

# Lecture Notes for ORIE 473: Empirical Methods in Financial Engineering

**by D. Ruppert**



# Contents

<b>1</b>	<b>Introduction: 3/21/01</b>	<b>1</b>
<b>2</b>	<b>Review of Prob and Stats: 3/12/01</b>	<b>5</b>
2.1	Densities, CDF's, means, variances, and correlation . . . . .	5
2.2	Best Linear Prediction . . . . .	6
2.2.1	Prediction Error . . . . .	7
2.3	Conditional Distributions . . . . .	7
2.4	The Normal Distribution . . . . .	8
2.4.1	Conditional expectations and variance . . . . .	9
2.5	Linear Functions of Random Variables . . . . .	9
2.6	Maximum Likelihood Estimation . . . . .	11
2.7	Likelihood Ratio Tests . . . . .	12
<b>3</b>	<b>Returns: 3/12/01</b>	<b>15</b>
3.1	Prices and returns . . . . .	15
3.2	Log returns . . . . .	16
3.3	Behavior of returns . . . . .	17
3.4	Common Model — IID Normal Returns . . . . .	18
3.5	The Lognormal Model . . . . .	18
3.6	Random Walk . . . . .	21
3.6.1	Geometric Random Walk . . . . .	22
3.7	Are log returns really normally distributed? . . . . .	23
3.7.1	Do the GE daily returns look like a geometric random walk? . . . . .	25
3.8	Portrait of an econometrician, Eugene Fama . . . . .	30

3.9	Other empirical work related to Fama's	32
3.10	Technical Analysis	34
3.11	Fundamental Analysis	35
3.12	Efficient Markets Hypothesis (EMH)	36
3.12.1	Three types of efficiency	36
3.12.2	Testing market efficiency	37
3.13	Summary	38
<b>4</b>	<b>Univariate Time Series Models: 3/12/01</b>	<b>41</b>
4.1	Time Series	41
4.2	Stationary Processes	41
4.2.1	Weak White Noise	42
4.2.2	Estimating parameters of a stationary process	43
4.3	AR(1) processes	43
4.3.1	Properties of a stationary AR(1) process	44
4.3.2	Nonstationary AR(1) processes	45
4.3.3	Estimation	51
4.4	AR(p) models	58
4.4.1	Example: GE daily returns	59
4.5	Moving Average (MA) Processes	59
4.5.1	MA(1) processes	59
4.5.2	General MA processes	60
4.6	ARIMA Processes	60
4.6.1	The backwards operator	60
4.6.2	ARMA Processes	60
4.6.3	The differencing operator	61
4.6.4	From ARMA processes to ARIMA process	61
4.7	Model Selection	64
4.7.1	AIC and SBC	64
4.7.2	Stepwise regression applied to AR processes	66
4.7.3	Using ARIMA in SAS: Cree data	70
4.8	Example: Three-month Treasury bill rates	77
4.9	Forecasting	92
4.9.1	GE daily returns	92

<b>5</b>	<b>Portfolio Selection: 3/12/01</b>	<b>95</b>
5.1	Trading off expected return and risk . . . . .	95
5.2	One risky asset and one risk-free asset . . . . .	96
5.2.1	Example . . . . .	98
5.2.2	Estimating $E(R)$ and $\sigma_R$ . . . . .	100
5.3	Two risky assets . . . . .	100
5.3.1	Estimating means, standard deviations, and covari- ances . . . . .	101
5.4	Combining two risky assets with a risk-free asset . . . . .	103
5.4.1	Tangency portfolio with two risky assets . . . . .	104
5.4.2	Effect of $\rho_{12}$ . . . . .	107
5.5	Harry Markowitz . . . . .	109
5.6	Risk-efficient portfolios with $N$ risky assets . . . . .	110
5.6.1	Efficient-set mathematics . . . . .	110
5.6.2	Selling short . . . . .	126
5.6.3	The Interior decorator fallacy . . . . .	128
5.6.4	Back to the math . . . . .	128
5.6.5	Example: $N = 2$ . . . . .	134
5.7	Is the theory useful? . . . . .	134
5.8	Example—Global Asset Allocation . . . . .	135
5.9	Quadratic programming . . . . .	137
<b>6</b>	<b>The Capital Asset Pricing Model: 3/26/01</b>	<b>141</b>
6.1	Introduction to CAPM . . . . .	141
6.2	The capital market line (CML) . . . . .	144
6.3	Betas and the Security Market Line . . . . .	147
6.3.1	Examples of betas . . . . .	151
6.3.2	Comparison of the CML with the SML . . . . .	151
6.4	The security characteristic line . . . . .	152
6.4.1	Reducing unique risk by diversification . . . . .	153
6.5	Some theory . . . . .	155
6.5.1	Contributions to the market portfolio's risk . . . . .	155
6.5.2	Derivation of the SML . . . . .	156
6.6	Estimation of beta and testing the CAPM . . . . .	158

6.6.1	Interpretation of alpha . . . . .	164
6.7	Summary . . . . .	164
<b>7</b>	<b>Pricing Options: 4/12/01</b>	<b>167</b>
7.1	Introduction . . . . .	167
7.2	Call options . . . . .	168
7.3	The law of one price . . . . .	169
7.3.1	Arbitrage . . . . .	171
7.4	Time value of money and present value . . . . .	171
7.5	A simple binomial example . . . . .	172
7.6	Two-step binomial option pricing . . . . .	176
7.7	Arbitrage pricing by expectation . . . . .	179
7.8	A general binomial tree model . . . . .	181
7.9	Martingales . . . . .	183
7.9.1	The risk-neutral world . . . . .	185
7.10	Trees to random walks to Brownian motion . . . . .	186
7.10.1	Getting more realistic . . . . .	186
7.10.2	A three-step binomial tree . . . . .	187
7.10.3	More time steps . . . . .	189
7.10.4	Properties of Brownian motion . . . . .	190
7.11	Geometric Brownian motion . . . . .	191
7.12	Using the Black-Scholes formula . . . . .	194
7.12.1	How does the option price depend on the inputs? . . . . .	194
7.12.2	An example — GE . . . . .	194
7.12.3	Early exercise of calls is never optimal . . . . .	196
7.12.4	Are there returns on non-trading days? . . . . .	198
7.12.5	Implied volatility . . . . .	199
7.13	Puts . . . . .	201
7.13.1	Pricing puts by binomial trees . . . . .	201
7.13.2	Why are puts different than calls? . . . . .	204
7.13.3	Put-call parity . . . . .	204
7.14	The evolution of option prices . . . . .	206
7.15	Intrinsic value and time value . . . . .	206
7.16	Black, Scholes, and Merton . . . . .	210

7.17	Summary . . . . .	212
7.18	References . . . . .	213
<b>8</b>	<b>GARCH models: 4/24/01</b>	<b>215</b>
8.1	Introduction . . . . .	215
8.2	Modeling conditional means and variances . . . . .	216
8.3	ARCH(1) processes . . . . .	217
8.3.1	Example . . . . .	219
8.4	The AR(1)/ARCH(1) model . . . . .	221
8.5	ARCH( $q$ ) models . . . . .	223
8.6	GARCH( $p, q$ ) models . . . . .	224
8.7	Heavy-tailed distributions . . . . .	226
8.8	Comparison of ARMA and GARCH processes . . . . .	228
8.9	Fitting GARCH models . . . . .	229
8.9.1	Example: S&P 500 returns . . . . .	232
8.10	I-GARCH models . . . . .	238
8.10.1	What does it mean to have an infinite variance? . . . . .	242
8.11	GARCH-M processes . . . . .	246
8.12	E-GARCH . . . . .	248
8.13	Back to the S&P 500 example . . . . .	250
8.14	The GARCH zoo . . . . .	253
8.15	Applications of GARCH in finance . . . . .	253
8.16	Summary . . . . .	254
8.17	References . . . . .	257
<b>9</b>	<b>Fixed Income Securities: 4/30/01</b>	<b>259</b>
9.1	Introduction . . . . .	259
9.2	Zero coupon bonds . . . . .	260
9.2.1	Price and returns fluctuate with the interest rate . . . . .	261
9.3	Coupon bonds . . . . .	262
9.4	Yield to maturity . . . . .	264
9.4.1	Spot rates . . . . .	266
9.5	Term structure . . . . .	267
9.6	Continuous compounding . . . . .	272
9.6.1	Continuous forward rates . . . . .	274

9.7	Summary . . . . .	276
9.7.1	Introduction . . . . .	276
9.7.2	Zero coupon bonds . . . . .	277
9.7.3	Risk due to interest rate changes . . . . .	277
9.7.4	Coupon bonds . . . . .	278
9.7.5	Term structure of interest rates . . . . .	279
9.7.6	Continuous compounding . . . . .	280
9.8	References . . . . .	281
<b>10</b>	<b>Behavioral finance: 5/1/01</b>	<b>283</b>
10.1	Introduction . . . . .	283
10.2	Defense of EMH . . . . .	284
10.3	Challenges to the EMH . . . . .	285
10.4	Can arbitrageurs save the day? . . . . .	286
10.5	What do the data say? . . . . .	287
10.6	References . . . . .	289



# Chapter 1

## Introduction: 3/21/01

The title of this course is “Empirical Research Methods in Financial Engineering.”

“Empirical” means derived from experience, observation, or experiment — so we are going to work with data. We’ll be doing statistics.

Financial engineering is the construction of financial products such as stock options.

Financial engineering uses probability models, e.g., those used to derive the famous Black-Scholes formula.

- are these models supported by financial markets data?
- how are the parameters in these models estimated?

Let’s look ahead to the Black-Scholes formula for the price of a European call option. “Now” is called time 0. The maturity date of the option is  $T$ . The option gives us the right to purchase one share of stock for  $E$  dollars at time  $T$ . Let  $S_T$  be the price of the stock at time  $T$ . At time 0,  $T$  and  $E$  are known but  $S$  is unknown.

At time  $T$ ,  $S_T$  will become known. If at time  $T$  we learn that  $S_T > E$  then

we will exercise the option and purchase one share. We can immediately sell the share for  $S_T$  dollars and earn a profit of  $S_T - E$  dollars.

If at time  $T$ ,  $S_T < E$  then we do not exercise the option. The option expires and we lose the original cost of the option, but no more.

The value of the option at time  $T$  is, therefore,  $\max\{0, S - E\}$ . But right now at time 0, what is the value of the option, i.e., the price for which it should sell on the market?

Prior to the 1970's, options were priced by "seat of pants". Then Black, Scholes, and Merton deduced the correct price of a call option from a mathematical model (and much hard thinking).

They assumed that one can lend and borrow at a risk-free rate  $r$ . Thus, if  $B_t$  is the price at time  $t$  of a risk-free bond purchased for \$1 at time 0, then

$$B_t = \exp(rt).$$

Let  $S_t$  be the price of the underlying stock. They assumed that

$$S_t = S_0 \exp(\mu t + \sigma W_t),$$

where  $\mu$  is a "drift" or "growth rate,"  $W_t$  is a Brownian motion stochastic process, and  $\sigma$  is a standard deviation that measures the volatility of the stock. In this course, you will learn exactly what this model means. Right now, the "take home" message is that there are precise mathematical models of stock price movements that we can check against the data. Also, there are important parameters such as  $\mu$  and  $\sigma$  that must be estimated from data.

The Black-Scholes formula is

$$C = \Phi(d_1)S_0 - \Phi(d_2)E \exp(-rT)$$

where  $C$  is the price of the option at time 0,  $\Phi$  is the standard normal CDF,

$$d_1 = \frac{\log(S_0/E) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}, \quad \text{and} \quad d_2 = d_1 - \sigma\sqrt{T}.$$

The formula is, quite obviously, complicated and it not easy to derive, but it is easy to compute and was hard-wired into calculators almost immediately; the Black-Scholes formula and hand-held calculators both emerged in the early 1970's.

We will be interested in the underlying assumptions behind the formula. Remember: GI – GO (garbage in, garbage out). If the assumptions don't hold, then there is no reason to trust the Black-Scholes formula, despite the impressive mathematics behind it.

The equation  $B_t = \exp(rt)$  of continuous compounding is the solution to the differential equation

$$\frac{dB_t}{dt} = rB_t.$$

The general solution is  $B_t = B_0 \exp(rt)$  and  $B_0 = 1$  since we have assumed that the bond can be purchased for \$1 at time 0.

Where does

$$S_t = S_0 \exp(\sigma W_t + \mu t)$$

come from? If  $\sigma$  were 0, then this would be exponential growth,  $S_t = S_0 \exp(\mu t)$ , just like the bond price  $B_t$ . The term  $\sigma W_t$  comes from the random behavior of stock prices.  $\sigma$  is a standard deviation, essentially of the changes in the stock prices. The random process  $W_t$  is something we will need to learn much more about; and we will.

In this course we will

- study models fo financial markets (they are complex but fascinating)
- learn to test the models — do they fit financial markets data adequately?
- estimate parameters in the models such as  $\mu$  and  $\sigma$  that are essential for correct pricing of financial products such as a call option.

**Key question:** How do the prices of stocks and other financial assets behave?

Looking ahead to where this course is going

- We will start by defining “returns” on the prices of a stock
- We will then look at “ARIMA models”
  - these are models for “time series,” which are sequences of data sampling over time
  - ARIMA models are stochastic processes
- After looking at returns and time series models of return behavior we will look at optimal portfolios of risky assets (e.g., stocks) and of risky assets and risk-free bonds (e.g., US Treasury bills).
  - This will take us to the famous Capital Asset Pricing Model (CAPM)

Looking even farther ahead, we will later return to the pricing of stock options by the Black-Scholes formula and cover other areas of financial engineering such as the term structure of interest rates.

But before we get into applications of probability and statistics in financial engineering, we need to review probability and statistics so that we are all up to speed.

# Chapter 2

## Review of Prob and Stats: 3/12/01

### 2.1 Densities, CDF's, means, variances, and correlation

random variable — large set of possible values but only one will actually occur

continuous random variable —  $X$  is a continuous random variable if it has a p.d.f.  $f_X$  such that

$$P(X \in A) = \int_A f_X(x) dx \quad \text{for all sets } A$$

The CDF of  $X$  is

$$F_X(x) := \int_{-\infty}^x f_X(u) du$$

The expectation of  $X$  is

$$E(X) := \int_{-\infty}^{+\infty} x f_X(x) dx$$

The variance of  $X$  is

$$\sigma_X^2 := \int \{x - E(X)\}^2 f_X(x) dx = E\{X - E(X)\}^2$$

Useful formula:  $\sigma_X^2 = E(X^2) - \{E(X)\}^2$ .

The standard deviation is the square root of the variance.

A pair of random variables,  $(X, Y)$ , as a bivariate density  $f_{XY}(x, y)$

Covariance:

$$\sigma_{XY} = E[\{X - E(X)\}\{Y - E(Y)\}] = \int \{x - E(X)\}\{y - E(Y)\}f_{XY}(x, y) dx dy.$$

Useful formulas:

- $\sigma_{XY} = E(XY) - E(x)E(y)$
- $\sigma_{XY} = E[\{X - E(X)\}Y]$
- $\sigma_{XY} = E[\{Y - E(Y)\}X]$
- $\sigma_{XY} = E(XY)$  if  $E(X) = 0$  or  $E(Y) = 0$

The correlation coefficient between  $X$  and  $Y$  is  $\rho_{XY} := \sigma_{XY}/\sigma_X\sigma_Y$ .

## 2.2 Best Linear Prediction

Suppose we observe  $X$  and want to predict  $Y$ ; this can be done if  $\rho_{XY}$  is not zero.

Best linear prediction means finding  $\beta_0$  and  $\beta_1$  so that

$$H(\beta_0, \beta_1) := E\{Y - (\beta_0 + \beta_1 X)\}^2$$

is minimized.

$$H(\beta_0, \beta_1) = E(Y^2) - 2\beta_0 E(Y) - 2\beta_1 E(XY) + (\beta_0 + \beta_1 X)^2.$$

Setting the partial derivatives to zero we get

$$\begin{aligned} 0 &= -E(Y) + \beta_0 + \beta_1 E(X) \text{ and} \\ 0 &= -E(XY) + \beta_0 E(X) + \beta_1 E(X^2). \end{aligned}$$

After some algebra we find that

$$\beta_1 = \sigma_{XY}/\sigma_X^2 \quad (2.1)$$

and

$$\beta_0 = E(Y) - \beta_1 E(X) = E(Y) - \sigma_{XY}/\sigma_X E(X).$$

Thus, the best linear predictor of  $Y$  is

$$\hat{Y} := \beta_0 + \beta_1 X = E(Y) + \sigma_{XY}/\sigma_X^2 \{X - E(X)\}$$

### 2.2.1 Prediction Error

The prediction error is  $Y - \hat{Y}$ . It is easy to prove that  $E\{Y - \hat{Y}\} = 0$  so that the prediction is “unbiased.” With a little algebra we can show that the expected squared prediction error is

$$E\{Y - \hat{Y}\}^2 = \sigma_Y^2 - \frac{\sigma_{XY}^2}{\sigma_X^2} = \sigma_Y^2(1 - \rho_{XY}^2).$$

If we do not observe  $X$ , then the best predictor of  $Y$  is  $E(Y)$  and the expected squared prediction error is  $\sigma_Y^2$ . Therefore,  $\rho_{XY}^2$  is the fraction by which the prediction error is reduced when  $X$  is known.

**Example:** If  $\rho_{XY} = .5$ , then the prediction error is reduced by 25% by observing  $X$ . If  $\sigma_Y^2 = 3$ , then the expected squared prediction error is 3 if  $X$  is unobserved by only 2 1/4 if  $X$  is observed.

## 2.3 Conditional Distributions

Let  $f_{XY}(x, y)$  be the joint density of a pair of random variables,  $(X, Y)$ .

The marginal density of  $X$  is  $f_X(x) := \int f_{XY}(x, y)dy$  and similarly for  $f_Y$ .

The conditional density of  $Y$  given  $X$  is

$$f_{Y|X}(y|x) = \frac{f_{XY}(x, y)}{f_X(x)}.$$

The conditional expectation of  $Y$  given  $X$  is just the expectation calculated using  $f_{Y|X}(y|x)$ :

$$E(Y|X = x) = \int y f_{Y|X}(y|x) dy$$

which is, of course, a function of  $x$ .

The conditional variance of  $Y$  given  $X$  is

$$\text{Var}(Y|X = x) = \int \{y - E(Y|X = x)\}^2 f_{Y|X}(y|x) dy.$$

**Example:** Suppose  $f_{XY}(x, y) = 2$  on  $0 < x < 1$  and  $x < y < 1$ .

Then the marginal density of  $X$  is  $f_X(x) = 2(1 - x)$ .

The conditional density of  $Y$  given  $X$  is  $f_{Y|X}(y|x) = (1 - x)^{-1}$  for  $x < y < 1$ .

