Matrix Computation

Definitions and Basic Results

- Let $A \equiv [\, a_{ij}\,]_{1 \leq i \leq m, 1 \leq j \leq n}$, or simply $A \in \boldsymbol{R}^{m \times n}$, denote an $m \times n$ matrix.
- It can also be represented as $[\, a_1, a_2, \dots, a_n\,]$ where $a_i \in \boldsymbol{R}^m$ are vectors.
  - Vectors are column vectors unless stated otherwise.
- A is a square matrix when $m = n$.
- The rank of a matrix is the largest number of linearly independent columns.
- An $m \times n$ matrix is rank deficient if its rank is less than $\min(m, n)$; otherwise, it has full rank.

Definitions and Basic Results (continued)

- A square matrix A is said to be symmetric if $A^{\mathrm{T}} = A$.
  - Such matrices are nonsingular.
- A diagonal $m \times n$ matrix $D \equiv [\, d_{ij}\,]_{i,j}$ may be denoted by $\mathrm{diag}[\, D_1, D_2, \dots, D_q\,]$, where $q \equiv \min(m, n)$ and $D_i = d_{ii}$ for $1 \leq i \leq q$.
- The identity matrix is the square matrix

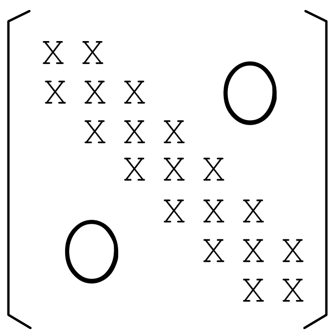
$$I \equiv \mathrm{diag}[\, 1, 1, \dots, 1\,].$$

Diagonal Matrices

$$\begin{bmatrix} \times & 0 & 0 & 0 & 0 \\ 0 & \times & 0 & 0 & 0 \\ 0 & 0 & \times & 0 & 0 \end{bmatrix} \quad \begin{bmatrix} \times & 0 & 0 \\ 0 & \times & 0 \\ 0 & 0 & \times \end{bmatrix} \quad \begin{bmatrix} \times & 0 & 0 \\ 0 & \times & 0 \\ 0 & 0 & \times \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Definitions and Basic Results (concluded)

- A matrix has full column rank if its columns are linearly independent.
- A real symmetric matrix A is positive definite if $x^{\mathrm{T}} A x = \sum_{i,j} a_{ij} x_i x_j > 0$ for any nonzero vector x.
- It is known that a matrix A is positive definite if and only if there exists a matrix W such that $A = W^{\mathrm{T}} W$ and W has full column rank.



Decompositions

- Positive definite matrices can be factored as

$$A = LL^{\mathrm{T}},$$

called the Cholesky decomposition.

Orthogonal and Orthonormal Matrices

- A vector set $\{ x_1, x_2, \ldots, x_p \}$ is orthogonal if all its vectors are nonzero and the inner products $x_i^{\mathrm{T}} x_j$ equal zero for $i \neq j$.
- It is orthonormal if, furthermore,

$$x_i^{\mathrm{T}} x_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

- A real square matrix Q is orthogonal if $Q^{\mathrm{T}} Q = I$.
- For such matrices, $Q^{-1} = Q^{\mathrm{T}}$ and $QQ^{\mathrm{T}} = I$.

Eigenvalues and Eigenvectors

- An eigenvalue of a square matrix A is a complex number λ such that $Ax = \lambda x$ for some nonzero vector x called an eigenvector.
- The eigenvalues for a real symmetric matrix are real.
- For them, the Schur decomposition theorem[a] says that there exists a real orthogonal matrix Q such that $Q^{\mathrm{T}} A Q = \mathrm{diag}[\lambda_1, \lambda_2, \dots, \lambda_n]$.
  - Q's ith column is the eigenvector corresponding to λ_i, and the eigenvectors form an orthonormal set.
- The eigenvalues of positive definite matrices are positive.

[a]Also called the principal-axes theorem or the spectral theorem.

Generation of Multivariate Normal Distribution

- Let $\boldsymbol{x} \equiv [x_1, x_2, \dots, x_n]^{\mathrm{T}}$ be a vector random variable with a positive definite covariance matrix C.
- As usual, assume $E[\boldsymbol{x}] = \mathbf{0}$.
- This distribution can be generated by $P\boldsymbol{y}$.
  - $C = PP^{\mathrm{T}}$ is the Cholesky decomposition of C.
  - $\boldsymbol{y} \equiv [y_1, y_2, \dots, y_n]^{\mathrm{T}}$ is a vector random variable with a covariance matrix equal to the identity matrix.
- Reason (see text):

$$\mathrm{Cov}[P\boldsymbol{y}] = P\,\mathrm{Cov}[\boldsymbol{y}]\,P^{\mathrm{T}} = PP^{\mathrm{T}} = C.$$

Generation of Multivariate Normal Distribution (concluded)

- Suppose we want to generate the multivariate normal distribution with a covariance matrix $C = PP^{\mathrm{T}}$.
- We start with independent standard normal distributions $y_1, y_2, \dots, y_n$.
- Then $P[y_1, y_2, \dots, y_n]^{\mathrm{T}}$ has the desired distribution.

