t+1}.$$  

  (80)

• The pair $(y_t, h_t^2)$ completely describes the current state.
• The conditional mean and variance of $y_{t+1}$ are clearly

  $$E[y_{t+1} | y_t, h_t^2] = y_t + r - \frac{h_t^2}{2},$$

  (81)

  $$\text{Var}[y_{t+1} | y_t, h_t^2] = h_t^2.$$  

  (82)
The Ritchken-Trevor (RT) Algorithm

- The GARCH model is a continuous-state model.
- To approximate it, we turn to trees with discrete states.
- Path dependence in GARCH makes the tree for asset prices explode exponentially.
- We need to mitigate this combinatorial explosion somewhat.

*Ritchken and Trevor (1999).*

The Ritchken-Trevor Algorithm (continued)

- It remains to pick the jump size and the three branching probabilities.
- The role of $\sigma$ in the Black-Scholes option pricing model is played by $h_t$ in the GARCH model.
- As a jump size proportional to $\sigma/\sqrt{n}$ is picked in the BOPM, a comparable magnitude will be chosen here.
- Define $\gamma \equiv h_0$, though other multiples of $h_0$ are possible, and $\gamma_n \equiv \frac{\gamma}{\sqrt{n}}$.
- The jump size will be some integer multiple $\eta$ of $\gamma_n$.
- We call $\eta$ the jump parameter (p. 727).

Partition a day into $n$ periods.
- Three states follow each state $(y_t, h_t^2)$ after a period.
- As the trinomial model combines, $2n + 1$ states at date $t + 1$ follow each state at date $t$ (recall p. 550).
- These $2n + 1$ values must approximate the distribution of $(y_{t+1}, h_{t+1}^2)$.
- So the conditional moments (81)–(82) at date $t + 1$ on p. 723 must be matched by the trinomial model to guarantee convergence to the continuous-state model.

The seven values on the right approximate the distribution of logarithmic price $y_{t+1}$. 

(0, 0)  (1, 0)  (1, −1)

$\eta \gamma_n$

(1, 1)
The Ritchken-Trevor Algorithm (continued)

- The middle branch does not change the underlying asset’s price.

- The probabilities for the up, middle, and down branches are

\[
pu = \frac{h_t^2}{2\eta^2\gamma^2} + \frac{r - (h_t^2/2)}{2\eta\gamma\sqrt{n}},
\]

\[
p_m = 1 - \frac{h_t^2}{\eta^2\gamma^2},
\]

\[
p_d = \frac{h_t^2}{2\eta^2\gamma^2} - \frac{r - (h_t^2/2)}{2\eta\gamma\sqrt{n}}.
\]

---

The Ritchken-Trevor Algorithm (continued)

- We can dispense with the intermediate nodes between dates to create a \((2n + 1)\)-nomial tree (p. 731).

- The resulting model is multinomial with \(2n + 1\) branches from any state \((y_t, h_t^2)\).

- There are two reasons behind this manipulation.
  - Interdate nodes are created merely to approximate the continuous-state model after one day.
  - Keeping the interdate nodes results in a tree that is \(n\) times as large.

---

The Ritchken-Trevor Algorithm (continued)

- It can be shown that:
  - The trinomial model takes on \(2n + 1\) values at date \(t + 1\) for \(y_{t+1}\).
  - These values have a matching mean for \(y_{t+1}\).
  - These values have an asymptotically matching variance for \(y_{t+1}\).

- The central limit theorem thus guarantees the desired convergence as \(n\) increases.

---

This heptanomial tree is the outcome of the trinomial tree on p. 727 after its intermediate nodes are removed.
The Ritchken-Trevor Algorithm (continued)

- A node with logarithmic price $y_t + \ell \eta n$ at date $t + 1$ follows the current node at date $t$ with price $y_t$ for some $-n \leq \ell \leq n$.
- To reach that price in $n$ periods, the number of up moves must exceed that of down moves by exactly $\ell$.
- The probability that this happens is
  
  $$P(\ell) = \sum_{j_u,j_m,j_d} \frac{n!}{j_u! j_m! j_d!} p_u^{j_u} p_m^{j_m} p_d^{j_d},$$

  with $j_u, j_m, j_d \geq 0$, $n = j_u + j_m + j_d$, and $\ell = j_u - j_d$.

The Ritchken-Trevor Algorithm (continued)

- The updating rule (79) on p. 721 must be modified to account for the adoption of the discrete-state model.
- The logarithmic price $y_t + \ell \eta n$ at date $t + 1$ following state $(y_t, h_{2t})$ at date $t$ has a variance equal to
  
  $$h_{2t+1}^2 = \beta_0 + \beta_1 h_t^2 + \beta_2 h_t^2 (\epsilon_{t+1} - c)^2,$$

  where
  
  $$\epsilon_{t+1} = \frac{\ell \eta n - (r - h_t^2 / 2)}{h_t}, \quad \ell = 0, \pm 1, \pm 2, \ldots, \pm n,$$

  is a discrete random variable with $2n + 1$ values.