The conditional expectation of  $Y$  is

$$E(Y|X = x) = \frac{1 + x}{2}.$$

The conditional variance of  $Y$  is

$$\text{Var}(Y|X = x) = \frac{(1 - x)^2}{12}.$$

## 2.4 The Normal Distribution

The standard normal distribution has density

$$\phi(x) := \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right).$$

The  $N(\mu, \sigma^2)$  density is

$$\frac{1}{\sigma} \phi\left(\frac{x - \mu}{\sigma}\right).$$

The standard normal CDF is

$$\Phi(x) := \int_{-\infty}^x \phi(u) du.$$



$\Phi$  can be evaluated using tables (ugh!) or more easily using software such as MATLAB or MINITAB.

If  $X \sim N(\mu, \sigma^2)$  then  $P(X \leq x) = \Phi\{(x - \mu)/\sigma\}$ .

**Example:** If  $X \sim N(5, 4)$  then what is  $P(X \leq 7)$ . *Answer:*  $\Phi(1) = .8413$ . In MATLAB, "cdfn(1)" gives "ans = 0.8413".

### 2.4.1 Conditional expectations and variance

The calculation of conditional expectations and variances can be difficult for some probability distributions, but it is quite easy for a pair  $(X, Y)$  that has a bivariate normal distribution.

For a bivariate normal pair, the conditional expectation of  $Y$  given  $X$  equals the best linear predictor of  $Y$  given  $X$ :

$$E(Y|X) = E(Y) + \frac{\sigma_{XY}}{\sigma_X^2}\{X - E(X)\}.$$

The conditional variance of  $Y$  given  $X$  is the expectation squared prediction error:

$$\text{Var}(Y|X) = \sigma_Y^2(1 - \rho_{XY}^2)$$

## 2.5 Linear Functions of Random Variables

$$E(aY + b) = aE(Y) + b$$

where  $Y$  is a random variable and  $a$  and  $b$  are constants. Also,

$$\text{Var}(aY + b) = a^2\text{Var}(Y).$$

If  $X$  and  $Y$  are random variables and  $w_1$  and  $w_2$  are constants, then

$$E(w_1X + w_2Y) = w_1E(X) + w_2E(Y),$$

and

$$\text{Var}(w_1X + w_2Y) = w_1^2\text{Var}(X) + 2w_1w_2\text{Cov}(X, Y) + w_2^2\text{Var}(Y).$$

Check that

$$\text{Var}(w_1X + w_2Y) = \begin{pmatrix} w_1 & w_2 \end{pmatrix} \begin{pmatrix} \text{Var}(X) & \text{Cov}(X, Y) \\ \text{Cov}(X, Y) & \text{Var}(Y) \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}.$$

Let  $\mathbf{X} = (X_1, \dots, X_N)^\top$  be a random vector. We define the expectation vector of  $\mathbf{X}$  to be

$$\begin{pmatrix} E(X_1) \\ \vdots \\ E(X_N) \end{pmatrix}.$$

The covariance matrix of  $\mathbf{X}$  is

$$\text{COV}(\mathbf{X}) = \begin{pmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_2) & \cdots & \text{Cov}(X_1, X_N) \\ \text{Cov}(X_2, X_1) & \text{Var}(X_2) & \cdots & \text{Cov}(X_2, X_N) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_N, X_1) & \text{Cov}(X_N, X_2) & \cdots & \text{Var}(X_N) \end{pmatrix}.$$

Let  $\mathbf{w} = (w_1, \dots, w_N)^\top$  be a vector of weights. Then

$$\mathbf{w}^\top \mathbf{X} = \sum_{i=1}^N w_i X_i$$

is a weighted average of the components of  $\mathbf{X}$ ; it is a random variable.

One can show that

$$E(\mathbf{w}^\top \mathbf{X}) = \mathbf{w}^\top \{E(\mathbf{X})\}.$$

Also

$$\text{Var}(\mathbf{w}^\top \mathbf{X}) = \sum_{i=1}^N \sum_{j=1}^N w_i w_j \text{Cov}(X_i, X_j)$$

This result can be expressed more simply using vector/matrix notation:

$$\text{Var}(\mathbf{w}^\top \mathbf{X}) = \mathbf{w}^\top \text{COV}(\mathbf{X}) \mathbf{w}. \quad (2.2)$$

**Important fact:** If  $\mathbf{X}$  has a multivariate normal distribution, then  $\mathbf{w}^\top \mathbf{X}$  is a normally distributed random variable.

**Example:** Suppose that  $E(X_1) = 1$ ,  $E(X_2) = 1.5$ ,  $\sigma_{X_1}^2 = 1$ ,  $\sigma_{X_2}^2 = 2$ , and  $\text{Cov}(X_1, X_2) = .5$ . Find  $E(.3X_1 + .7X_2)$  and  $\text{Var}(.3X_1 + .7X_2)$ .

If  $(X_1 \ X_2)^\top$  is bivariate normal, find  $P(.3X_1 + .7X_2 < 2)$ .

*Answer:*  $E(.3X_1 + .7X_2) = 1.35$ ,  $\text{Var}(.3X_1 + .7X_2) = 1.28$ , and  $P(.3X_1 + .7X_2 < 2) = \Phi\{(2 - 1.35)/\sqrt{1.28}\} = \Phi(.5745) = .7172$ .

## 2.6 Maximum Likelihood Estimation

Maximum likelihood is the most important and widespread method of estimation. Many well-known estimators such as the sample mean and the least-squares estimator in regression are maximum likelihood estimators. Maximum likelihood is a very useful in practice and tends to give more precise estimates than other methods of estimation.

Let  $\mathbf{Y} = (Y_1, \dots, Y_n)^\top$  be a vector of data and let  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)^\top$  be a vector of parameters. Suppose that  $f(\mathbf{y}; \boldsymbol{\theta})$  is the density of  $\mathbf{Y}$  which depends on the parameters.

**Example:** Suppose that  $Y_1, \dots, Y_n$  are IID  $N(\mu, \sigma^2)$ . Then  $\boldsymbol{\theta} = (\mu, \sigma^2)$ . Also,

$$f(\mathbf{y}; \boldsymbol{\theta}) = \prod_{i=1}^n \frac{1}{\sigma} \phi\left(\frac{Y_i - \mu}{\sigma}\right) = \frac{1}{\sigma^n (2\pi)^{n/2}} \exp\left\{\frac{-1}{2\sigma^2} \sum_{i=1}^n (Y_i - \mu)^2\right\}.$$

$L(\boldsymbol{\theta}) := f(\mathbf{Y}; \boldsymbol{\theta})$  is called the “likelihood function” and is the density evaluated at the observed data. It tells us the likelihood of what was actually observed. The maximum likelihood estimator (MLE) is the value of  $\boldsymbol{\theta}$  that maximizes the likelihood function. In other words, the MLE is the value of  $\boldsymbol{\theta}$  that maximizes the likelihood of the data that was observed. We will denote the MLE by  $\hat{\boldsymbol{\theta}}_{ML}$ . Often it is mathematically easier to maximize  $\log\{L(\boldsymbol{\theta})\}$ ; since the log function is increasing, maximizing  $\log\{L(\boldsymbol{\theta})\}$  is equivalent to maximizing  $L(\boldsymbol{\theta})$ .

**Example:** In the example above, it is an easy calculus exercise to show that  $\hat{\mu}_{ML} = \bar{Y}$ . Also, with  $\mu$  fixed at its MLE, the MLE of  $\sigma^2$  solves

$$0 = \frac{d}{d\sigma^2} \log\{L(\hat{\mu}_{ML}, \sigma^2)\} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (Y_i - \hat{\mu}_{ML})^2.$$

The solution to this equation is

$$\hat{\sigma}_{ML}^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2.$$

The MLE of  $\sigma^2$  has a small bias. The “bias-corrected” MLE is the so-called “sample variance” defined as

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2.$$

In a “textbook example” such as the one above, it is possible to find an explicit formula for the MLE. With more complex models, there is no explicit formula for the MLE. Rather, one writes a program to compute  $\log\{L(\boldsymbol{\theta})\}$  for any value of  $\boldsymbol{\theta}$  and then using optimization software to maximize this function numerically. For some models such as the ARIMA time series models discussed in Chapter 4, there are software packages, e.g, MINITAB and SAS, that compute the MLE; the computation of the log-likelihood function has been pre-programmed.

## 2.7 Likelihood Ratio Tests

Likelihood ratio tests, like maximum likelihood estimation, are a convenient, all-purpose tool. We will consider likelihood ratio tests when we wish to test a restriction on a subset of the parameters. Let

$$\boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 \end{pmatrix}$$

be a partitioning of the parameter vector into two components. Suppose we want to test a hypothesis about  $\boldsymbol{\theta}_1$  without making any hypothesis

about the value of  $\theta_2$ . For example, we might want to test that a population mean is zero; then  $\theta_1 = \mu$  and  $\theta_2 = \sigma^2$ .

Let  $\theta_{1,0}$  be the hypothesized value of  $\theta_1$ , e.g.,  $\theta_{1,0} = 0$  if we want to test that  $\mu$  is zero. Then the hypotheses are

$$H_0 : \theta_1 = \theta_{1,0} \quad \text{and} \quad H_1 : \theta_1 \neq \theta_{1,0}.$$

For example, if we are testing that  $\mu$  is zero then the hypotheses are

$$H_0 : \mu = 0 \quad \text{and} \quad H_1 : \mu \neq 0.$$

Let  $\hat{\theta}_{ML}$  be the maximum likelihood estimator and let  $\hat{\theta}_{2,0}$  be the value of  $\theta_2$  that maximizes  $L(\theta)$  when  $\theta_1 = \theta_{1,0}$ .

**Idea:** If  $H_0$  is true, then  $L(\theta_{1,0}, \hat{\theta}_{2,0})$  should be similar to  $L(\hat{\theta})$ . Otherwise,  $L(\theta_{1,0}, \hat{\theta}_{2,0})$  should be *smaller* than  $L(\hat{\theta})$ .

The likelihood ratio test rejects  $H_0$  if

$$2 \left[ \log\{L(\hat{\theta}_{ML})\} - \log\{L(\theta_{1,0}, \hat{\theta}_{2,0})\} \right] \geq \chi_{\alpha; \dim(\theta_1)}^2.$$

Here  $\dim(\theta_1)$  is the dimension (number of components) of  $\theta_1$  and  $\chi_{\alpha, k}^2$  is the  $\alpha$  upper-probability value of the chi-squared distribution with  $k$  degrees of freedom; the probability above  $\chi_{\alpha, k}^2$  is  $\alpha$ .

**Example:** Suppose again that  $Y_1, \dots, Y_n$  are IID  $N(\mu, \sigma^2)$  and  $\theta = (\mu, \sigma^2)$ . We want to test that  $\mu$  is zero. Note that

$$\log(L) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (Y_i - \mu)^2.$$

If we evaluate  $\log(L)$  at the MLE, we get

$$\log\{L(\bar{Y}, \hat{\sigma}_{ML}^2)\} = -\frac{n}{2} \{1 + \log(2\pi) + \log(\hat{\sigma}_{ML}^2)\}.$$

The value of  $\sigma^2$  that maximizes  $L$  when  $\mu = 0$  is

$$\hat{\sigma}_0^2 = \frac{1}{n} \sum_{i=1}^n Y_i^2.$$

Therefore,

$$2[\log\{L(\bar{Y}, \hat{\sigma}_{ML}^2)\} - \log\{L(0, \hat{\sigma}_0^2)\}] = n \log \left( \frac{\hat{\sigma}_0^2}{\hat{\sigma}_{ML}^2} \right) = n \log \left( \frac{\sum_{i=1}^n Y_i^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \right).$$

The likelihood ratio test rejects  $H_0$  if

$$n \log \left( \frac{\sum_{i=1}^n Y_i^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \right) > \chi_{\alpha, 1}^2.$$

# Chapter 3

## Returns: 3/12/01

### 3.1 Prices and returns

Let  $P_t$  be the price of an asset at time  $t$ .

Assuming no dividends the *net return* is

$$R_t = \frac{P_t}{P_{t-1}} - 1 = \frac{P_t - P_{t-1}}{P_{t-1}}$$

The *simple gross return* is

$$\frac{P_t}{P_{t-1}} = 1 + R_t$$

**Example:** If  $P_t = 2$  and  $P_{t+1} = 2.1$  then

$$1 + R_t = 1.05 \quad \text{and} \quad R_t = .05.$$

The gross return over the most recent  $k$  periods ( $t - k$  to  $t$ ) is

$$\begin{aligned} 1 + R_t(k) &:= \\ \frac{P_t}{P_{t-k}} &= \left( \frac{P_t}{P_{t-1}} \right) \left( \frac{P_{t-1}}{P_{t-2}} \right) \cdots \left( \frac{P_{t-k+1}}{P_{t-k}} \right) \\ &= (1 + R_t) \cdots (1 + R_{t-k+1}) \end{aligned}$$

Returns are scale-free, meaning that they do not depend on units (dollars, cents, etc.). Returns are *not* unitless. Their unit is time; they depend on the units of  $t$  (hour, day, etc.).

**Example:**

Time	$t - 2$	$t - 1$	$t$	$t + 1$
$P$	200	210	206	212
$R$		1.05	.981	1.03
$R_t(2)$			1.03	1.01
$R_t(3)$				1.06

## 3.2 Log returns

Continuously compounded returns, also known as “log returns” are:

$$r_t := \log(1 + R_t) = \log\left(\frac{P_t}{P_{t-1}}\right) = p_t - p_{t-1}$$

where

$$p_t := \log(P_t)$$

[**Notation:**  $\log(x)$  will mean the natural logarithm of  $x$  throughout these notes.  $\log_{10}(x)$  will be used to denote the logarithm to base ten, if it is needed.]

Advantage — simplicity of multiperiod returns

$$\begin{aligned} r_t(k) &:= \log\{1 + R_t(k)\} \\ &= \log\{(1 + R_t) \cdots (1 + R_{t-k+1})\} \\ &= \log(1 + R_t) + \cdots + \log(1 + R_{t-k+1}) \\ &= r_t + r_{t-1} + \cdots + r_{t-k+1} \end{aligned}$$



### 3.3 Behavior of returns

What can we say about returns?

- They cannot be perfectly predicted — i.e., they are random.
- If we were ancient Greeks, we would think of the returns as determined by the Gods or Fates (three Goddesses of destiny). The Greeks did not seem to realize that random phenomena do exhibit some regularities such as the law of large numbers and the central limit theorem.

Peter Bernstein has written an interesting popular book “Against the Gods: The Remarkable Story of Risk.” He chronicles the developments of probability theory and our understanding of risk.

It took a surprisingly long time for probability theory to develop. The ancient Greeks did not have probability theory.

Probability arose out of gambling during the Renaissance.

University of Chicago economist Frank Knight (1916 Cornell PhD) distinguishes between

- measurable uncertainty or “risk proper” (e.g., games of chance) where the probabilities are known
- unmeasurable uncertainty (e.g., finance) where the probabilities are unknown

At time  $t - 1$   $P_t$  and  $R_t$  are not only unknown, but we do not know their probability distributions.

However, we can estimate this probability distribution *if* we are willing to make an assumption.

#### Leap of Faith

Future returns will be similar to past returns.

More precisely, the probability distribution of  $P_t$  can be determined from past data

With this (big) assumption, we can get somewhere — and we will!

Asset pricing models (e.g. CAPM) use the joint distribution of a cross-section  $\{R_{1t}, \dots, R_{Nt}\}$  of returns on  $N$  assets at a single time  $t$

Other models use the time series  $\{R_1, R_2, \dots, R_t\}$  of returns on a single asset. We will start with a single asset.

### 3.4 Common Model — IID Normal Returns

Here  $R_1, R_2, \dots$  are the returns from a single asset. A common model is that they are

1. mutually independent
2. identically distributed, i.e., they have the same mean and variance
3. normally distributed

IID = independent and identically distributed

There are (at least) two problems with this model:

- The model implies the possibility of unlimited losses, but liability is usually limited;  $R_t \geq -1$  since you can lose no more than your investment
- $1 + R_t(k) = (1 + R_t)(1 + R_{t-1}) \cdots (1 + R_{t-k+1})$  is not normal — sums of normals are normal but not so with products.

### 3.5 The Lognormal Model

A second model assumes that the continuously compounded single-period returns, a.k.a. the log returns and denoted by  $r_t$ , are IID. Recall that the log

return is

$$r_t = \log(1 + R_t)$$

where  $1 + R_t$  is the simple gross return

Thus, we assume that

$$\log(1 + R_t) \sim N(\mu, \sigma^2)$$

so that  $1 + R_t = \exp(\text{normal r.v.}) \geq 0$  so that  $R_t \geq -1$ . This solves the first problem.

Also,

$$\begin{aligned} 1 + R_t(k) &= (1 + R_t) \cdots (1 + R_{t-k+1}) \\ &= \exp(r_t) \cdots \exp(r_{t-k+1}) \\ &= \exp(r_t + \cdots + r_{t-k+1}). \end{aligned}$$

Therefore,

$$\log\{1 + R_t(k)\} = r_t + \cdots + r_{t-k+1}$$

Sums of normals are normal  $\Rightarrow$  the second problem is solved — normality of single period returns implies normality of multiple period returns.

The lognormal distribution goes back to Louis Bachelier (1900).

- dissertation at Sorbonne called *The Theory of Speculation*
- Poincarè: “M. Bachelier has evidenced an original and precise mind [but] the subject is somewhat remote from those our other candidates are in the habit of treating.”
- Bachelier was awarded “mention honorable” rather than “mention très honorable” — Bachelier never found a decent academic job.
- Bachelier anticipated Einstein’s (1905) theory of Brownian motion.

In 1827, Brown, a Scottish Botanist, observed the erratic, unpredictable motion of pollen grains under a microscope

Einstein (1905) — movement due to bombardment by water molecules — Einstein developed a mathematical theory giving precise quantitative predictions.

Later, Norbert Wiener, an MIT mathematician, developed a more precise mathematical model of Brownian motion. This model is now called the Wiener process.

[Aside: 1905 was a good year for Einstein. He published:

- the paper on introducing special relativity
- a paper on quantization of light which led quantum theory (which he never embraced — “God does not play dice with the world”)
- the paper on Brownian motion]

Bachelier stated that

- “The mathematical expectation of the speculator is zero” (this is essentially true of short-term speculation but not of long term investing)
- “It is evident that the present theory solves the majority of problems in the study of speculation by the calculus of probability”

Bachelier’s thesis came to light accidentally more than 5 years after he wrote it. Jimmie Savage found a book by Bachelier in the U. Chicago library and asked other economists about it. Paul Samuelson found Bachelier’s thesis in the MIT library. The English translation was published in 1964 in *The Random Character of Stock Market Prices*, an edited volume.