Multivariate Derivatives Pricing

- Generating the multivariate normal distribution is essential for the Monte Carlo pricing of multivariate derivatives.
- The rainbow option on k assets has payoff

$$\max(\max(S_1, S_2, \dots, S_k) - X, 0)$$

at maturity.

- The closed-form formula is a multi-dimensional integral.[a]

[a]Johnson (1987).

Time Series Analysis

The historian is a prophet in reverse.
— Friedrich von Schlegel (1772–1829)

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, ϵ_t is a stationary, uncorrelated process with zero mean and constant variance σ^2.
- The conditional variance,

$$\text{Var}[\,X_t \,|\, X_{t-1}, X_{t-2}, \ldots\,],$$

equals σ^2, which is smaller than the *unconditional* variance $\text{Var}[\,X_t\,] = \sigma^2/(1-a^2)$.

Conditional Variance Models for Price Volatility (continued)

- In the Black-Scholes model, past information 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.
  - $\text{Var}[\,X_t \,|\, V_t = v_t\,] = v_t^2$.

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

$$\text{Cov}[\, X_t, X_{t+\tau} \,] = 0 \tag{74}$$

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

$$\begin{aligned} & E\left[\, (X_t - \mu)^2 \,\right] \\ = \; & E\left[\, V_t^2 U_t^2 \,\right] \\ = \; & E\left[\, V_t^2 \,\right] E\left[\, U_t^2 \,\right] \\ = \; & E\left[\, V_t^2 \,\right]. \end{aligned}$$

- This makes $\{ X_t \}$ stationary.

Conditional Variance Models for Price Volatility (concluded)

- Suppose we assume that conditional variances are deterministic functions of past returns:

$$V_t = f(X_{t-1}, X_{t-2}, \ldots)$$

for some function f.

- Then V_t can be computed given the information set of past returns:

$$I_{t-1} \equiv \{ X_{t-1}, X_{t-2}, \ldots \}.$$

ARCH Models[a]

- 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. (74) on p. 641.
- Assume furthermore that $\{ U_t \}$ is a Gaussian stationary, uncorrelated process.
- Then $X_t \,|\, I_{t-1} \sim N(\mu, V_t^2)$.

[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.

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 $\text{Var}[\, X_t \,] = a_0/(1 - a_1)$.

GARCH Models[a]

- 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 $t-1$.

[a] Bollerslev (1986) and Taylor (1986).

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 Δt.

GARCH Option Pricing (continued)

- Adopt the following risk-neutral process for the price dynamics:[a]

$$\ln \frac{S_{t+1}}{S_t} = r - \frac{h_t^2}{2} + h_t \epsilon_{t+1},$$

where

$$\begin{aligned} h_{t+1}^2 &= \beta_0 + \beta_1 h_t^2 + \beta_2 h_t^2 (\epsilon_{t+1} - c)^2, & (75)\\ \epsilon_{t+1} &\sim N(0,1) \text{ given information at date } t, \\ r &= \text{daily riskless return}, \\ c &\geq 0. \end{aligned}$$

[a]Duan (1995).

GARCH Option Pricing (continued)

- The five unknown parameters of the model are c, h_0, β_0, β_1, and β_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

$$y_{t+1} = y_t + r - \frac{h_t^2}{2} + h_t \epsilon_{t+1}.$$

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

$$\begin{aligned} E[\,y_{t+1} \,|\, y_t, h_t^2\,] &= y_t + r - \frac{h_t^2}{2}, & (76)\\ \text{Var}[\,y_{t+1} \,|\, y_t, h_t^2\,] &= h_t^2. & (77) \end{aligned}$$

The Ritchken-Trevor (RT) Algorithm[a]

- 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.

[a]Ritchken and Trevor (1999).

The Ritchken-Trevor Algorithm (continued)

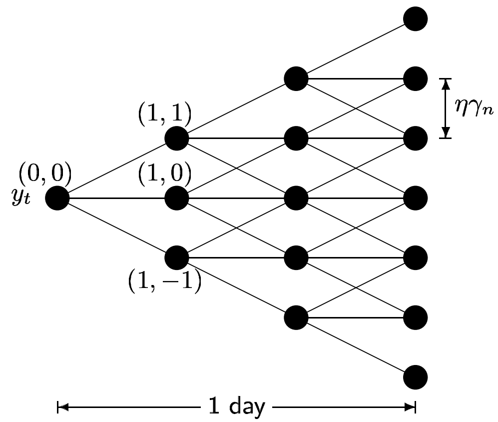
- Partition a day into n periods.
- Three states follow each state (y_t, h_t^2) after a period.
- As the trinomial model combines, $2n+1$ states at date $t+1$ follow each state at date t (recall p. 550).
- These $2n+1$ values must approximate the distribution of (y_{t+1}, h_{t+1}^2).
- So the conditional moments (76)–(77) at date $t+1$ on p. 652 must be matched by the trinomial model to guarantee convergence to the continuous-state model.