The Ritchken-Trevor Algorithm (continued)

- Different conditional variances $h_t^2$ may require different $\eta$ so that the probabilities calculated by Eqs. (83)–(85) on p. 728 lie between 0 and 1.
- This implies varying jump sizes.
- The necessary requirement $p_m \geq 0$ implies $\eta \geq h_t / \gamma$.
- Hence we try
  
  $$\eta = \lfloor h_t / \gamma \rfloor, \lfloor h_t / \gamma \rfloor + 1, \lfloor h_t / \gamma \rfloor + 2, \ldots,$$

  until valid probabilities are obtained or until their nonexistence is confirmed.
The Ritchken-Trevor Algorithm (continued)

- The sufficient and necessary condition for valid probabilities to exist is
  \[
  \left| r - \left( \frac{h^2_t}{2} \right) \right| \leq \frac{h^2_t}{2\eta^2 \gamma^2} \leq \min \left( 1 - \left| r - \left( \frac{h^2_t}{2} \right) \right| , \frac{1}{2} \right).
  \]
- Obviously, the magnitude of \( \eta \) tends to grow with \( h_t \).
- The plot on p. 737 uses \( n = 1 \) to illustrate our points for a 3-day model.
- For example, node (1,1) of date 1 and node (2,3) of date 2 pick \( \eta = 2 \).

The Ritchken-Trevor Algorithm (continued)

- The topology of the tree is not a standard combining multinomial tree.
- For example, a few nodes on p. 737 such as nodes (2,0) and (2,-1) have multiple jump sizes.
- The reason is the path dependence of the model.
  - Two paths can reach node (2,0) from the root node, each with a different variance for the node.
  - One of the variances results in \( \eta = 1 \), whereas the other results in \( \eta = 2 \).

The Ritchken-Trevor Algorithm (concluded)

- The possible values of \( h^2_t \) at a node are exponential nature.
- To address this problem, we record only the maximum and minimum \( h^2_t \) at each node.a
- Therefore, each node on the tree contains only two states \( (y_t, h^2_{\text{max}}) \) and \( (y_t, h^2_{\text{min}}) \).
- Each of \( (y_t, h^2_{\text{max}}) \) and \( (y_t, h^2_{\text{min}}) \) carries its own \( \eta \) and set of \( 2n + 1 \) branching probabilities.

---

aÇakici and Topyan (2000).
Negative Aspects of the Ritchken-Trevor Algorithm\textsuperscript{a}

- A small $n$ may yield inaccurate option prices.
- But the tree will grow exponentially if $n$ is large enough.
  - Specifically, $n > (1 - \beta_1)/\beta_2$ when $r = c = 0$.
- A large $n$ has another serious problem: The tree cannot grow beyond a certain date.
- Thus the choice of $n$ may be limited in practice.
- The RT algorithm can be modified to be free of exponential complexity and shortened maturity.\textsuperscript{b}

\textsuperscript{a}Lyuu and Wu (2003).
\textsuperscript{b}Lyuu and Wu (2005).

Numerical Examples (continued)

- Let $h^2_{\text{max}}(i, j)$ denote the maximum variance at node $(i, j)$.
- Let $h^2_{\text{min}}(i, j)$ denote the minimum variance at node $(i, j)$.
- Initially, $h^2_{\text{max}}(0, 0) = h^2_{\text{min}}(0, 0) = h^2_0$.
- The resulting three-day tree is depicted on p. 743.

Numerical Examples

- Assume $S_0 = 100$, $y_0 = \ln S_0 = 4.60517$, $r = 0$,
  $h^2_0 = 0.0001096$, $\gamma = h_0 = 0.010469$, $n = 1$,
  $\gamma_n = \gamma/\sqrt{n} = 0.010469$, $\beta_0 = 0.000006575$, $\beta_1 = 0.9$,
  $\beta_2 = 0.04$, and $c = 0$.
- A daily variance of 0.0001096 corresponds to an annual volatility of $\sqrt{365 \times 0.0001096} \approx 20\%$.
- Let $h^2(i, j)$ denote the variance at node $(i, j)$.
- Initially, $h^2(0, 0) = h^2_0 = 0.0001096$. 

\begin{center}
\begin{tabular}{|c|c|}
\hline
$i$  & $h^2(i, j)$ \\
\hline
0  & 0.0001096 \\
1  & 0.0001096 \\
2  & 0.0001096 \\
\hline
\end{tabular}
\end{center}
A top (bottom) number inside a gray box refers to the minimum (maximum, respectively) variance $h_{\text{min}}^2$ ($h_{\text{max}}^2$, respectively) for the node. Variances are multiplied by 100,000 for readability. A top (bottom) number inside a white box refers to $\eta$ corresponding to $h_{\text{min}}^2$ ($h_{\text{max}}^2$, respectively).