**Example:** A simple gross return,  $(1 + R)$ , is lognormal $(0, (.1)^2)$ . What is  $P(1 + R < .9)$ ?

**Answer:** Since  $\log(.9) = -.105$ ,

$$P(1+R < .9) = P(\log(1+R) < \log(.9)) = \Phi\{(-.105-0)/.1\} = \Phi(-1.05) = .1469$$

In MATLAB,  $\text{cdfn}(-1.05) = .1469$ .

**Example:**

Assume again that  $1 + R$  is log normal( $0, (.1)^2$ ). Find the probability that a simple gross two-period return is less than .9.

*Answer:* The two-period gross return is log normal( $0, 2(.1)^2$ ) so this probability is

$$\Phi \left\{ \frac{\log(.9)}{(\sqrt{2})(.1)} \right\} = \Phi(-.745) = .2281.$$

Let's find a general formula for the  $k$ th period returns:

Assume that

- $1 + R_t(k) = (1 + R_t) \cdots (1 + R_{t-k+1})$ .
- $\log(1 + R_i) \sim N(\mu, \sigma^2)$  for all  $i$
- The  $\{R_i\}$  are mutually independent.

Then  $\log\{1 + R_t(k)\}$  is the sum of  $k$  independent  $N(\mu, \sigma^2)$  random variables, so that  $\log(1 + R_t(k)) \sim N(k\mu, k\sigma^2)$ .

$$P(1 + R_t(k) < x) = \Phi \left\{ \frac{\log(x) - k\mu}{\sqrt{k}\sigma} \right\}.$$

## 3.6 Random Walk

Let  $Z_1, Z_2, \dots$  be IID with mean  $\mu$  and standard deviation  $\sigma$ .  $Z_0$  is an arbitrary starting point. Let  $S_0 = Z_0$  and

$$S_t := Z_0 + Z_1 + \cdots + Z_t, \quad t \geq 1.$$

$S_0, S_1, \dots$  is called a random walk. We have  $E(S_t|Z_0) = Z_0 + \mu t$  and  $\text{Var}(S_t|Z_0) = \sigma^2 t$ .

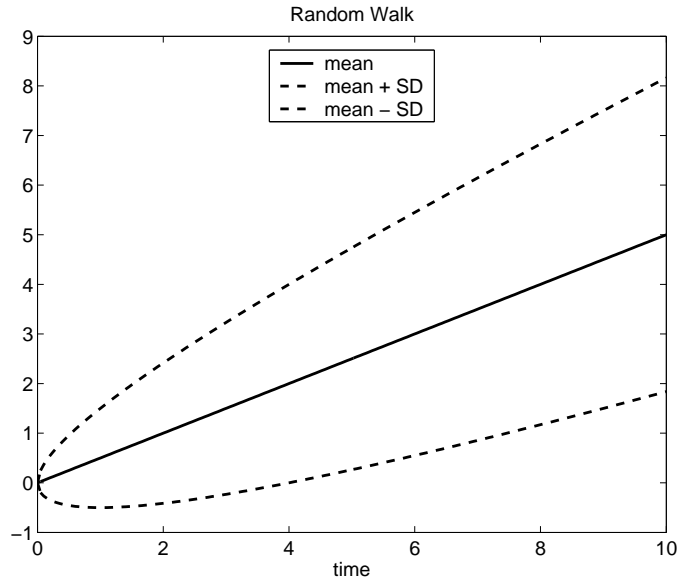


Figure 3.1: Mean and probability bounds on a random walk with  $S_0 = 0$ ,  $\mu = .5$  and  $\sigma = 1$ . At any given time, the probability of being between the probability bounds (dashed curves) is 68%.

### 3.6.1 Geometric Random Walk

Recall that  $\log\{1 + R_t(k)\} = r_t + \dots + r_{t-k+1}$ . Therefore

$$\frac{P_t}{P_{t-k}} = 1 + R_t(k) = \exp(r_t + \dots + r_{t-k+1})$$

so taking  $k = t$  we have

$$P_t = P_0 \exp(r_t + r_{t-1} + \dots + r_1).$$

**Conclusion:** If the log returns are IID normals, then the process  $\{P_t : t = 1, 2, \dots\}$  is the exponential of a random walk. We call such a process a “geometric random walk”.

If  $r = \log(1 + R)$  is  $N(\mu, \sigma^2)$ , then the median of  $R$  is  $\exp(\mu) - 1$  since

$$P(R < \exp(\mu) - 1) = P(1 + R < \exp(\mu)) = P(r < \mu) = P(N(\mu, \sigma^2) < \mu) = \frac{1}{2}.$$

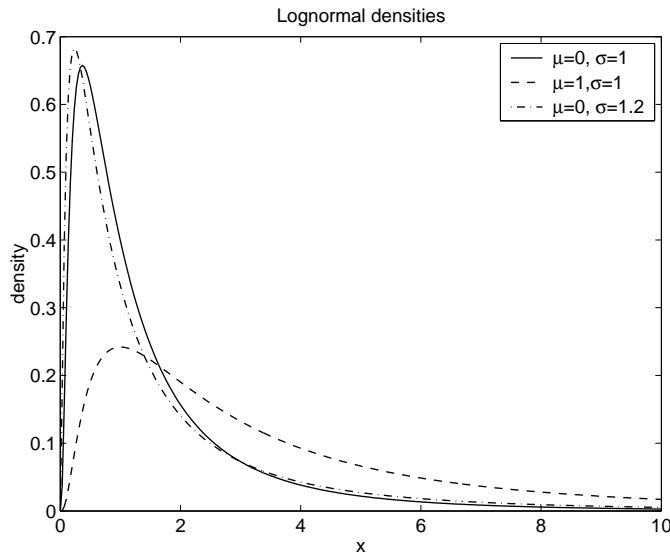


Figure 3.2: Log normal densities.

### 3.7 Are log returns really normally distributed?

There are several ways to check whether log returns are really normally distributed. One way is to look at a normal probability plot of the log returns to see if the plot is approximately a straight line. Another method is to look at the sample skewness and kurtosis of the log returns and to check if their values are near those of the normal distribution; any normal distribution has a skewness coefficient of 0 and a kurtosis of 3.

Suppose we have a time series of log returns,  $r_1, \dots, r_t, \dots, r_T$  on some asset. The sample skewness, denoted by  $\hat{S}$ , is

$$\hat{S} = \frac{1}{T} \sum_{t=1}^T \left( \frac{r_t - \hat{\mu}}{\hat{\sigma}} \right)^3.$$

The sample kurtosis is The sample skewness, denoted by  $\hat{K}$ , is

$$\hat{K} = \frac{1}{T} \sum_{t=1}^T \left( \frac{r_t - \hat{\mu}}{\hat{\sigma}} \right)^4.$$

The “excess kurtosis” is  $\widehat{K} - 3$ . Both the sample skewness and the excess kurtosis should be near 0 if the log returns are normally distributed.

Table 1.1 of Campbell et al. gives  $\widehat{S}$  and  $\widehat{K} - 3$  for several market indices and common stocks. In that table,  $\widehat{S}$  is generally close to zero, which indicates that log returns are not very skewed. However, the excess kurtosis is typically rather large for daily returns and positive though not as large for monthly returns. By the CLT, the distribution of log returns over longer periods should approach the normal distribution. Therefore, the smaller excess kurtosis for monthly log returns, in contrast to daily log returns, is expected. The large kurtosis of daily returns indicates that they are “heavy-tailed.”

Normal probability plots can be supplemented by tests of normality based on the sample CDF,  $\widehat{F}$ .  $\widehat{F}(x)$  is defined to be the proportion of the sample that is less than or equal to  $x$ ; if 10 out of 40 data points are 3 or less then  $\widehat{F}(3) = .25$ . Normality is tested by comparing the sample CDF with the normal CDF with mean and variance equal to the sample mean and variance, i.e., with compare  $\widehat{F}(x)$  with  $\Phi\{(x - \widehat{\mu})/s\}$ . Three common tests of normality that compare the sample CDF with the normal CDF are the Anderson-Darling test, the Shapiro-Wilks test, and the Kolmogorov-Smirnov test. All three are available on MINITAB. Actually, MINITAB uses the Ryan-Joiner test which is close to the Shapiro-Wilks test. In MINITAB, go to “Stat,” then “Basic Statistics,” and then “Normality test.” You will need to choose one of the three tests. The output is a normal plot plus the results of the test. You can re-run the procedure to run the other tests.

The Kolmogorov-Smirnov test is based on the maximum distance between the sample CDF and the normal CDF.

The Shapiro-Wilks test is closely tied to the normal probability plot, since it is based on the correlation between the normal quantiles and the sample quantiles. The correlation measures how close the normal plot is to being a straight line.



CDF's and quantiles are closely related. In fact, quantiles are given by the inverse of the CDF function; if a random variable  $X$  has CDF  $F$  then the  $p$ th quantile of  $X$  is  $F^{-1}(p)$  since  $P\{X \leq F^{-1}(p)\} = F\{F^{-1}(p)\} = p$ .

Let's look at daily returns for GE common stock from December 1999 to December 2000. The daily price  $P_t$  is taken to be the average of the high and the low for the day. It might have been better to use the closing price for each day. Why?

As can be seen in Figure 3.3, the net returns  $R$  and the log returns  $R$  are *very* similar. A normal plot is roughly linear.

The log return have a sample mean, standard deviation, skewness, and excess kurtosis of .00014, .0176, -.094, and .094, respectively. The values of the sample skewness and excess kurtosis suggest that the log returns are approximately normally distributed.

From MINITAB, the Kolmogorov-Smirnov, Anderson-Darling, and Ryan-Joiner tests of normality have a p-values of .15, .40, and .10, respectively. Since each p-value exceeds .05, each test would accept the null hypothesis of normality at  $\alpha = .05$ .

### 3.7.1 Do the GE daily returns look like a geometric random walk?

Figure 3.4 shows five independent simulated geometric random walks with the same parameters as the GE daily log returns. Note that the geometric random walks seem to have "patterns" and "momentum" even though they do not. The GE log returns look similar to the geometric random walks.

It is somewhat difficult to distinguish between a random walk and a geometric random walk. Figure 3.5 shows three independent simulated time series. For each pair, the log price series (a random walk) is plotted on

the left while the price series (a geometric random walk) is plotted on the right. Note the subtle differences between the prices and the log prices.

We prefer the geometric random walk model to the random walk model, because the geometric random walk model is more realistic: the geometric random walk implies non-negative prices and net returns that are at least  $-1$ .

This graphical comparison of GE prices to geometric random walks is not strong evidence in favor of the geometric random walk hypothesis. This hypothesis implies that the log returns are *mutually independent* and, therefore, uncorrelated. Therefore we should check for evidence that the log returns are correlated. If we find no such evidence, then we have more reason to believe the geometric random walk hypothesis.

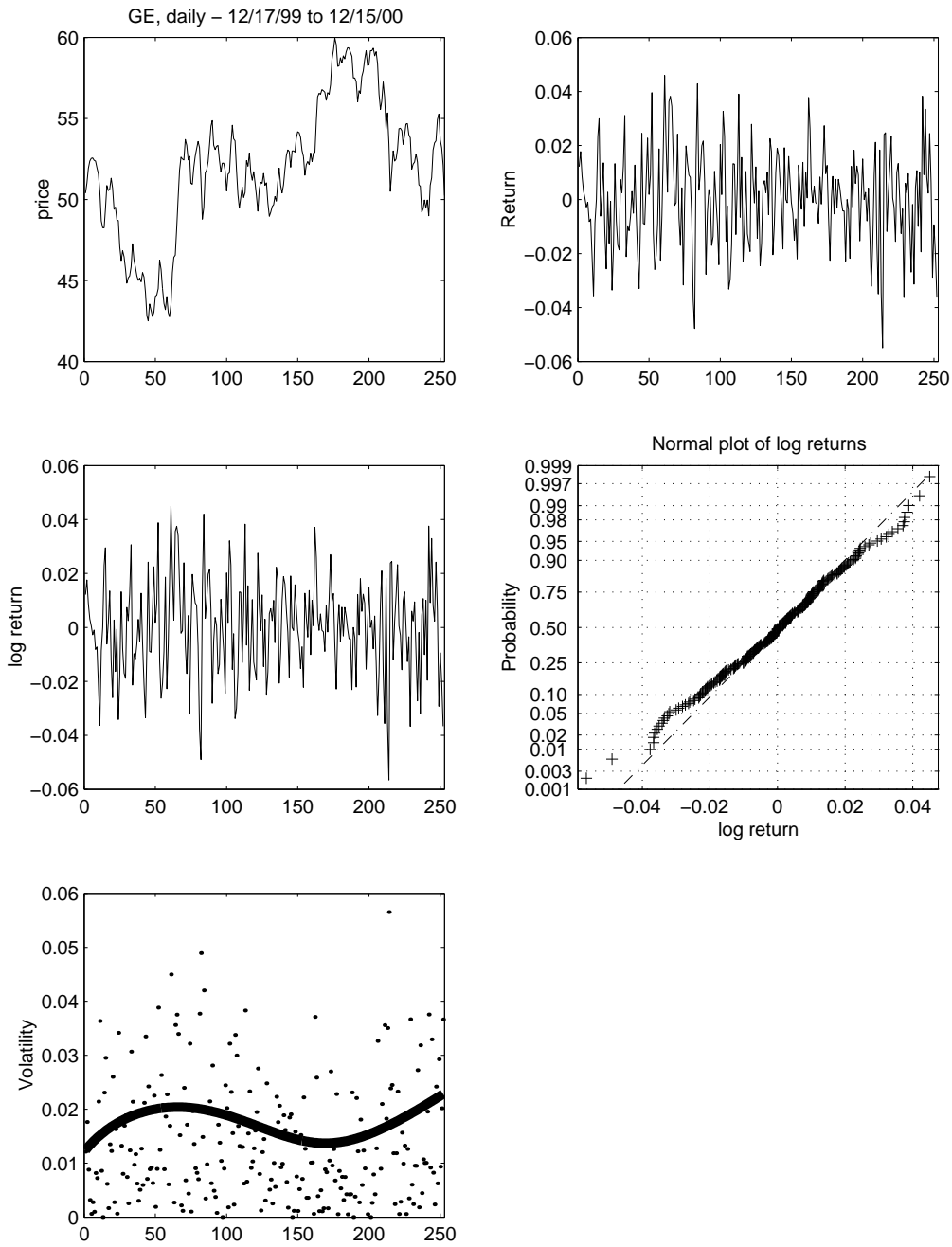


Figure 3.3: GE daily returns. The first plot is the prices. The second and third are the net returns and the log returns. The fourth plot is a normal probability plot of the log returns. The final plot is of the absolute log returns; there is a scatterplot smooth to help show whether the volatility is constant.

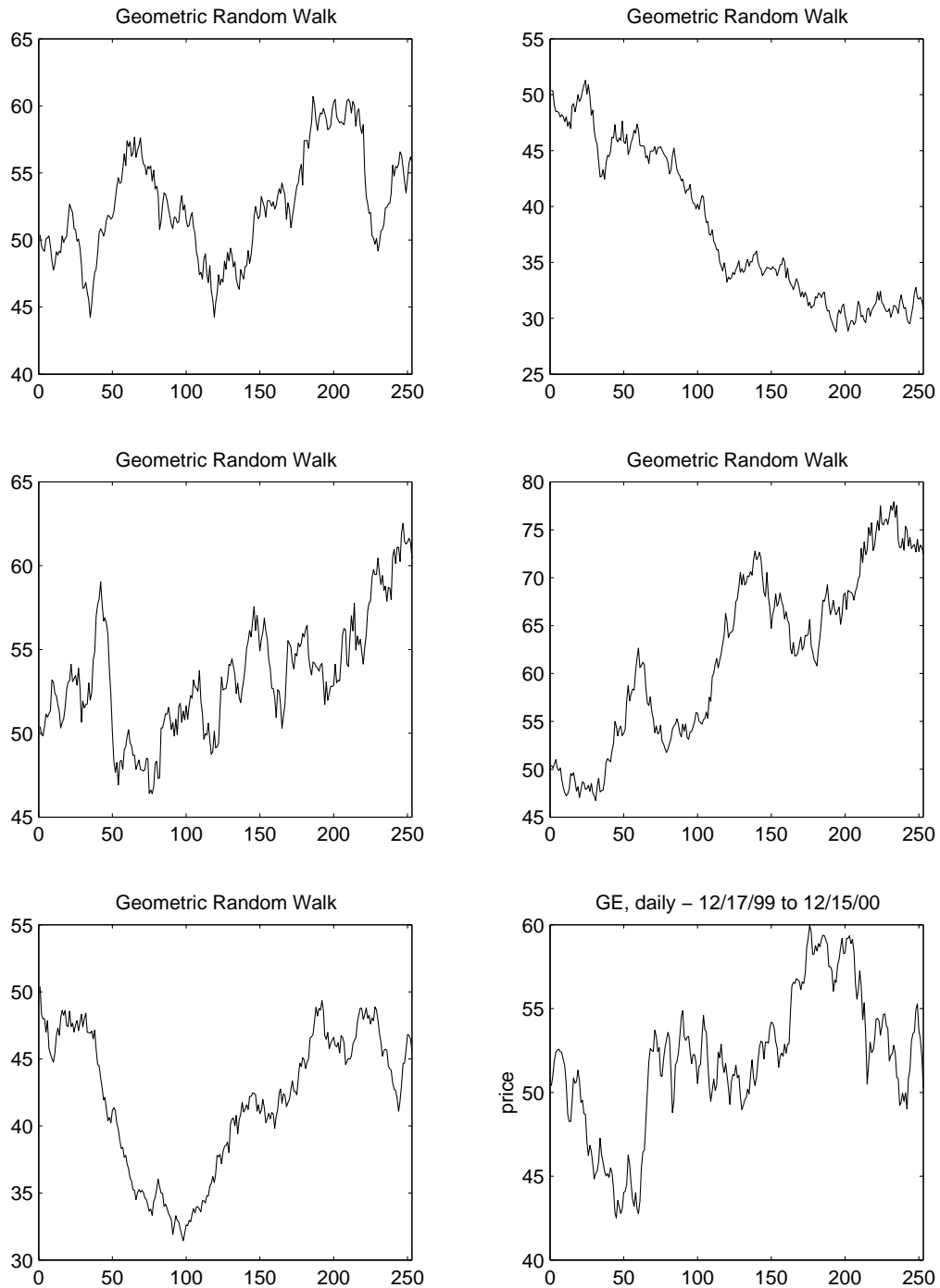


Figure 3.4: Five independent geometric random walks and GE daily log returns. The geometric random walks have the same expected log return, volatility, and starting point as the GE log returns.