The Ritchken-Trevor Algorithm (continued)

- It remains to pick the jump size and the three branching probabilities.
- The role of σ 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 η of γ_n.
- We call η the jump parameter (see p. 656).



The seven values on the right approximate the distribution of logarithmic price y_{t+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

$$p_u = \frac{h_t^2}{2\eta^2\gamma^2} + \frac{r - (h_t^2/2)}{2\eta\gamma\sqrt{n}}, \tag{78}$$

$$p_m = 1 - \frac{h_t^2}{\eta^2\gamma^2}, \tag{79}$$

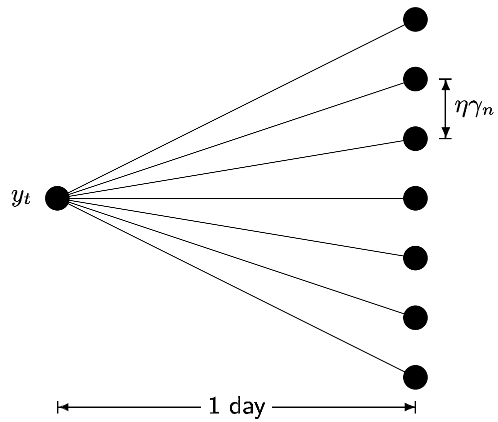
$$p_d = \frac{h_t^2}{2\eta^2\gamma^2} - \frac{r - (h_t^2/2)}{2\eta\gamma\sqrt{n}}. \tag{80}$$

The Ritchken-Trevor Algorithm (continued)

- It can be shown that the trinomial model takes on $2n+1$ values at date $t+1$ with a matching mean and variance for y_{t+1}.
- The central limit theorem thus guarantees the desired convergence as n increases.

The Ritchken-Trevor Algorithm (continued)

- We can dispense with the intermediate nodes *between* dates to create a $(2n+1)$-nomial tree (see p. 660).
- 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.



This heptanomial tree is the outcome of the trinomial tree on p. 656 after its intermediate nodes are removed.

The Ritchken-Trevor Algorithm (continued)

- A node with logarithmic price $y_t + \ell\eta\gamma_n$ 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 ℓ.
- The probability that this happens is

$$P(\ell) \equiv \sum_{j_u, j_m, j_d} \frac{n!}{j_u!\, j_m!\, j_d!}\, p_u^{j_u} p_m^{j_m} p_d^{j_d},$$

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)

- A particularly simple way to calculate the $P(\ell)$s starts by noting that

$$(p_u x + p_m + p_d x^{-1})^n = \sum_{\ell=-n}^{n} P(\ell)\, x^{\ell}.$$

- So we expand $(p_u x + p_m + p_d x^{-1})^n$ and retrieve the probabilities by reading off the coefficients.
- It can be computed in $O(n^2)$ time.

The Ritchken-Trevor Algorithm (continued)

- The updating rule (75) on p. 650 must be modified to account for the adoption of the discrete-state model.
- The logarithmic price $y_t + \ell\eta\gamma_n$ at date $t+1$ following state (y_t, h_t^2) at date t has a variance equal to

$$h_{t+1}^2 = \beta_0 + \beta_1 h_t^2 + \beta_2 h_t^2 (\epsilon'_{t+1} - c)^2, \qquad (81)$$

  - Above,

$$\epsilon'_{t+1} = \frac{\ell\eta\gamma_n - (r - h_t^2/2)}{h_t}, \quad \ell = 0, \pm 1, \pm 2, \ldots, \pm n,$$

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

The Ritchken-Trevor Algorithm (continued)

- Different conditional variances h_t^2 may require different η so that the probabilities calculated by Eqs. (78)–(80) on p. 657 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

$$\eta = \lceil h_t/\gamma \rceil, \lceil h_t/\gamma \rceil + 1, \lceil h_t/\gamma \rceil + 2, \ldots$$

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

$$\frac{|\, r-(h_t^2/2)\,|}{2\eta\gamma\sqrt{n}} \leq \frac{h_t^2}{2\eta^2\gamma^2} \leq \min\left(1-\frac{|\, r-(h_t^2/2)\,|}{2\eta\gamma\sqrt{n}}, \frac{1}{2}\right).$$

- Obviously, the magnitude of η tends to grow with h_t.
- The plot on p. 666 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 $\eta = 2$.