### Numerical Examples (continued)

- Let us see how the numbers are calculated.
- Start with the root node, node (0, 0).
- Try $\eta = 1$ in Eqs. (83)–(85) on p. 728 first to obtain
  \[
  p_u = 0.4974, \\
  p_m = 0, \\
  p_d = 0.5026.
  \]
- As they are valid probabilities, the three branches from the root node use single jumps.

- Move on to node (1, 1).
- It has one predecessor node—node (0, 0)—and it takes an up move to reach the current node.
- So apply updating rule (87) on p. 734 with $\ell = 1$ and $h_t^2 = h^2(0, 0)$.
- The result is $h^2(1, 1) = 0.000109645$.

### Numerical Examples (continued)

- Because $\lfloor h(1, 1)/\gamma \rfloor = 2$, we try $\eta = 2$ in Eqs. (83)–(85) on p. 728 first to obtain
  \[
  p_u = 0.1237, \\
  p_m = 0.7499, \\
  p_d = 0.1264.
  \]
- As they are valid probabilities, the three branches from node (1, 1) use double jumps.
Numerical Examples (continued)

• Carry out similar calculations for node \((1, 0)\) with \(\ell = 0\) in updating rule (87) on p. 734.
• Carry out similar calculations for node \((1, -1)\) with \(\ell = -1\) in updating rule (87).
• Single jump \(\eta = 1\) works in both nodes.
• The resulting variances are
  \[
  h^2(1, 0) = 0.000105215, \\
  h^2(1, -1) = 0.000109553.
  \]

Numerical Examples (continued)

• Now move on to the other predecessor node \((1, -1)\).
• Because it takes an up move to reach the current node, apply updating rule (87) on p. 734 with \(\ell = 1\) and \(h^2 = h^2(1, -1)\).
• The result is \(h_{t+1}^2 = 0.000109603\).
• We hence record
  \[
  h_{\text{min}}^2(2, 0) = 0.000101269, \\
  h_{\text{max}}^2(2, 0) = 0.000109603.
  \]

Numerical Examples (continued)

• Node \((2, 0)\) has 2 predecessor nodes, \((1, 0)\) and \((1, -1)\).
• Both have to be considered in deriving the variances.
• Let us start with node \((1, 0)\).
• Because it takes a middle move to reach the current node, we apply updating rule (87) on p. 734 with \(\ell = 0\) and \(h^2 = h^2(1, 0)\).
• The result is \(h_{t+1}^2 = 0.000101269\).

Numerical Examples (continued)

• Consider state \(h_{\text{max}}^2(2, 0)\) first.
• Because \(\lfloor h_{\text{max}}(2, 0)/\gamma \rfloor = 2\), we first try \(\eta = 2\) in Eqs. (83)–(85) on p. 728 to obtain
  \[
  p_u = 0.1237, \\
  p_m = 0.7500, \\
  p_d = 0.1263.
  \]
• As they are valid probabilities, the three branches from node \((2, 0)\) with the maximum variance use double jumps.
Numerical Examples (continued)

- Now consider state $h_{\text{min}}^2(2,0)$.
- Because $\lfloor h_{\text{min}}(2,0)/\gamma \rfloor = 1$, we first try $\eta = 1$ in Eqs. (83)–(85) on p. 728 to obtain

\begin{align*}
  p_u &= 0.4596, \\
  p_m &= 0.0760, \\
  p_d &= 0.4644.
\end{align*}

- As they are valid probabilities, the three branches from node $(2,0)$ with the minimum variance use single jumps.

Numerical Examples (continued)

- Now move on to predecessor node $(1,0)$.
- Because it also takes a down move to reach the current node, we apply updating rule (87) on p. 734 with $\ell = -1$ and $h_t^2 = h^2(1,0)$.
- The result is $h_{t+1}^2 = 0.000105609$.

Numerical Examples (continued)

- Finally, consider predecessor node $(1,-1)$.
- Because it takes a middle move to reach the current node, we apply updating rule (87) on p. 734 with $\ell = 0$ and $h_t^2 = h^2(1,-1)$.
- The result is $h_{t+1}^2 = 0.000105173$.
- We hence record

\begin{align*}
  h_{\text{min}}^2(2,-1) &= 0.000105173, \\
  h_{\text{max}}^2(2,-1) &= 0.0001227.
\end{align*}
Numerical Examples (continued)

• Consider state $h_{\text{max}}^2(2, -1)$.