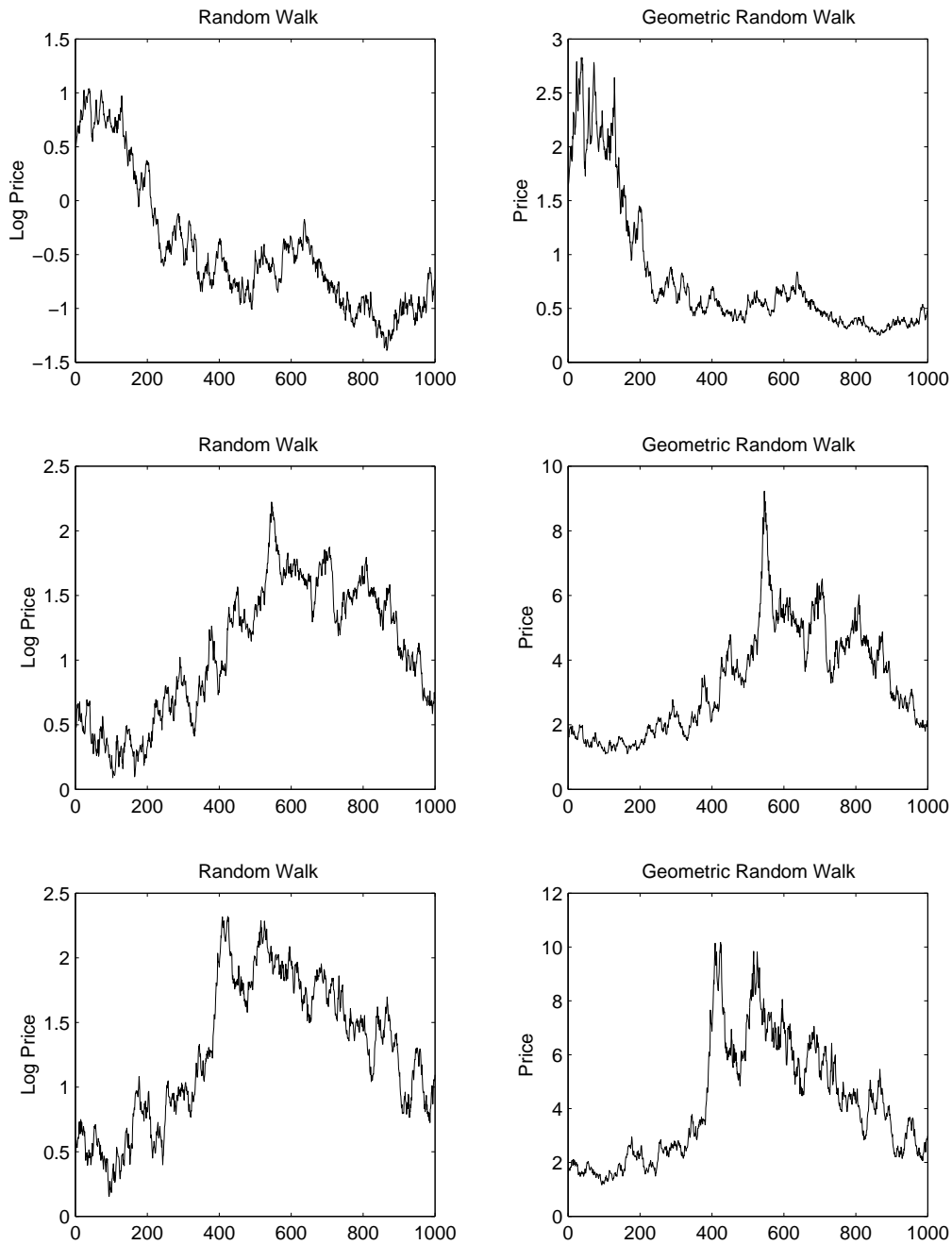


Figure 3.5: Three independent simulated price series. On left: log prices. On right: prices.

### 3.8 Portrait of an econometrician, Eugene Fama

This material is taken from Chapter 7 of *Capital Ideas* by Peter Bernstein.

Fama was born in 1939 in Boston, majored in French at Tufts, and was an outstanding student-athlete.

In college, Fama earned extra money working for Harry Ernst who published a stock market newsletter:

- Fama's job was to find workable buy and sell signals.
- Ernst believed that trends, once in place, would continue because of "price momentum."
- Bernstein writes that "Fama's efforts to develop profitable trading rules were by no means unsuccessful" but "the ones he found worked only on the old data, not on the new."
  - like many other investors Fama found that rules that worked well on "backtests" couldn't beat the market when applied in real time.
  - the market environment would shift or too many people would be using the same strategy

Fama decided to go to business school to learn what was really going on.

- 1964 doctorate at University of Chicago.
- he thought of going to Harvard but was told that he was "more intellectual than the typical Harvard type"

Fama stayed at Chicago where he taught finance.

- scholars at Chicago were keenly interested in collecting facts (empirical research!)

- at Chicago, James Lorie and Lawrence Fisher were demonstrating what the computer could offer to economic research
  - 1964: Lorie and Fisher published a “bombshell” — \$1000 invested in 1926 would grow to almost \$30,000 in 1960, a growth of over 9% a year ( $\log(30)/35 = .097$ )
    - \* Remember: 1929 was the great crash and the ensuing great depression lasted until the US entered WW II in the 40’s. This was not exactly a favorable time for investing.
    - \* These findings increased the interest in stocks as long-term investments
- 1965: Fama published “The Behavior of Stock Market Prices” (his thesis) in *Journal of Business*.
  - a less technical version was published in 1966 as “Random Walks in Stock Market Prices” in *Financial Analysts Journal*.
  - the less technical version was reprinted in *Institutional Investor*.
- Fama’s first target was “technical analysis” as practiced by so-called “chartists.”
  - technical analysts believe that future prices can be predicted from past patterns
  - Charting stock prices was once fashionable
    - \* I remember as a young child my grandmother explaining to me how to chart stock prices.
  - Fama: “The chartist must admit that the evidence in favor of the random walk model is both consistent and voluminous, whereas there is precious little published in discussion of rigorous empirical test of various technical theories.”
- Fama’s next target was “fundamental analysis” as practiced by securities analysts.

- Fundamental analysts examine accounting data, interview management, and look at economic forecasts, interest rates, and political trends.
- Selecting stocks by fundamental analysis seems to do no better than using a dartboard
- Of course, good management, favorable economic trends, etc. influence the prices of assets, but Fama claimed that this information is already fully reflected in stock prices by the time we learn it — markets react instantaneously to information.
- Security analysis is essential in order for stocks to be priced correctly, but ironically it means that there are few discrepancies between actual prices and the values of stocks
- William Sharpe discussed the antagonism of professional investors to the random walk theories of Fama and other academics. He stated that “Interestingly, professional economists seem to think more highly of professional investors than do other professional investors.” (Later we will learn more about Sharpe, the economist who developed the CAPM and winner of the Nobel Prize.)

### 3.9 Other empirical work related to Fama’s

Fama’s work was preceded by that of other researchers.

- In 1933 Alfred Cowles published “Can stock market forecasters forecast?” The three-word abstract stated “It is doubtful.” The article appeared in the brand-new journal *Econometrica*. *Econometrica* is now the leading journal in econometrics.
  - Cowles analyzed the track records of:
    - \* 16 leading financial services that furnished their subscribers with selected lists of common stocks



- \* purchases and sales of stock by 20 leading fire insurance companies
  - \* 24 publications by financial services, financial weeklies, bank letters, etc.
  - \* editorials in *The Wall Street Journal* by William Peter Hamilton, an expounder of the "Dow Theory" due to Charles Dow (the Dow of Dow-Jones). Dow compared stock prices to tides and ocean waves; the tides were a way to explain "price momentum."
  - Cowles found that only 6 of 16 financial services had achieved any measure of success
    - \* even the best record could not be definitely attributed to skill rather than luck (one needs statistical analysis to reach such a conclusion)
  - In 1944, Cowles published a new study with basically the same conclusions.
- In 1936, Holbrook Working published a paper in *The Journal of the American Statistical Association* on commodity prices.
    - These were once believed to have rhythms and trends.
    - Working found that he could not distinguish the price *changes* from an independent sequence of random changes.
    - Perturbed, Working took his data to professional commodity traders.
      - \* He also showed them graphs of random series.
      - \* The professionals could not distinguish the random series from real commodity prices.
      - \* of course, Working's study does not prove anything about stock returns, but it is an interesting example of a financial time series where momentum was thought to exist, but where no evidence of momentum was found in a statistical analysis.

- Maurice Kendall published the paper “The analysis of economic time series” in the *Journal of the Royal Statistical Society* in 1953.
  - Kendall wrote “the patterns of events in the price series was much less systematic than is generally believed,” and
  - “Investors can, perhaps, make money on the Stock Exchange, but not, apparently by watching price movements and coming in on what looks like a good thing ... But it is unlikely that anything I say or demonstrate will destroy the illusion that the outside investor can make money by playing the markets, so let us leave him to his own devices.”

There is no question as to whether one can make money in the stock market. Over the long haul, stocks outperform bonds which outperform savings accounts. The question is rather whether anyone can “beat the market.”

### 3.10 Technical Analysis

“A Random Walk Down Wall Street” was written by Burton G. Malkiel, a professor of economics at Princeton. It is a perennial best seller and has been revised several times. It contains much sensible advice for the small investor. This book is also quite humorous, and the discussion of technical analysts is particularly amusing (unless you are a technical analyst).

Malkiel writes

I, personally, have never known a successful technician, but I have seen the wrecks of several unsuccessful ones. (This is, of course, in terms of following their own technical advice. Commissions from urging customers to act on their recommendations are very lucrative.)

Malkiel describes many of the technical theories, including the Dow Theory, the Filter System, and the Relative-Strength system, which advises

buying stocks that have done well recently. There is also the hemline theory which predicts price changes by the lengths of women's dresses and the super bowl indicator which says that "a victory by an NFL team predicts a bull market, whereas a victory by a former AFL team is bad news for stock-market investors." There is also the odd-lot theory. It is based on the impeccable logic that a person who is *always* wrong is a reliable source of information—just negate whatever that person says. The believe is that the odd-lot trader is precisely that sort of person. It turns out that the odd-lotter isn't such a dolt after all.

Human nature seems to find randomness very hard to accept. For example, sports fans have many theories of streaks in athletics, e.g., the "hot hand" theory of basketball. Extensive testing of basketball players' performances have show no evidence of streaks beyond what would be expected by pure chance. The point is that streaks will occur by chance, but you cannot make money on the basis of random streaks since you cannot predict if they will continue.

Why are technicians hired? Malkiel has the skeptical view that it is because their theories recommend a lot of trading. "The technicians do not help produce yachts for the customers, but they do help generate the trading that provides yachts for the brokers."

### 3.11 Fundamental Analysis

The practitioners of fundamental analysis are called *security analysts*. Their job is basically to predict future earnings of companies, since it is future earnings that ultimately drive prices.

Although few on Wall Street still have much faith in technical analysis, there is much faith in fundamental analysis. However, some academics studying the financial markets data have come to the conclusion that security analysts can do no better than blindfolded monkeys that throw darts at the *Wall Street Journal*.

## 3.12 Efficient Markets Hypothesis (EMH)

As evidence accumulated that stock price fluctuated like random walks, economists sought a theory as to why that would be so. In 1965 Paul Samuelson published a paper "Proof that properly anticipated prices fluctuate randomly." The idea is that random walk behavior is due to the very efficiency of the market.

- A market is information efficient if prices "fully reflect" available information
- A market is "efficient with respect to an information set" if prices would be unchanged by revealing that information to *all* participants
  - this implies that it is impossible to make economic profits by trading on the basis of this information set
- This last idea is the key to testing (empirically) the EMH.

### 3.12.1 Three types of efficiency

**weak-form efficiency** the information set includes only the history of prices or returns

**semi-strong efficiency** the information set includes all information that is publically available

**strong-form efficiency** the information set includes all information known to any market participant

Weak-form efficiency  $\Rightarrow$  technical analysis will not make money

Semistrong-form efficiency  $\Rightarrow$  fundamental analysis will not help the average investor

### 3.12.2 Testing market efficiency

The research of Fama, Cowles, Working, and Kendall just described tests the various forms of the EMH. Cowles's work supports the semi-strong and perhaps the strong form of the EMH.

In their book *Investments*, Bodie, Kane, and Marcus discuss some of the issues involved when testing the EMH. One is the magnitude issue. No one believes that markets are perfectly efficient. The small inefficiencies might be important to the manager of a large portfolio. If one is managing a \$5 billion portfolio, beating the market by .1% results in a \$5 million increase in profit. This is clearly worth achieving. Yet, no statistical test is likely to undercover a .1% inefficiency amidst typical market fluctuations. The S&P 500 index has a 20% standard deviation in annual returns.

Another issue is selection bias. If there is someone who can consistently beat the market, they probably are keeping that a secret. We can only test market efficiency by testing methods of technical or fundamental analysis that are publicized. These may be the ones that don't reveal market inefficiencies.

Another problem is that for any time periods, *by chance* there will be some investment managers that consistently beat the the market.

- if 2,000 people each toss a coin 10 times, it is likely that at least one will get 10 heads since

$$2,000 * 2^{-10} = 1.95.$$

Using the Poisson approximation to the binomial, the probability that no one tosses 10 heads is  $\exp(-1.95) = .14$ .

If some does toss 10 heads, it would be a mistake to say that that person has skill in tossing heads.

- Peter Lynch's Magellan Fund outperformed the S&P 500 in 11 of 13 years ending in 1989. Was Lynch a skilled investment manager or

just lucky? (If he really was skilled, then this is evidence against the semi-strong form of the EMH.)

Campbell, Lo, and MacKinlay and Bodie, Kane, and Marcus discuss much of the empirical literature on testing the EMH and give references to the original studies.

Fama was written a review article:

Fama, E. (1970), "Efficient Capital Markets: A Review of Theory and Empirical Work," *Journal of Finance*, 25, 383–417.

There is a sequel as well:

Fama, E., (1991), "Efficient Capital Markets: II," *Journal of Finance*, 46, 1575–1618.

The *Journal of Finance*, as well as many other journals in economics and finance, are available online at JStor:

<http://www.jstor.org/cgi-bin/jstor/listjournal>

However, the most recent five years of these journals are not available online.

**Good course project:** Read one or more of the studies of the EMH and prepare a report summarizing the work. The two review articles by Fama could help you find studies that would interest you. Using some new financial markets data, try to replicate some of original work.

### 3.13 Summary

Let  $P_t$  be the price of an asset at time  $t$ . Then  $P_t/P_{t-1}$  is the simple gross return and  $R_t = P_t/P_{t-1} - 1$  is the simple net return. ("Simple" means one period.) The gross return over the last  $k$  periods is  $1 + R_t(k) = P_t/P_{t-k}$ . Let  $p_t = \log(P_t)$ . The (one-period) log return is  $r_t = p_t - p_{t-1}$ .  $R_t \approx r_t$ .

Log returns are often modeled as geometric random walks. This model implies that log returns are mutually independent; one cannot predict future returns from past returns. The model also implies that  $R_t$  is lognormally distributed.

Empirical research by Eugene Fama, Alfred Cowles, Holbrook Working, Maurice Kendall, and other econometricians supports the geometric random walk model.

The geometric random walk suggests the efficient market hypothesis (EMH) that states that all valuable information is reflected in the market prices; price changes occur because of unanticipated new information. There are three forms of the EMH, the weak form, the semi-strong form, and the strong form.





# Chapter 4

## Univariate Time Series Models: 3/12/01

### 4.1 Time Series

A univariate time series is a sequence of observations taking over time, for example, a sequence of daily returns on a stock. A multivariate time series is a sequence of vectors of observations taking in time, for example, the sequence of vectors of returns on a fixed set of stocks.

In this chapter, we will study statistical models for univariate times series. These models are widely used in econometrics as well as in other business and OR applications. For example, time series models are routinely used to model the output of simulations.

### 4.2 Stationary Processes

A process is stationary if its behavior is unchanged by shifts in time. More precisely  $X_1, X_2, \dots$  is a weakly stationary process if

- $E(X_i) = \mu$  (a constant) for all  $i$
- $\text{Var}(X_i) = \sigma^2$  (a constant) for all  $i$

- $\text{Corr}(X_i, X_j) = \rho(|i - j|)$  for all  $i$  and  $j$

Thus, the mean and variance do not change with time and the correlation between two observations depends only on the time distance between them. For example, if the process is stationary then the correlation between  $X_2$  and  $X_5$  is the same as the correlation between  $X_7$  and  $X_{10}$ , since each pair are separated from each other by three units of time.

$\rho$  is called the correlation function of the process. Note that  $\rho(h) = \rho(-h)$ . Why?

The covariance between  $X_t$  and  $X_{t+h}$  is denoted by  $\gamma(h)$ .  $\gamma(\cdot)$  is called the autocovariance function. Note that  $\gamma(h) = \sigma^2\rho(h)$  and that  $\gamma(0) = \sigma^2$  since  $\rho(0) = 1$ .

### 4.2.1 Weak White Noise

White noise is the simplest example of a stationary process.  $X_1, X_2, \dots$  is a  $\text{WN}(0, \sigma^2)$  process (weak white noise process) if

- $E(X_i) = 0$  for all  $i$
- $\text{Var}(X_i) = \sigma^2$  (a constant) for all  $i$
- $\text{Corr}(X_i, X_j) = 0$  for all  $i \neq j$

If in addition,  $X_1, X_2, \dots$  are independent normal random variables, then the process is called a *Gaussian white noise process*. (The normal distribution is sometimes called the Gaussian distribution.)

A weak white noise process is weakly stationary with

$$\begin{aligned}\rho(0) &= 1 \\ \rho(t) &= 0 \text{ if } t \neq 0.\end{aligned}$$

### Properties of Gaussian white noise

$$E(X_{i+t}|X_1, \dots, X_i) = 0 \text{ for all } t \geq 1.$$

(You cannot predict the future, because the future is independent of the past and present.)

To us, “white noise” will mean weak white noise, which includes Gaussian white noise as a special case.

White noise (either weak or Gaussian) is uninteresting in itself but is the building block of important time series models used for economic data.

#### 4.2.2 Estimating parameters of a stationary process

Suppose we observe  $y_1, \dots, y_n$  from a stationary process. To estimate the mean  $\mu$  and variance  $\sigma^2$  of the process we use the sample mean  $\bar{y}$  and sample variance  $s^2$ .

To estimate the autocovariance function we use

$$\hat{\gamma}(h) = n^{-1} \sum_{j=1}^{n-h} (y_{j+h} - \bar{y})(y_j - \bar{y}).$$

To estimate  $\rho(\cdot)$  we use the sample autocorrelation function (SACF) defined as

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}.$$

### 4.3 AR(1) processes

Let  $\epsilon_1, \epsilon_2, \dots$  be  $\text{WH}(0, \sigma_\epsilon^2)$ . We say that  $y_1, y_2, \dots$  is an AR(1) process if for some constant parameters  $\mu$  and  $\phi$

$$y_t - \mu = \phi(y_{t-1} - \mu) + \epsilon_t \quad (4.1)$$

for all  $t$ .

If  $|\phi| < 1$ , then  $y_1, \dots$  is a weakly stationary process. Its mean is  $\mu$ . Simple algebra shows that (4.1) can be rewritten as

$$y_t = (1 - \phi)\mu + \phi y_{t-1} + \epsilon. \quad (4.2)$$

Remember the linear regression model,  $y_t = \beta_0 + \beta_1 x_t + \epsilon_t$  from your statistics courses. (4.2) is just a linear regression model with  $\beta_0 = (1 - \phi)\mu$  and  $\beta_1 = \phi$ . If it is assumed that  $\mu = 0$ , then  $\beta_0 = 0$  as well. Linear regression with  $\beta_0 = 0$  is the “linear regression through the origin model.” The term *autoregression* refers to the regression of the process on its own past values.

When  $|\phi| < 1$  then

$$y_t = \mu + \epsilon_t + \phi\epsilon_{t-1} + \phi^2\epsilon_{t-2} + \dots = \mu + \sum_{h=0}^{\infty} \phi^h \epsilon_{t-h} \quad (4.3)$$

(infinite moving average (MA( $\infty$ )) representation).

### 4.3.1 Properties of a stationary AR(1) process

When  $|\phi| < 1$  (stationarity), then

1.

$$E(y_t) = \mu \quad \forall t$$

2.

$$\gamma(0) = \text{Var}(y_t) = \frac{\sigma_\epsilon^2}{1 - \phi^2} \quad \forall t$$

3.

$$\gamma(h) = \text{Cov}(y_t, y_{t+h}) = \frac{\sigma_\epsilon^2 \phi^{|h|}}{1 - \phi^2} \quad \forall t.$$

4.

$$\rho(h) = \text{Corr}(y_t, y_{t+h}) = \phi^{|h|} \quad \forall t.$$

**It is important to remember that these formulas hold only if  $|\phi| < 1$  and only for AR(1) processes.** If  $|\phi| \geq 1$ , then the AR(1) process is nonstationary, and the mean, variance, and correlation are not constant.

These formulas can be proved using (4.3). For example

$$\text{Var}(y_t) = \text{Var}\left(\sum_{h=0}^{\infty} \phi^h \epsilon_{t-h}\right) = \sigma_\epsilon^2 \sum_{h=0}^{\infty} \phi^{2h} = \frac{\sigma_\epsilon^2}{1 - \phi^2}.$$

Also, for  $h > 0$

$$\text{Cov}\left(\sum_{i=0}^{\infty} \epsilon_{t-i} \phi^i, \sum_{j=0}^{\infty} \epsilon_{t+h-j} \phi^j\right) = \frac{\sigma_\epsilon^2 \phi^{|h|}}{1 - \phi^2}.$$

Be sure to distinguish between  $\sigma_\epsilon^2$  which is the variance of the stationary white noise process  $\epsilon_1, \epsilon_2, \dots$  and  $\gamma(0)$  which is the variance of the AR(1) process  $y_1, y_2, \dots$ . We can see from the result above that  $\gamma(0)$  is bigger than  $\sigma_\epsilon^2$  unless  $\phi = 0$  in which case  $y_t = \epsilon_t$ .

### 4.3.2 Nonstationary AR(1) processes

#### Random Walk

If  $\phi = 1$  then

$$y_t = y_{t-1} + \epsilon$$

and the process is **not** stationary. This is the random walk process we saw in Chapter 3.

It is easy to see that

$$y_t = y_0 + \epsilon_1 + \dots + \epsilon_t.$$

Suppose we start at the process at an arbitrary point  $y_0$ . Then  $E(y_t|y_0) = y_0$  for all  $t$ , which is constant but depends entirely on the arbitrary starting point. Moreover,  $\text{Var}(y_t|y_0) = t\sigma_\epsilon^2$  which is not stationary but rather increases linearly with time. The increasing variance makes the random walk “wander.”

### AR(1) processes when $|\phi| > 1$

When  $|\phi| > 1$ , an AR(1) process has explosive behavior. This can be seen in Figure 4.1. This figure shows simulations of 200 observations from AR(1) processes with various values of  $\phi$ . The explosive case where  $\phi = 1.02$  clearly is different than the other cases where  $|\phi| \leq 1$ . However, the case where  $\phi = 1$  is not that much different than  $\phi = .9$  even though the former is non-stationary while the latter is stationary.

The ability to distinguish the three types of AR(1) processes (stationary, random walk, and explosive) depends on the length of the observed series. For short AR(1), it is very difficult to tell if the process is stationary, random walk, or explosive. For example, in Figure 4.2, we see 30 observations from processes with the same parameter values as in Figure 4.1. If we observe the AR processes for longer than 200 observations, then the behavior of  $\phi = .9$  and  $\phi = 1$  processes would not look as similar as in Figure 4.1. For example, in Figure 4.3 there are 1,000 observations from each of the processes. Now the processes with  $\phi = .9$  and  $\phi = 1$  look dissimilar. The stationary process  $\phi = .9$  continues to return to its mean of zero. The random walk ( $\phi = 1$ ) wanders without tending to return to zero.

Suppose an explosive AR(1) process starts at  $y_0$  and has  $\mu = 0$ . Then

$$y_t = \phi y_{t-1} + \epsilon_t = \phi(y_{t-2} + \epsilon_{t-1}) + \epsilon_t = \cdots = \epsilon_t + \phi\epsilon_{t-1} + \phi^2\epsilon_{t-2} + \cdots + \phi^t\epsilon_0.$$

Therefore,

$$\text{Var}(y_t) = \sigma^2(1 + \phi^2 + \phi^4 + \cdots + \phi^{2t}) = \sigma^2 \frac{\phi^{2(t+1)} - 1}{\phi^2 - 1}.$$

This variance increase geometrically fast at  $t \rightarrow \infty$ .

Explosive AR processes are not widely used in econometrics since economic growth is usually not explosive, though these processes may serve as good models of rabbit populations.

```

      /|      _
      / |      ,~ /
      Y :| // /
      | jj /(. ^
      >-"~"-v"
      /      Y
      jo o |
      ( ~T~ j
      >.-' _./
      / " ~ " |
      Y      _' |
      /| i-~ - 1
      / 1/ ,-"~ \
      \//\//      .- \
      Y      / Y*
      l      I !
      l\      _\ /"\
      (" ~----( ~ Y. )
      ~~~~~~
    
```

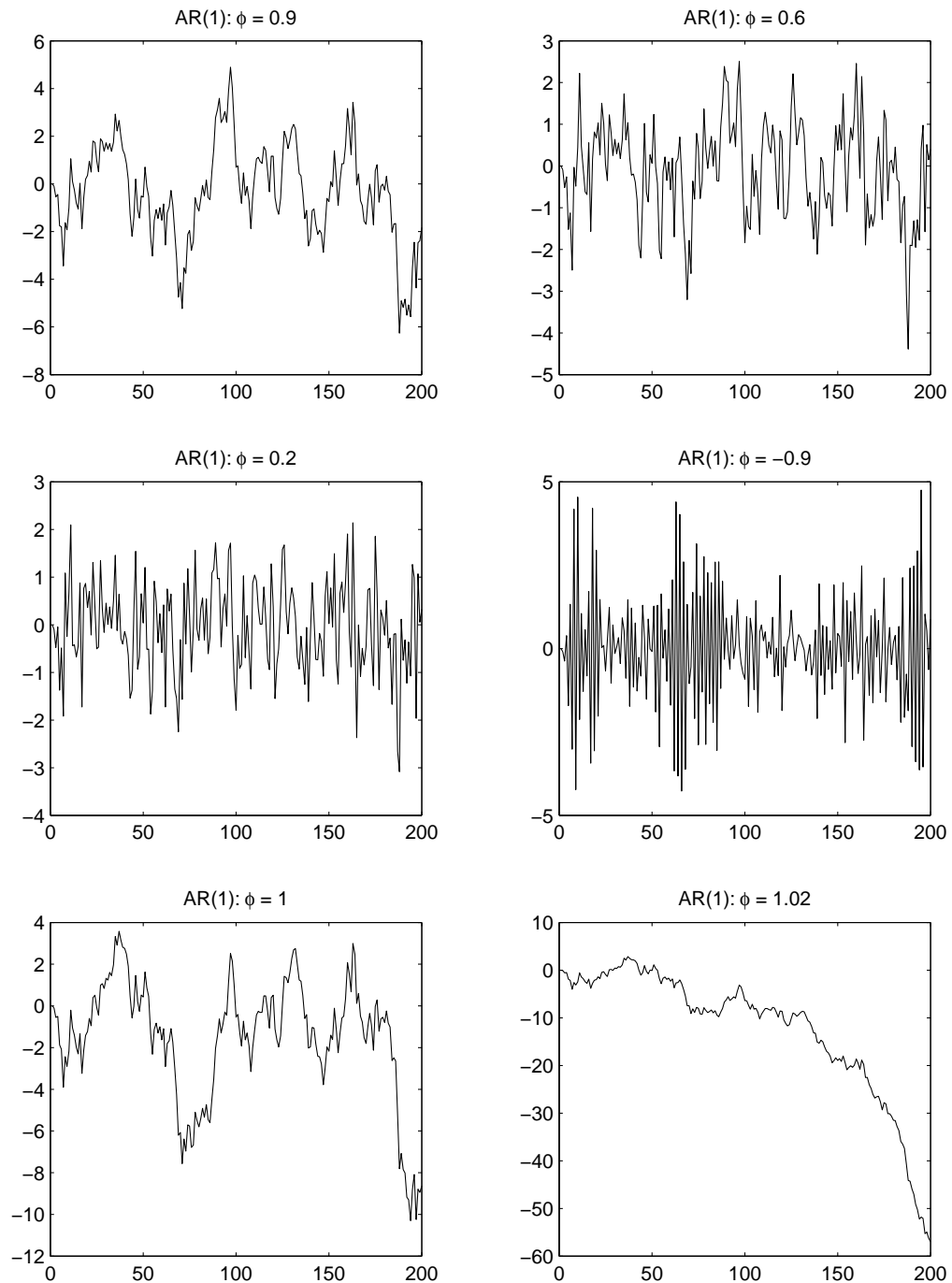


Figure 4.1: Simulations of 200 observations from AR(1) processes with various values of  $\phi$  and  $\mu = 0$ . The white noise “residual” or “error” process  $\epsilon_1, \epsilon_2, \dots$  is the same for all six AR(1) processes.



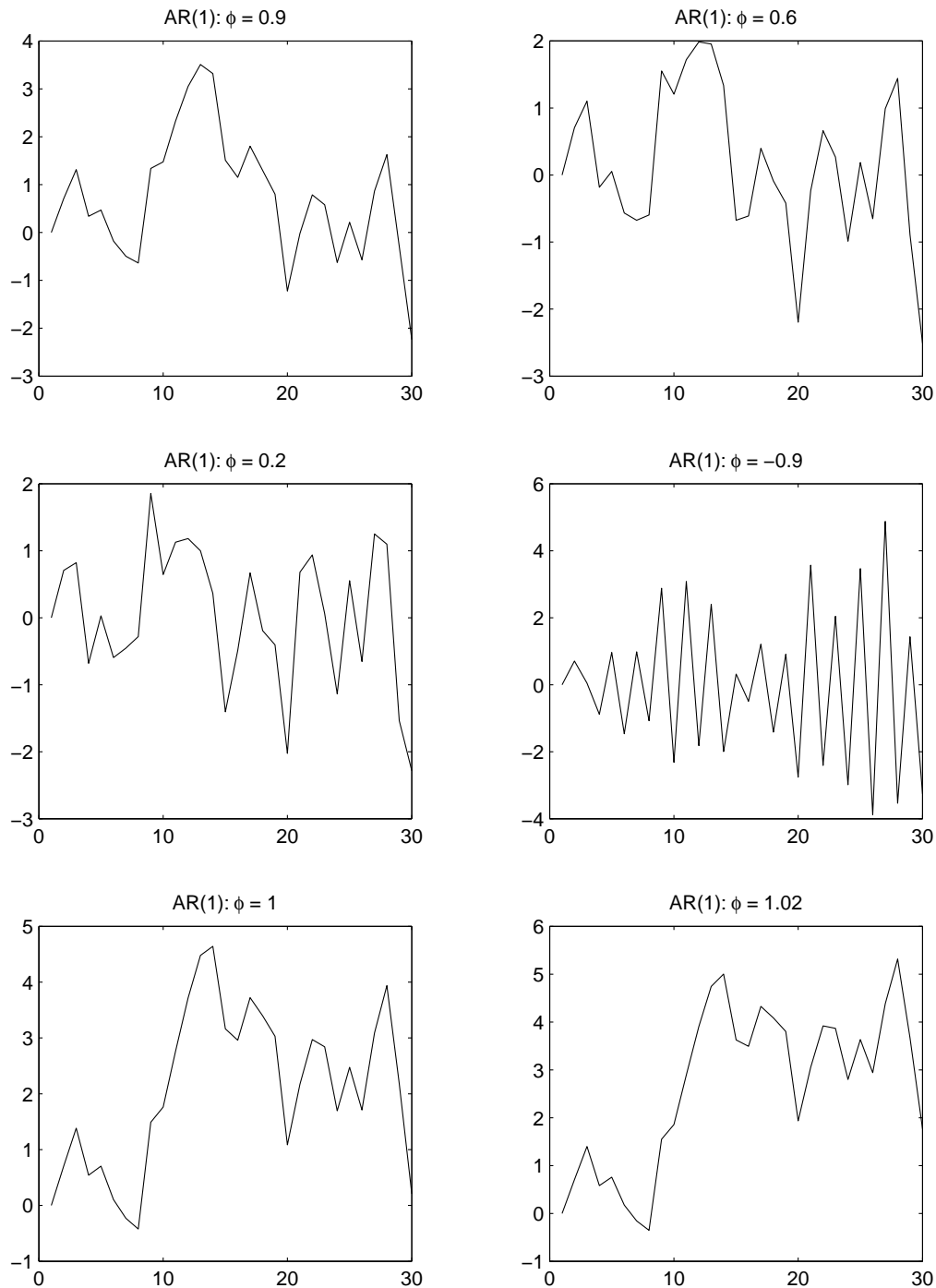


Figure 4.2: Simulations of 30 observation from AR(1) processes with various values of  $\phi$  and  $\mu = 0$ . The white noise “residual” or “error” process  $\epsilon_1, \epsilon_2, \dots$  is the same for all six AR(1) processes.

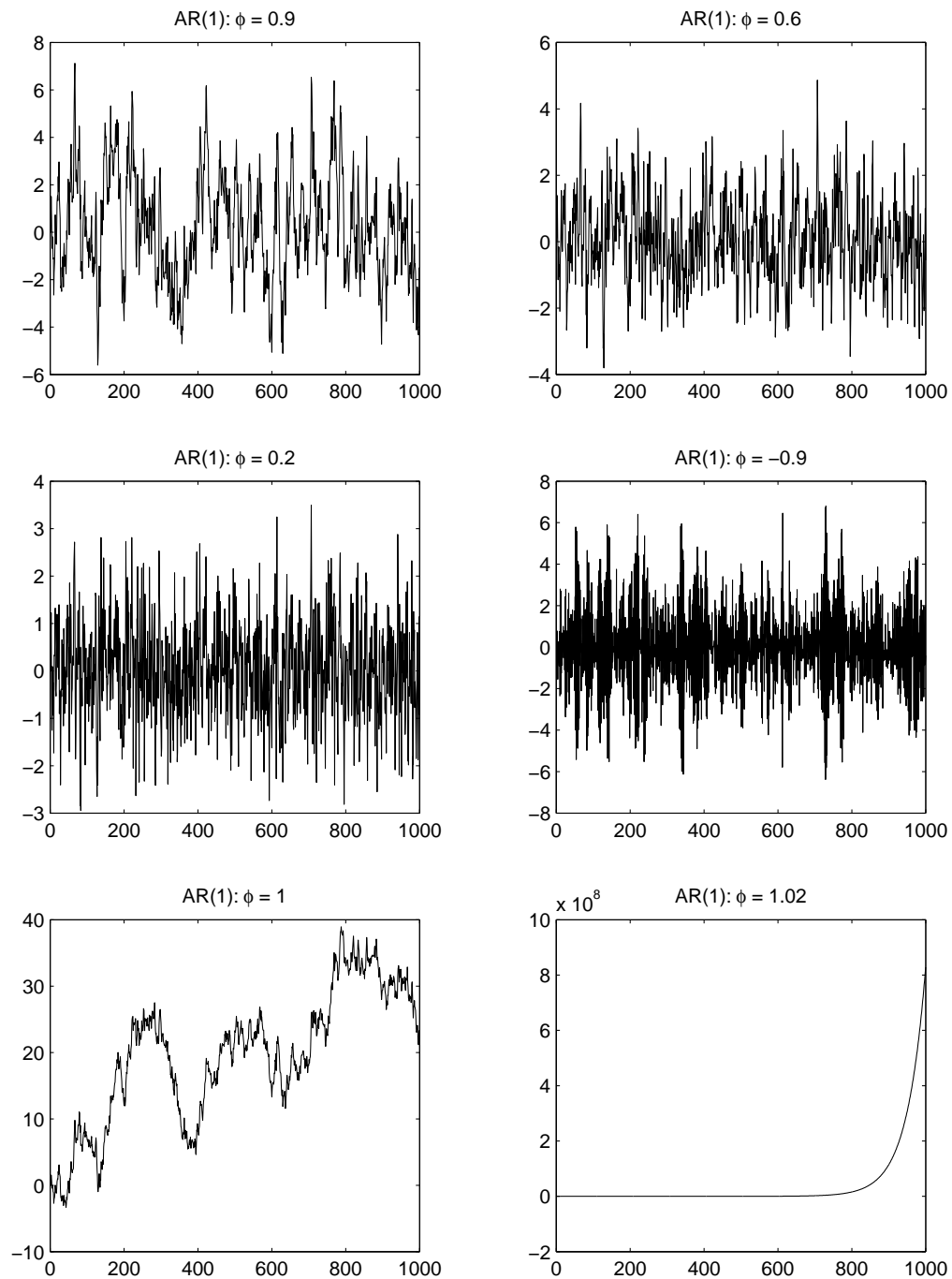


Figure 4.3: Simulations of 1000 observation from AR(1) processes with various values of  $\phi$  and  $\mu = 0$ . The white noise “residual” or “error” process  $\epsilon_1, \epsilon_2, \dots$  is the same for all six AR(1) processes.