• Because $\lceil h_{\text{max}}(2, -1)/\gamma \rceil = 2$, we first try $\eta = 2$ in Eqs. (83)–(85) on p. 728 to obtain

$$p_u = 0.1385,$$
$$p_m = 0.7201,$$
$$p_d = 0.1414.$$ 

• As they are valid probabilities, the three branches from node $(2, -1)$ with the maximum variance use double jumps.

Numerical Examples (continued)

• Next, consider state $h_{\text{min}}^2(2, -1)$.

• Because $\lceil h_{\text{min}}(2, -1)/\gamma \rceil = 1$, we first try $\eta = 1$ in Eqs. (83)–(85) on p. 728 to obtain

$$p_u = 0.4773,$$
$$p_m = 0.0404,$$
$$p_d = 0.4823.$$ 

• As they are valid probabilities, the three branches from node $(2, -1)$ with the minimum variance use single jumps.

Negative Aspects of the RT Algorithm Revisited

• Other nodes at dates 2 and 3 can be handled similarly.

• In general, if a node has $k$ predecessor nodes, then $2^k$ variances will be calculated using the updating rule.
  – This is because each predecessor node keeps two variance numbers.

• But only the maximum and minimum variances will be kept.

• Recall the problems mentioned on p. 740.

• In our case, combinatorial explosion occurs when

$$n > \frac{1 - \beta_1}{\beta_2} = \frac{1 - 0.9}{0.04} = 2.5.$$ 

• Suppose we are willing to accept the exponential running time and pick $n = 100$ to seek accuracy.

• But the problem of shortened maturity forces the tree to stop at date 9!

\textsuperscript{a}Lyuu and Wu (2003).
Backward Induction on the RT Tree

- After the RT tree is constructed, it can be used to price options by backward induction.
- Recall that each node keeps two variances $h_{\text{max}}^2$ and $h_{\text{min}}^2$.
- We now increase that number to $K$ equally spaced variances between $h_{\text{max}}^2$ and $h_{\text{min}}^2$ at each node.
- Besides the minimum and maximum variances, the other $K - 2$ variances in between are linearly interpolated.\(^a\)

\(^a\)In practice, log-linear interpolation works better; Lyuu and Wu (2005). Log-cubic interpolation works even better; Liu (2005).

Backward Induction on the RT Tree (continued)

- For example, if $K = 3$, then a variance of $10.5436 \times 10^{-6}$ will be added between the maximum and minimum variances at node $(2,0)$ on p. 743.
- In general, the $k$th variance at node $(i,j)$ is

\[
    h_{\text{min}}^2(i,j) + k \frac{h_{\text{max}}^2(i,j) - h_{\text{min}}^2(i,j)}{K - 1},
\]

$k = 0, 1, \ldots, K - 1$.
- Each interpolated variance’s jump parameter and branching probabilities can be computed as before.

Backward Induction on the RT Tree (concluded)

- During backward induction, if a variance falls between two of the $K$ variances, linear interpolation of the option prices corresponding to the two bracketing variances will be used as the approximate option price.
- The above ideas are reminiscent of the ones on p. 319, where we dealt with arithmetic average-rate options.
Numerical Examples

- We next use the numerical example on p. 743 to price a European call option with a strike price of 100 and expiring at date 3.
- Recall that the riskless interest rate is zero.
- Assume $K = 2$; hence there are no interpolated variances.
- The pricing tree is shown on p. 765 with a call price of $0.66346$.
  - The branching probabilities needed in backward induction can be found on p. 766.

Numerical Examples (continued)

- Let us derive some of the numbers on p. 765.
- The option price for a terminal node at date 3 equals $\max(S_3 - 100, 0)$, independent of the variance level.
- Now move on to nodes at date 2.
- The option price at node $(2, 3)$ depends on those at nodes $(3, 5), (3, 3), \text{ and } (3, 1)$.
  - It therefore equals $0.1387 \times 5.37392 + 0.7197 \times 3.19054 + 0.1416 \times 1.05240 = 3.19054$.
- Option prices for other nodes at date 2 can be computed similarly.
• For node $(1, 1)$, the option price for both variances is

$$0.1237 \times 3.19054 + 0.7499 \times 1.05240 + 0.1264 \times 0.14573 = 1.20241. \quad \star \star \star$$

• Node $(1, 0)$ is most interesting.
• We knew that a down move from it gives a variance of $0.000105609$.
• This number falls between the minimum variance $0.000105173$ and the maximum variance $0.0001227$ at node $(2, -1)$ on p. 743.

• The option price corresponding to the minimum variance is 0.
• The option price corresponding to the maximum variance is $0.14573$.
• The equation

$$x \times 0.000105173 + (1 - x) \times 0.0001227 = 0.000105609$$

is satisfied by $x = 0.9751$.
• So the option for the down state is approximated by

$$x \times 0 + (1 - x) \times 0.14573 = 0.00362.$$