### 4.3.3 Estimation

Depending upon the application, one will want to fit an AR(1) to either one of the variables in the raw data or a variable that has been constructed from the raw data. In finance applications, one often has the prices as the raw data but wants to fit an AR(1) to the log returns. To create the log returns, one first log-transforms the prices and then differences the log prices. MINITAB and SAS both have functions to do differencing. For example, in MINITAB, go to the “Stat” menu, then the “Time Series” menu, and then select “differences.” Once a variable containing the log returns has been created, one then can fit an AR(1) model to it.

Let’s assume we have a time series  $y_1, \dots, y_n$  and we want to fit an AR(1) model to this series. Since an AR(1) model is a linear regression model, it can be analyzed using linear regression software. One creates a lagged variable in  $y_t$  and uses this as the “ $x$ -variable” in the regression. MINITAB and SAS both support lagging. For example, in MINITAB, go to the “Stat” menu, then the “Time Series” menu, and then select “lag.”

The least squares estimation of  $\rho$  and  $\mu$  minimize

$$\sum_{t=2}^n [\{y_t - \mu\} - \{\phi(y_{t-1} - \mu)\}]^2.$$

If errors  $(\epsilon_1, \dots, \epsilon_n)$  are *Gaussian* white noise then the least-squares estimate is also the MLE.

Moreover, both MINITAB or SAS have special procedure for fitting AR models.

In MINITAB, go the the “Stat” menu, then the “Time Series” menu, and then choose ARIMA. Use 1 autoregressive parameter, 0 differencing, and 0 moving average parameters.

In SAS, use the “AUTOREG” or the “ARIMA” procedure.

Once  $\phi$  has been estimated, one can calculate the residuals,  $\hat{\epsilon}_1, \hat{\epsilon}_2, \dots, \hat{\epsilon}_n$ ,

defined by

$$\hat{\epsilon}_t = y_t - \hat{\mu} - \hat{\phi}(y_{t-1} - \hat{\mu}).$$

The residuals estimate  $\epsilon_1, \epsilon_2, \dots, \epsilon_n$  and can be used to check the assumption that  $y_1, y_2, \dots, y_n$  is an AR(1) process; any autocorrelation in the residuals is evidence that against the assumption of an AR(1) process.

To test for residual autocorrelation one can use the “test bounds” provided by MINITAB’s or SAS’s autocorrelation plots. One can also use the Ljung-Box test that simultaneously tests that all autocorrelations up to a specified lag are zero.

#### **Example: GE daily returns**

Autoregressive models can be analyzed in both MINITAB and SAS.

The MINITAB output was obtained by running MINITAB interactively. Here is the MINITAB output.

### 4.3. AR(1) PROCESSES

53

2/2/01 10:45:25 AM

Welcome to Minitab, press F1 for help.  
Retrieving worksheet from file: C:\COURSES\OR473\MINITAB\GE\_DAILY.MTW  
# Worksheet was saved on Wed Jan 10 2001

Results for: GE\_DAILY.MTW

ARIMA Model: logR

ARIMA model for logR

Estimates at each iteration

Iteration	SSE	Parameters	
0	2.11832	0.100	0.090
1	0.12912	0.228	0.015
2	0.07377	0.233	0.001
3	0.07360	0.230	0.000
4	0.07360	0.230	-0.000
5	0.07360	0.230	-0.000

Relative change in each estimate less than 0.0010

Final Estimates of Parameters

Type	Coef	SE Coef	T	P
AR 1	0.2299	0.0621	3.70	0.000
Constant	-0.000031	0.001081	-0.03	0.977
Mean	-0.000040	0.001403		

Number of observations: 252

Residuals: SS = 0.0735911 (backforecasts excluded)  
MS = 0.0002944 DF = 250

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	12	24	36	48
Chi-Square	23.0	33.6	47.1	78.6
DF	10	22	34	46
P-Value	0.011	0.054	0.066	0.002

The SAS output comes from running the following program.

```
options linesize = 72 ;
comment Restrict the linesize to 72 characters ;
data ge ; comment Start the data step ;
infile 'c:\courses\or473\data\ge.dat' ;
comment Specify the input data set ;
input close ;
comment Create a new variable ;
D_p = dif(close);
comment Take first differences ;
logP = log(close) ;
logR = dif(logP) ;
comment logR = log returns ;
run ;
title 'GE - Daily prices, Dec 17, 1999 to Dec 15, 2000' ;
title2 'AR(1)' ;
proc autoreg ;
model logR =/nlag = 1 ;
run ;
```



The AUTOREG Procedure					
Yule-Walker Estimates					
SSE	0.07359998	DFE			250
MSE	0.0002944	Root MSE			0.01716
SBC	-1324.6559	AIC			-1331.7148
Regress R-Square	0.0000	Total R-Square			0.0518
Durbin-Watson	1.9326				

Variable	DF	Estimate	Standard Error	t Value	Approx Pr >  t
Intercept	1	-0.000040	0.001394	-0.03	0.9773

From MINITAB we see that  $\hat{\phi} = .2299$  and the estimated standard deviation of  $\hat{\phi}$  is 0.0621. The t-value for testing  $H_0 : \phi = 0$  versus  $H_1 : \phi \neq 0$  is  $.2299/0.0621 = 3.70$  and the p-value is .000 (zero to three decimals). Since the p-value is so small, we reject the null hypothesis.

[Note: Recall from your statistics course that small p-values are significant; we reject the null hypothesis if the p-value is less than  $\alpha$ , e.g., less than .05.]

The null hypothesis is that the log returns are white noise and the alternative is that they are correlated. Thus, we have evidence against the geometric random walk hypothesis. However,  $\phi = .2299$  is not large. Since  $\rho(h) = \phi^h$ , the correlation between successive log returns is .2299 and the squared correlation is only .0528 — only about five percent of the variation in a log return can be predicted by the previous days return.

We have seen that an AR(1) process fits the GE log returns better than a white noise model. Of course, this is not proof that the AR(1) fits these data, only that it fits better than a white noise model. To check that the AR(1) fits well, one looks at the sample autocorrelation function (SACF) of the residuals. A plot of the residual SACF can be requested when using either MINITAB or SAS.



The SACF of the residuals from the GE daily log returns shows high negative autocorrelation at lag 6;  $\hat{\rho}(6)$  is outside the test limits so is “significant” at  $\alpha = .05$ ; see Figure 4.4. This is disturbing.

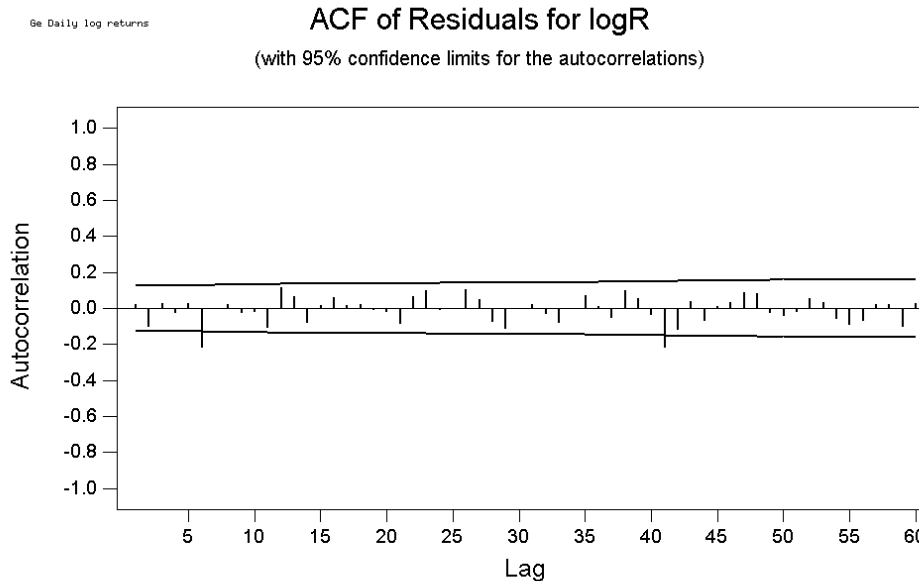


Figure 4.4: SACF of residuals from an AR(1) fit to the GE daily log returns. Notice the large negative residual autocorrelation at lag 6. This is a sign that the AR(1) model does not fit well.

Moreover, the more conservative Ljung-Box “simultaneous” test that  $\rho(1) = \dots = \rho(12) = 0$  has  $p = .011$ . Since the AR(1) model does not fit well, one might consider more complex models. These will be discussed in the following sections.

The SAS estimate of  $\phi$  is  $-.2254$ . SAS uses the model

$$y_t = -\phi y_{t-1} + \epsilon_t$$

so SAS's  $\phi$  is the negative of  $\phi$  as we, and MINITAB, define it. The difference, .2299 versus .2254, between MINITAB and SAS is due to slight variation in the estimation algorithm.

We can also estimate  $\mu$  and test that  $\mu$  is zero. From the MINITAB output, we see  $\hat{\mu}$  is nearly zero, the t-value for testing that  $\mu$  is zero is very small while the p-value is near one. Remember that **small** values of the p-value are significant; since the p-value is **large** we accept the null hypothesis that  $\mu$  is zero.

## 4.4 AR(p) models

$y_t$  is an AR( $p$ ) process if

$$(y_t - \mu) = \phi_1(y_{t-1} - \mu) + \phi_2(y_{t-2} - \mu) + \cdots + \phi_p(y_{t-p} - \mu) + \epsilon_t$$

where  $\epsilon_1, \dots, \epsilon_n$  is WN( $0, \sigma_\epsilon^2$ ).

This is a multiple linear regression model with lagged values of the time series as the "x-variables." The model can be reexpressed as

$$y_t = \beta_0 + \phi y_{t-1} + \cdots + \phi_p y_{t-p} + \epsilon_t,$$

where  $\beta_0 = \{1 - (\phi_1 + \cdots + \phi_p)\}\mu$ .

The least-squares estimator minimizes

$$\sum_{t=p+1}^n \{y_t - (\beta_0 + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p})\}^2.$$

The least-squares estimator can be calculated using a multiple linear regression program but one must create "x-variables" by lagging the time series with lags 1 through  $p$ . It is easier to use the ARIMA command in MINITAB or SAS or SAS's AUTOREG procedure; these procedures do the lagging automatically.

### 4.4.1 Example: GE daily returns

The SAS program shown above was rerun with

```
model logR =/nlag = 1
```

replaced by

```
model logR =/nlag = 6 .
```

The output is on the course's web site as "GE DAILY, AR(6) (SAS)."

The autoregression coefficients (the  $\phi_i$ ) are "significant" at lags 1 and 6 but not at lags 2 through 5. Here "significant" means at  $\alpha = .05$  which corresponds to absolute t-value bigger than 2. MINITAB will not allow  $p > 5$  but SAS does not have such a constraint.

## 4.5 Moving Average (MA) Processes

### 4.5.1 MA(1) processes

The moving average process of order [MA(1)] is

$$y_t - \mu = \epsilon_t - \theta\epsilon_{t-1},$$

where as before the  $\epsilon_t$ 's are  $WH(0, \sigma_\epsilon^2)$ .

One can show that

$$E(y_t) = \mu,$$

$$\text{Var}(y_t) = \sigma_\epsilon^2(1 + \theta^2),$$

$$\gamma(1) = \theta\sigma_\epsilon^2,$$

$$\gamma(h) = 0 \text{ if } |h| > 1,$$

$$\rho(1) = \frac{-\theta}{1 + \theta^2},$$

and

$$\rho(h) = 0 \text{ if } |h| > 1.$$

### 4.5.2 General MA processes

The MA( $q$ ) process is

$$y_t - \mu = \epsilon_t - \theta_1 \epsilon_{t-1} - \cdots - \theta_q \epsilon_{t-q}.$$

One can show that  $\gamma(h) = 0$  and  $\rho(h) = 0$  if  $|h| > q$ .

## 4.6 ARIMA Processes

Stationary time series with complex autocorrelation behavior are better modeled by mixed autoregressive and moving average (ARMA) processes than by either a pure AR or pure MA process. ARIMA (autoregressive, integrated, moving average) processes are based on ARMA processes and are models for nonstationary time series.

ARIMA processes are more easily described if we introduce the “backwards” operator,  $B$ .

### 4.6.1 The backwards operator

The backwards operator  $B$  is defined by

$$B y_t = y_{t-1}$$

and, more generally,

$$B^k y_t = y_{t-k}.$$

Note that  $Bc = c$  for any constant  $c$  since a constant does not change with time.

### 4.6.2 ARMA Processes

The ARMA( $p, q$ ) process satisfies the equation

$$(1 - \phi B - \cdots - \phi_p B^p)(y_t - \mu) = (1 - \theta_1 B - \cdots - \theta_q B^q)\epsilon_t.$$

A white noise process is ARMA(0,0).

### 4.6.3 The differencing operator

The differencing operator is  $\Delta = 1 - B$  so that

$$\Delta y_t = y_t - B y_t = y_t - y_{t-1}.$$

Thus, differencing a time series produces a new time series consisting of the changes in the original series. For example, if  $p_t = \log(P_t)$  is the log price, then the log return is

$$r_t = \Delta p_t.$$

Differencing can be iterated. For example,

$$\begin{aligned} \Delta^2 y_t = \Delta(\Delta y_t) &= \Delta(y_t - y_{t-1}) = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) \\ &= y_t - 2y_{t-1} + y_{t-2}. \end{aligned}$$

### 4.6.4 From ARMA processes to ARIMA process

Often the first or second differences of nonstationary time series are stationary. For example, the first differences of random walk (nonstationary) are white noise (stationary).

A time series  $y_t$  is said to be ARIMA( $p, d, q$ ) if  $\Delta^d y_t$  is ARMA( $p, q$ ). Also, if log returns ( $r_t$ ) on an asset are ARMA( $p, q$ ), then the log prices ( $p_t$ ) are ARIMA( $p, 1, q$ ).

The ARIMA procedures in MINITAB and SAS allow one to specify  $p, d$ , and  $q$ .

Notice that an ARIMA( $p, 0, q$ ) model is the same as an ARMA( $p, q$ ) model. ARIMA( $(p, 0, 0)$ , ARMA( $p, 0$ ), and AR( $p$ ) models are the same. Also, ARIMA( $(0, 0, q)$ , ARMA( $0, q$ ), and MA( $q$ ) models are the same. A random walk is an ARIMA( $0, 1, 0$ ) model. Why?

The inverse of differencing is “integrating.” The integral of a process  $y_t$  is the process  $w_t$  where

$$w_t = w_{t_0} + y_{t_0} + y_{t_0+1} + \cdots + y_t.$$

where  $t_0$  is an arbitrary starting time point and  $w_{t_0}$  is the starting value of the  $w_t$  process.

Figure 4.5 shows an AR(1), its “integral” and its “second integral,” meaning the integral of its integral.

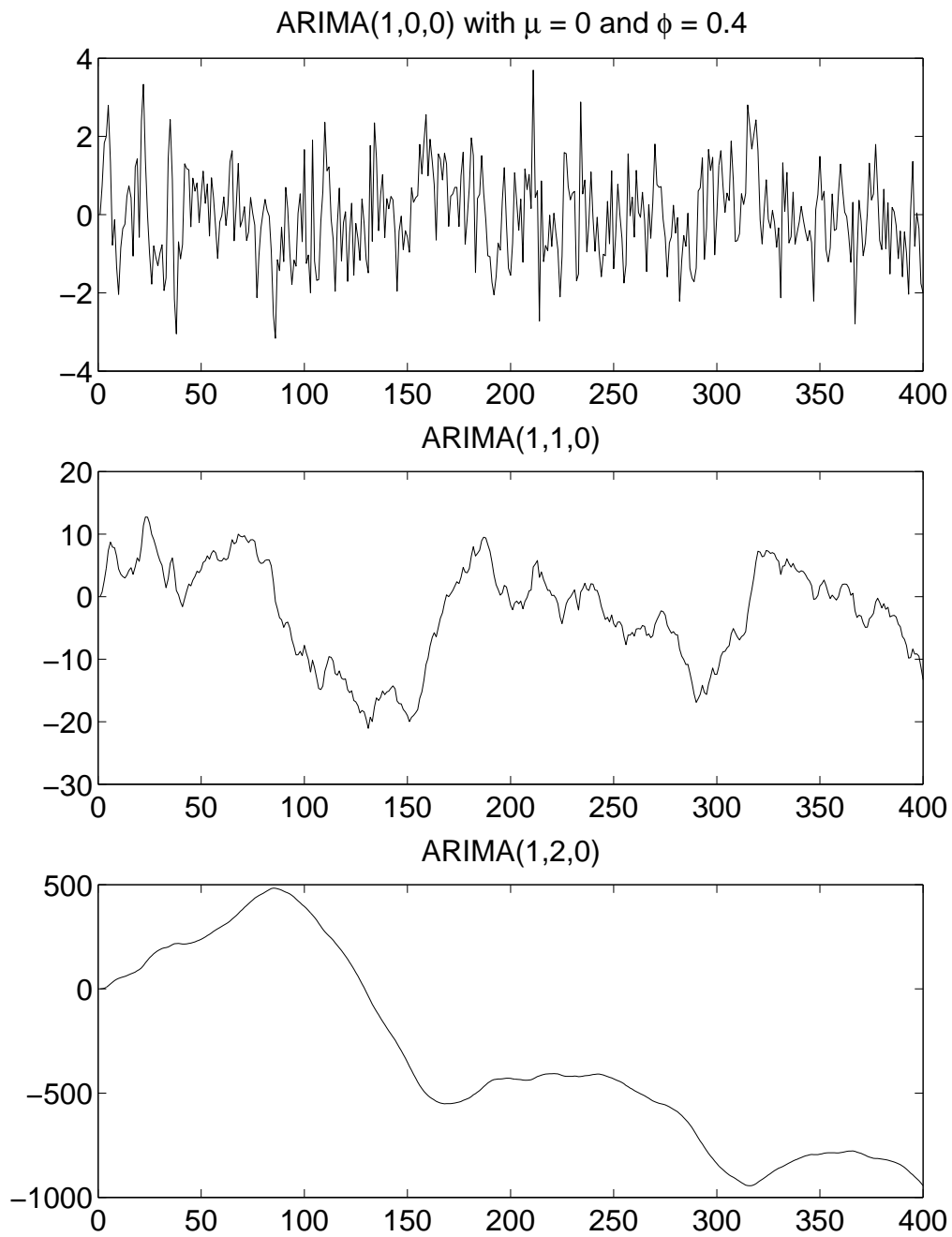


Figure 4.5: The top plot is of an AR(1) process with  $\mu = 0$  and  $\phi = 0.4$ . The middle and bottom plots are, respectively, the integral and second integral of this AR(1) process.

## 4.7 Model Selection

Once the parameters  $p$ ,  $d$ , and  $q$  of an ARIMA process have been selected, the AR and MA coefficients can be estimated by maximum likelihood. But how do we choose  $p$ ,  $d$ , and  $q$ ?

Generally,  $d$  is either 0, 1, or 2 and is chosen by looking at the SACF of  $y_t$ ,  $\Delta y_t$ , and  $\Delta^2 y_t$ .

A sign that a process is nonstationary is that its SACF decays to zero very slowly. If this is true of  $y_t$  then the original series is nonstationary and should be differenced at least once.

If the SACF of  $\Delta y_t$  looks stationary then we use  $d = 1$ . Otherwise, we look at the SACF of  $\Delta^2 y_t$ ; if this looks stationary we use  $d = 2$ .

I have never seen a real time series where  $\Delta^2 y_t$  did not look stationary, but if one were encountered then  $d > 2$  would be used.

Once  $d$  has been chosen, we know that we will fit an ARMA( $p, q$ ) process to  $\Delta^d y_t$ , but we still need to select  $p$  and  $q$ . This can be done by comparing various choices of  $p$  and  $q$  by some criterion that measures how well a model fits the data.

### 4.7.1 AIC and SBC

AIC and SBC are model selection criteria based on the log-likelihood.

Akaike's information criterion (AIC) is defined as

$$-2 \log(L) + 2(p + q),$$

where  $L$  is the likelihood evaluated at the MLE.

Schwarz's Bayesian Criterion (SBC) is also called the Bayesian Information



Criterion (BIC) and is defined as

$$-2 \log(L) + 2 \log(n)(p + q),$$

where  $n$  is the length of the time series.

The “best” model according to either criterion is the model that minimizes that criterion.

Either criteria will tend to select models with large values of the likelihood; this makes perfect sense since a large value of  $L$  means that the observed data are likely under that model.

The term  $2(p + q)$  in AIC or  $\log(n)(p + q)$  is a penalty on having too many parameters. Therefore, AIC and SBC both try to tradeoff a good fit to the data measured by  $L$  with the desire to use as few parameters as possible.

Note that  $\log(n) > 2$  if  $n \geq 8$ . Since most time series are much longer than 8, SBC penalizes  $p + q$  more than AIC. Therefore, AIC will tend to choose models with more parameters than SBC. Compared to SBC, with AIC the tradeoff is more in favor of a large value of  $L$  than a small value of  $p + q$ .

This difference between AIC and SBC is due to the way they were designed. AIC is designed to select the model that will predict best and is less concerned with having a few too many parameters. SBC is designed to select the true values of  $p$  and  $q$  exactly.

In practice the best AIC model is usually close to the best SBC model; often they are the same model.

Two model can be compared by likelihood ratio testing when one model is “bigger” than the other. Therefore, AIC and SBC are closely connected with likelihood ratio tests.

### 4.7.2 Stepwise regression applied to AR processes

Stepwise regression is a way of looking at a variety of regression models to see which ones fit the data well. You may encounter stepwise regression if you take an advanced regression course. In backwards regression, sometimes called backstepping, one starts with all possible x-variables and eliminates them one at a time until all remaining variables are “significant” by some criterion.

Stepwise regression can, of course, be applied to AR models since these are a type of multiple regression model. SAS’s AUTOREG procedure allows backstepping as an option.

The following SAS program starts with an AR(6) model and backsteps.

```
options linesize = 72 ;
comment Restrict the linesize to 72 characters ;
data ge ; comment Start the data step ;
infile 'c:\courses\or473\data\ge_quart.dat' ;
comment Specify the input data set ;
input close ;
D_p = dif(close);
comment Take first differences ;
logP = log(close) ;
logR = dif(logP) ;
comment logR = log returns ;
run ;
title 'GE - Quarterly closing prices, Dec 1900 to Dec 2000' ;
title2 'AR(6) with backstepping' ;
proc autoreg ;
model logR =/nlag = 6 backstep ;
run ;
```

Here is the SAS output:

```

GE - Quarterly closing prices, Dec 1900 to Dec 2000          1
      AR(6) with backstepping
                                23:32 Tuesday, January 30, 2001

      The AUTOREG Procedure

      Dependent Variable      logR

      Ordinary Least Squares Estimates

      SSE          0.15125546      DFE          38
      MSE          0.00398        Root MSE      0.06309
      SBC          -102.20076     AIC          -103.86432
      Regress R-Square      0.0000      Total R-Square      0.0000
      Durbin-Watson      2.0710

      Variable      DF      Estimate      Standard      t Value      Approx
                   DF      Estimate      Error      t Value      Pr > |t|

      Intercept      1      0.0627      0.0101      6.21      <.0001

      Estimates of Autocorrelations

      Lag      Covariance      Correlation

      0      0.00388      1.000000
      1      -0.00014      -0.036627
      2      -0.00023      -0.059114
      3      0.00152      0.392878
      4      -0.00014      -0.035792
      5      -0.00075      -0.193269
      6      0.000337      0.086919

      Estimates of Autocorrelations

      Lag      -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

      0      |                                     | ***** |
      1      |                                     | *      |
      2      |                                     | *      |
      3      |                                     | ***** |
      4      |                                     | *      |
      5      |                                     | ****  |
      6      |                                     | **    |

```

4.7. MODEL SELECTION

AR(6) with backstepping  
 23:32 Tuesday, January 30, 2001

The AUTOREG Procedure

Backward Elimination of  
 Autoregressive Terms

Lag	Estimate	t Value	Pr >  t
4	0.020648	0.12	0.9058
2	-0.023292	-0.14	0.8921
1	0.035577	0.23	0.8226
6	0.082465	0.50	0.6215
5	0.170641	1.13	0.2655

Preliminary MSE 0.00328

Estimates of Autoregressive Parameters

Lag	Coefficient	Standard Error	t Value
3	-0.392878	0.151180	-2.60

Expected  
 Autocorrelations

Lag	Autocorr
0	1.0000
1	0.0000
2	0.0000
3	0.3929

Yule-Walker Estimates

SSE	0.12476731	DFE	37
MSE	0.00337	Root MSE	0.05807
SBC	-105.5425	AIC	-108.86962
Regress R-Square	0.0000	Total R-Square	0.1751
Durbin-Watson	1.9820		

GE - Quarterly closing prices, Dec 1900 to Dec 2000 3  
 AR(6) with backstepping  
 23:32 Tuesday, January 30, 2001

The AUTOREG Procedure

Variable	DF	Estimate	Standard Error	t Value	Approx Pr >  t
Intercept	1	0.0632	0.0146	4.33	0.0001

Expected  
Autocorrelations

Lag	Autocorr
0	1.0000
1	0.0000
2	0.0000
3	0.3929

### 4.7.3 Using ARIMA in SAS: Cree data

Daily returns of Cree from December 1999 to December 2000 are shown in Figure 4.6.

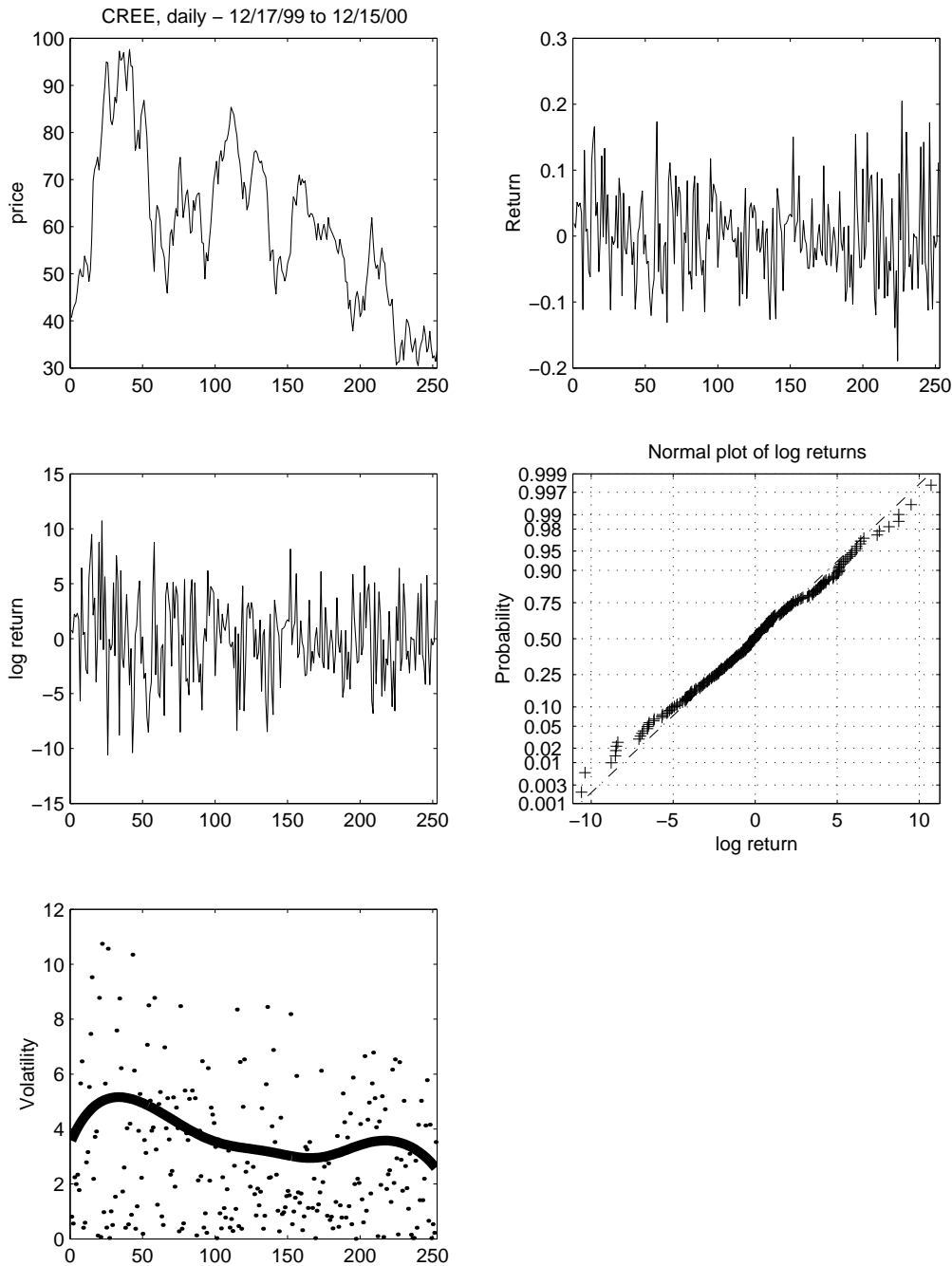


Figure 4.6: Cree daily returns. The first plot is the prices. The second and third are the net returns and the log returns. The fourth plot is a normal probability plot of the log returns. The final plot is of the absolute log returns; there is a scatterplot smooth to help show whether the volatility is constant.

In this example, we will illustrate fitting an ARMA model in SAS. We use daily log returns on Cree from December 1999 to December 2000. The SAS program is:

```
options linesize = 72 ;
data cree ;
infile 'U:\courses\473\data\cree_daily.dat ' ;
input month day year volume high low close ;
logP = log(close) ;
logR = dif(logP) ;
run ;
title Cree daily log returns ;
title2 ARMA(1,1) ;
proc arima ;
identify var=logR ;
estimate p=1 q=1 ;
run ;
```



The “identify” statement specifies the input series and tells SAS to compute the SACF. It can also be used to specify the amount of differencing; “identify var=logP(1) ;” would tell SAS to use the first differences of the log prices as input.

Here is the SAS output. The result is that the Cree log returns appear to be white noise since  $\phi_1$  (denoted by AR1,1 in SAS),  $\theta_1$  (denoted by MA1,1) and  $\mu$  not significantly different from zero.

```

Cree daily log returns                                1
              ARMA(1,1)  15:18 Friday, February 2, 2001

The ARIMA Procedure

Name of Variable = logR

Mean of Working Series      -0.00071
Standard Deviation          0.067473
Number of Observations      252

Autocorrelations

Lag Covariance Correlation -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
0  0.0045526      1.00000 |          | *****|
1  0.00031398     0.06897 |          |  .  *  . |
2 -0.0000160     -0.00351 |          |  .  | .   |
3 -5.5958E-6     -0.00123 |          |  .  | .   |
4 -0.0002213     -0.04862 |          |  .  * | .   |
5  0.00002748     0.00604 |          |  .  | .   |
6 -0.0000779     -0.01712 |          |  .  | .   |
7 -0.0000207     -0.00454 |          |  .  | .   |
8 -0.0003281     -0.07207 |          |  .  * | .   |
9  0.00015664     0.03441 |          |  .  * | .   |
10 0.00057077     0.12537 |          |  .  *** |
11 0.00023632     0.05191 |          |  .  *  . |
12 -0.0003475     -0.07633 |          |  . ** | .   |
13 -0.0001348     -0.02961 |          |  .  * | .   |
14 -0.0005590     -0.12278 |          |  . ** | .   |
15 0.00023425     0.05145 |          |  .  *  . |
16 -0.0001021     -0.02242 |          |  .  | .   |
17 -0.0000582     -0.01278 |          |  .  | .   |
18 -0.0007147     -0.15699 |          |  . *** | .   |
19 0.00006314     0.01387 |          |  .  | .   |
20 -0.0000466     -0.01024 |          |  .  | .   |
21 -0.0001681     -0.03692 |          |  .  * | .   |
22 -0.0001439     -0.03161 |          |  .  * | .   |
23 -0.0002135     -0.04690 |          |  .  * | .   |

```



```

 9      0.04628 |          . | * . |
10      0.11841 |          . | ** . |
11      0.03697 |          . | * . |
12     -0.09207 |         .** | . |
13     -0.01457 |          . | . |
14     -0.11485 |         .** | . |
Cree daily log returns                                3
ARMA(1,1) 15:18 Friday, February 2, 2001

```

The ARIMA Procedure

Partial Autocorrelations

```

Lag      Correlation   -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
15      0.07540 |          . | ** . |
16     -0.04385 |          . | * . |
17      0.00180 |          . | . |
18     -0.16594 |         *** | . |
19      0.05041 |          . | * . |
20     -0.06240 |          . | * . |
21     -0.02732 |          . | * . |
22     -0.05643 |          . | * . |
23      0.00111 |          . | . |
24      0.01957 |          . | . |

```

Autocorrelation Check for White Noise

```

To      Chi-      Pr >
Lag     Square  DF  ChiSq -----Autocorrelations-----
 6      1.91   6 0.9276  0.069 -0.004 -0.001 -0.049  0.006 -0.017
12     10.02  12 0.6143 -0.005 -0.072  0.034  0.125  0.052 -0.076
18     21.95  18 0.2344 -0.030 -0.123  0.051 -0.022 -0.013 -0.157
24     23.37  24 0.4978  0.014 -0.010 -0.037 -0.032 -0.047  0.016

```

Conditional Least Squares Estimation

Parameter	Estimate	Standard Error	t Value	Approx Pr >  t	Lag
MU	-0.0006814	0.0045317	-0.15	0.8806	0
MA1,1	-0.18767	0.88710	-0.21	0.8326	1
AR1,1	-0.11768	0.89670	-0.13	0.8957	1

```

Constant Estimate -0.00076
Variance Estimate 0.004585
Std Error Estimate 0.067712
AIC                -638.889

```

SBC -628.301  
 Number of Residuals 252  
 \* AIC and SBC do not include log determinant.  
 Cree daily log returns 4  
 ARMA(1,1) 15:18 Friday, February 2, 2001

## The ARIMA Procedure

## Correlations of Parameter Estimates

Parameter	MU	MA1,1	AR1,1
MU	1.000	0.005	0.006
MA1,1	0.005	1.000	0.998
AR1,1	0.006	0.998	1.000

## Autocorrelation Check of Residuals

To Lag	Chi- Square	DF	Pr > ChiSq	-----Autocorrelations-----						
6	0.75	4	0.9444	0.000	0.004	0.001	-0.049	0.010	-0.019	
12	8.54	10	0.5761	0.003	-0.075	0.032	0.118	0.050	-0.079	
18	21.12	16	0.1741	-0.014	-0.127	0.062	-0.029	0.001	-0.159	
24	22.48	22	0.4314	0.025	-0.011	-0.035	-0.026	-0.045	0.016	
30	32.65	28	0.2490	0.054	0.127	0.102	-0.023	-0.029	0.070	
36	38.16	34	0.2858	-0.055	-0.038	-0.026	-0.079	0.021	0.083	
42	47.23	40	0.2009	-0.061	-0.092	-0.004	-0.028	-0.118	-0.055	
48	49.15	46	0.3480	-0.032	-0.011	-0.004	0.027	0.054	-0.036	

## Model for variable logR

Estimated Mean -0.00068

## Autoregressive Factors

Factor 1: 1 + 0.11768 B\*\*(1)

## Moving Average Factors

Factor 1: 1 + 0.18767 B\*\*(1)

## 4.8 Example: Three-month Treasury bill rates

The efficient market hypothesis predicts that log returns will be white noise, and our empirical results are that log returns have little autocorrelation even if they are not exactly white noise. Other financial time series do have substantial autocorrelation, as is shown in this example.

The time series in this example is monthly interest rates on three-month US Treasury bills from December 1950 until February 1996. The data come from Example 16.1 of Pindyck and Rubin (1998), *Econometric Models and Economic Forecasts*. The rates are plotted in Figure 4.7. The first differences look somewhat stationary, and we will fit ARMA models to the first differences.

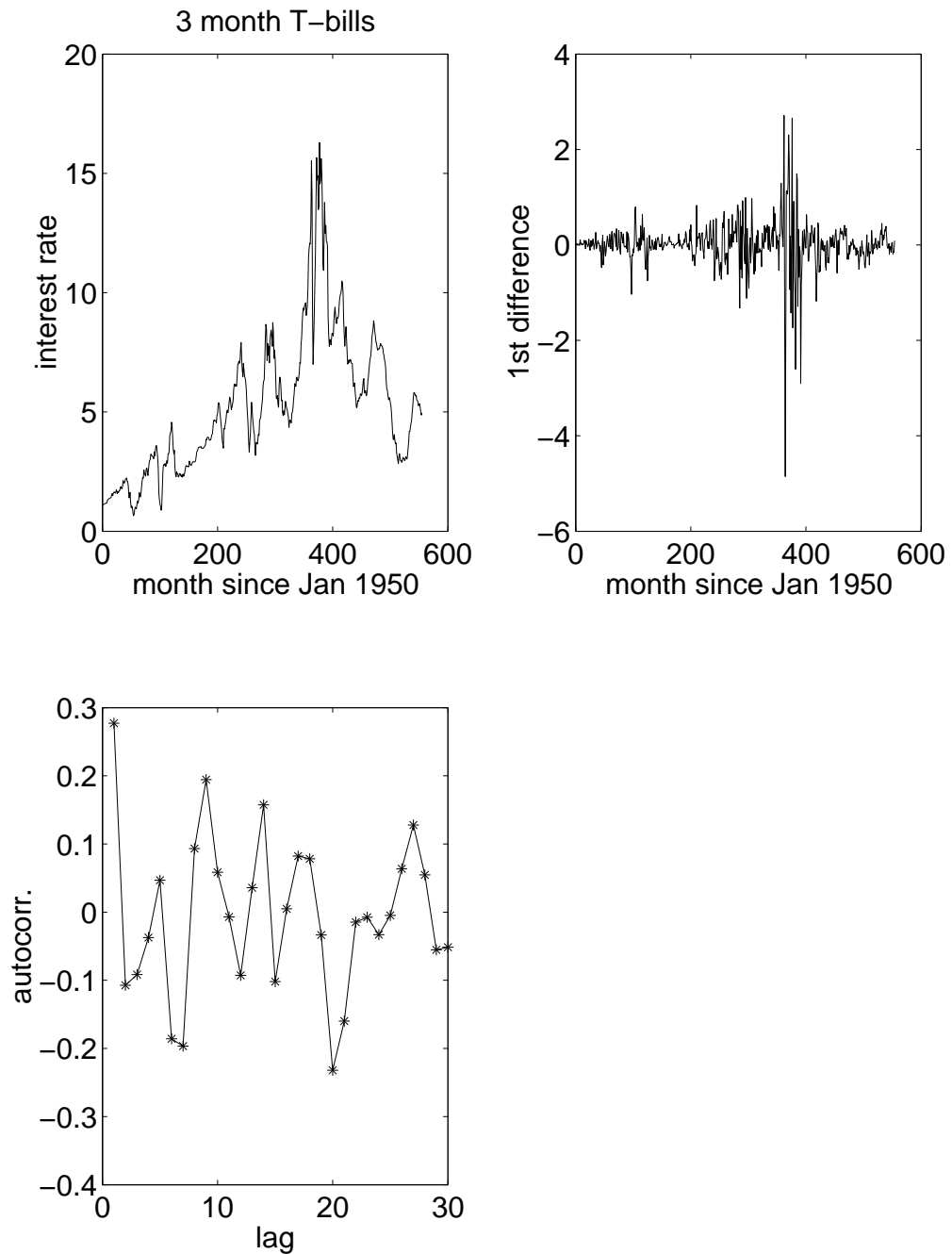


Figure 4.7: Time series plot of 3 month Treasury bill rates, plot of first differences, and sample autocorrelation function of first differences. Monthly values from January 1950 until March 1996.

First we try fitting an AR(10) model with ARIMA. Here is the SAS program. Note statement “identify var=z(1);” specifies that the model should be fit to the first differences of the variance  $z$ ;  $z$  is the interest rate.

```
options linesize = 72 ;
data ratel ;
infile 'c:\courses\or473\data\fygn.dat' ;
input date $ z;
title 'Three month treasury bills' ;
title2 'ARIMA model - to first differences' ;
proc arima ;
identify var=z(1) ;
estimate p=10 plot;
run ;
```

Here is the SAS output.

```

              Three month treasury bills                      1
              ARIMA model - to first differences
                                14:41 Saturday, February 3, 2001

              The ARIMA Procedure

              Name of Variable = z

              Period(s) of Differencing                      1
              Mean of Working Series                        0.006986
              Standard Deviation                          0.494103
              Number of Observations                      554
              Observation(s) eliminated by differencing    1

              Autocorrelations

Lag Covariance Correlation -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
0  0.244138    1.00000 |                                     |*****|
1  0.067690    0.27726 |                                     |*****|
2 -0.026212   -0.10736 |                                     |**|.  |
```





17	0.03736		.		*	.	
18	-0.05356		.		*	.	
19	0.07262		.		*	.	
20	0.03663		.		*	.	
21	0.03580		.		*	.	
22	0.02890		.		*	.	
23	0.00507		.		.	.	
24	0.00765		.		.	.	

Partial Autocorrelations

Lag	Correlation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1	
1	0.27726										.		*****										
2	-0.19958									****		.											
3	-0.00061									.		.											
4	-0.03172									.		*	.										
5	0.05661									.		*	.										
6	-0.25850									*****		.											
7	-0.05221									.		*	.										
8	0.14071									.		***											
9	0.08439									.		**											
10	-0.04699									.		*	.										
11	0.06148									.		*	.										
12	-0.11389									**		.											
13	0.05561									.		*	.										
14	0.13716									.		***											

Three month treasury bills 3

ARIMA model - to first differences

14:41 Saturday, February 3, 2001

The ARIMA Procedure

Partial Autocorrelations

Lag	Correlation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1	
15	-0.13273									***		.											
16	0.15741									.		***											
17	0.02301									.		.											
18	0.01777									.		.											
19	-0.13330									***		.											
20	-0.08447									**		.											
21	-0.07718									**		.											
22	-0.04553									.		*	.										
23	-0.01479									.		.	.										
24	-0.01071									.		.	.										

Autocorrelation Check for White Noise

To Lag	Chi-Square	DF	Pr > ChiSq	-----Autocorrelations-----					
6	75.33	6	<.0001	0.277	-0.107	-0.092	-0.037	0.047	-0.186
12	130.15	12	<.0001	-0.197	0.093	0.194	0.059	-0.007	-0.093
18	158.33	18	<.0001	0.036	0.157	-0.102	0.005	0.082	0.078
24	205.42	24	<.0001	-0.033	-0.232	-0.160	-0.015	-0.008	-0.033

## Conditional Least Squares Estimation

Parameter	Estimate	Standard Error	t Value	Approx Pr >  t	Lag
MU	0.0071463	0.02056	0.35	0.7283	0
AR1,1	0.33494	0.04287	7.81	<.0001	1
AR1,2	-0.16456	0.04501	-3.66	0.0003	2
AR1,3	0.01712	0.04535	0.38	0.7060	3
AR1,4	-0.10901	0.04522	-2.41	0.0163	4
AR1,5	0.14252	0.04451	3.20	0.0014	5
AR1,6	-0.21560	0.04451	-4.84	<.0001	6
AR1,7	-0.08347	0.04522	-1.85	0.0655	7
AR1,8	0.10382	0.04536	2.29	0.0225	8
AR1,9	0.10007	0.04502	2.22	0.0267	9
AR1,10	-0.04723	0.04290	-1.10	0.2714	10

Constant Estimate 0.006585

Variance Estimate 0.198648

Std Error Estimate 0.445699

Three month treasury bills

4

ARIMA model - to first differences

14:41 Saturday, February 3, 2001

## The ARIMA Procedure

AIC 687.6855

SBC 735.1743

Number of Residuals 554

\* AIC and SBC do not include log determinant.

## Correlations of Parameter Estimates

Parameter	MU	AR1,1	AR1,2	AR1,3	AR1,4	AR1,5
MU	1.000	0.001	-0.000	-0.001	-0.001	-0.000
AR1,1	0.001	1.000	-0.315	0.160	-0.020	0.095
AR1,2	-0.000	-0.315	1.000	-0.357	0.166	-0.033
AR1,3	-0.001	0.160	-0.357	1.000	-0.350	0.204
AR1,4	-0.001	-0.020	0.166	-0.350	1.000	-0.375
AR1,5	-0.000	0.095	-0.033	0.204	-0.375	1.000



4	0.0041792	0.02104		.		.	
5	-0.0030362	-.01528		.		.	
6	-0.0061377	-.03090		.*		.	
7	0.0071315	0.03590		.		.*	
8	-0.0001693	-.00085		.		.	
9	-0.0061781	-.03110		.*		.	
10	0.0036055	0.01815		.		.	
11	0.020788	0.10465		.		**	
12	-0.0078818	-.03968		.*		.	
13	-0.015171	-.07637		**		.	
14	0.035240	0.17740		.		****	
15	-0.022934	-.11545		**		.	
16	0.016000	0.08054		.		**	
17	0.0037288	0.01877		.		.	
18	0.0049781	0.02506		.		.*	
19	-0.012221	-.06152		.*		.	
20	-0.029590	-.14896		***		.	
21	-0.015566	-.07836		**		.	
22	-0.0050098	-.02522		.*		.	
23	-0.0048445	-.02439		.		.	
24	-0.0026174	-.01318		.		.	

"." marks two standard errors

Three month treasury bills

6

ARIMA model - to first differences

14:41 Saturday, February 3, 2001

The ARIMA Procedure

Inverse Autocorrelations

Lag	Correlation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1	
1	-0.04462												.*		.								
2	0.02988												.		.*								
3	0.02921												.		.*								
4	-0.04817												.*		.								
5	0.00308												.		.								
6	0.02072												.		.								
7	-0.02134												.		.								
8	-0.01272												.		.								
9	0.01308												.		.								
10	-0.02753												.*		.								
11	-0.10241												**		.								
12	0.03617												.		.*								
13	0.06350												.		.*								
14	-0.16306												***		.								
15	0.12298												.		**								
16	-0.08990												**		.								
17	-0.02141												.		.								
18	-0.00130												.		.								
19	0.04419												.		.*								

20	0.11901		.		**	
21	0.08929		.		**	
22	0.02613		.		*	
23	0.00628		.		.	
24	0.00879		.		.	

Partial Autocorrelations

Lag	Correlation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1	
1	0.00291		.		.		.		.		.		.		.		.		.		.		.
2	-0.01056		.		.		.		.		.		.		.		.		.		.		.
3	0.00351		.		.		.		.		.		.		.		.		.		.		.
4	0.02091		.		.		.		.		.		.		.		.		.		.		.
5	-0.01534		.		.		.		.		.		.		.		.		.		.		.
6	-0.03040		.		.		.		.		.		*		.		.		.		.		.
7	0.03569		.		.		.		.		.		.		*		.		.		.		.
8	-0.00204		.		.		.		.		.		.		.		.		.		.		.
9	-0.02966		.		.		.		.		.		*		.		.		.		.		.
10	0.01926		.		.		.		.		.		.		.		.		.		.		.
11	0.10200		.		.		.		.		.		.		**		.		.		.		.
12	-0.04035		.		.		.		.		.		*		.		.		.		.		.
13	-0.07248		.		.		.		.		.		*		.		.		.		.		.
14	0.17834		.		.		.		.		.		.		****		.		.		.		.

Three month treasury bills 7  
 ARIMA model - to first differences  
 14:41 Saturday, February 3, 2001

The ARIMA Procedure

Partial Autocorrelations

Lag	Correlation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1	
15	-0.13109		.		.		.		.		.		***		.		.		.		.		.
16	0.09936		.		.		.		.		.		.		**		.		.		.		.
17	0.02268		.		.		.		.		.		.		.		.		.		.		.
18	0.00293		.		.		.		.		.		.		.		.		.		.		.
19	-0.05597		.		.		.		.		.		*		.		.		.		.		.
20	-0.13881		.		.		.		.		.		***		.		.		.		.		.
21	-0.10044		.		.		.		.		.		**		.		.		.		.		.
22	-0.02905		.		.		.		.		.		*		.		.		.		.		.
23	-0.00750		.		.		.		.		.		.		.		.		.		.		.
24	-0.00979		.		.		.		.		.		.		.		.		.		.		.

Model for variable z

Estimated Mean 0.007146  
 Period(s) of Differencing 1

## Autoregressive Factors

Factor 1:  $1 - 0.33494 B^{**}(1) + 0.16456 B^{**}(2) - 0.01712 B^{**}(3) +$   
 $0.10901 B^{**}(4) - 0.14252 B^{**}(5) + 0.2156$   
 $B^{**}(6) + 0.08347 B^{**}(7) - 0.10382 B^{**}(8)$   
 $- 0.10007 B^{**}(9) + 0.04723 B^{**}(10)$

The AR(10) model does not fit well. Next we try an AR(24) model with backfitting. Here is the SAS program:

```
options linesize = 72 ;
data ratel ;
infile 'c:\courses\or473\data\fygn.dat' ;
input date $ z;
zdif=dif(z) ;
title 'Three month treasury bills' ;
title2 'AR(24) model to first differences with backfitting' ;

proc autoreg ;
model zdif= / nlag=24 backstep;
run ;
```

Here is the output.

```

              Three month treasury bills                      1
AR(24) model to first differences with backfitting
              10:32 Wednesday, February 14, 2001

              The AUTOREG Procedure

Dependent Variable      zdif

              Ordinary Least Squares Estimates

SSE              135.25253      DFE              553
MSE              0.24458       Root MSE         0.49455
SBC              797.34939     AIC              793.032225
Regress R-Square  0.0000       Total R-Square   0.0000
Durbin-Watson    1.4454

Variable          DF      Estimate      Standard      t Value      Approx
                  DF      Estimate      Error         Pr > |t|

Intercept         1       0.006986    0.0210        0.33         0.7397

              Estimates of Autocorrelations

Lag      Covariance      Correlation
```

0	0.2441	1.000000
1	0.0677	0.277260
2	-0.0262	-0.107364
3	-0.0224	-0.091587
4	-0.00911	-0.037332
5	0.0114	0.046690
6	-0.0453	-0.185710
7	-0.0480	-0.196558
8	0.0227	0.093118
9	0.0474	0.194318

Estimates of Autocorrelations

Lag	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1
0											*****										
1											*****										
2									**												
3								**													
4								*													
5										*											
6							****														
7							****														
8										**											
9										****											

Three month treasury bills 2  
 AR(24) model to first differences with backfitting  
 10:32 Wednesday, February 14, 2001

The AUTOREG Procedure

Estimates of Autocorrelations

Lag	Covariance	Correlation
10	0.0143	0.058501
11	-0.00171	-0.006997
12	-0.0226	-0.092572
13	0.00876	0.035897
14	0.0384	0.157393
15	-0.0249	-0.101930
16	0.00120	0.004923
17	0.0200	0.082117
18	0.0190	0.078001
19	-0.00816	-0.033427
20	-0.0565	-0.231618
21	-0.0389	-0.159520
22	-0.00358	-0.014653
23	-0.00185	-0.007563
24	-0.00806	-0.032995



Estimates of Autocorrelations

Lag	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1	
10												*										
11																						
12									**													
13												*										
14												***										
15								**														
16																						
17												**										
18												**										
19											*											
20								*****														
21								***														
22																						
23																						
24										*												

Three month treasury bills 3  
 AR(24) model to first differences with backfitting  
 10:32 Wednesday, February 14, 2001

The AUTOREG Procedure

Backward Elimination of  
 Autoregressive Terms

Lag	Estimate	t Value	Pr >  t
10	0.007567	0.16	0.8721
23	0.010212	0.22	0.8241
17	0.008951	0.19	0.8492
3	-0.014390	-0.32	0.7496
24	0.015798	0.40	0.6907
13	0.041434	0.92	0.3605
7	0.038880	0.85	0.3964
18	-0.037456	-0.90	0.3702
22	0.042555	1.02	0.3090
20	0.058230	1.31	0.1912
4	0.059903	1.48	0.1389
9	-0.058141	-1.42	0.1562

Preliminary MSE 0.1765

Estimates of Autoregressive Parameters

Standard

Lag	Coefficient	Error	t Value
1	-0.388246	0.040419	-9.61
2	0.200242	0.040438	4.95
5	-0.108069	0.040513	-2.67
6	0.249095	0.039719	6.27
8	-0.103462	0.039668	-2.61
11	-0.102896	0.040278	-2.55
12	0.119950	0.040704	2.95
14	-0.204702	0.040427	-5.06
15	0.223381	0.042441	5.26
16	-0.151917	0.040811	-3.72
19	0.103356	0.038847	2.66
21	0.108074	0.039511	2.74

Three month treasury bills 4  
 AR(24) model to first differences with backfitting  
 10:32 Wednesday, February 14, 2001

The AUTOREG Procedure

Expected  
 Autocorrelations

Lag	Autocorr
0	1.0000
1	0.2840
2	-0.1196
3	-0.0801
4	0.0273
5	0.0656
6	-0.1914
7	-0.1923
8	0.0880
9	0.1549
10	0.0223
11	-0.0229
12	-0.0737
13	0.0767
14	0.1628
15	-0.1000
16	-0.0017
17	0.0685
18	0.0437
19	-0.0638
20	-0.1968
21	-0.1296

Yule-Walker Estimates

SSE	97.7597462	DFE	541
MSE	0.18070	Root MSE	0.42509
SBC	695.767655	AIC	639.644514
Regress R-Square	0.0000	Total R-Square	0.2772
Durbin-Watson	2.0627		

Variable	DF	Estimate	Standard Error	t Value	Approx Pr >  t
Intercept	1	0.006664	0.0192	0.35	0.7289

Three month treasury bills 5  
AR(24) model to first differences with backfitting  
10:32 Wednesday, February 14, 2001

## The AUTOREG Procedure

Expected  
Autocorrelations

Lag	Autocorr
0	1.0000
1	0.2840
2	-0.1196
3	-0.0801
4	0.0273
5	0.0656
6	-0.1914
7	-0.1923
8	0.0880
9	0.1549
10	0.0223
11	-0.0229
12	-0.0737
13	0.0767
14	0.1628
15	-0.1000
16	-0.0017
17	0.0685
18	0.0437
19	-0.0638
20	-0.1968
21	-0.1296

## 4.9 Forecasting

ARIMA models are often used to forecast future values of a time series. Consider first an AR(1) process. Suppose that we have data  $y_1, \dots, y_n$  and estimates  $\hat{\mu}$  and  $\hat{\phi}$ . Then we estimate  $y_{n+1}$  by

$$\hat{y}_{n+1} := \hat{\mu} + \hat{\phi}(y_n - \hat{\mu})$$

and  $y_{t+2}$  by

$$\hat{y}_{n+2} := \hat{\mu} + \hat{\phi}(\hat{y}_{n+1} - \hat{\mu}) = \hat{\phi}\{\hat{\phi}(y_n - \hat{\mu})\},$$

etc. In general,  $\hat{y}_{n+k} = \hat{\mu} + \hat{\phi}^k(y_n - \hat{\mu})$ . If  $\hat{\phi} < 1$  as is expected for a stationary series, then as  $k$  increases the forecasts will decay exponentially fast to  $\hat{\mu}$ .

Forecasting general AR(p) processes is similar. For example, for an AR(2) process

$$\hat{y}_{n+1} := \hat{\mu} + \hat{\phi}_1(y_n - \hat{\mu}) + \hat{\phi}_2(y_{n-1} - \hat{\mu})$$

and

$$\hat{y}_{n+2} := \hat{\mu} + \hat{\phi}_1(\hat{y}_{n+1} - \hat{\mu}) + \hat{\phi}_2(y_n - \hat{\mu}).$$

Forecasting ARMA and ARIMA processes is only slightly more complicated than forecasting AR processes and is discussed in time series courses such as ORIE 563. Moreover, the forecasts can be generated automatically by MINITAB and SAS, so you don't need to know the details in order to forecast.

### 4.9.1 GE daily returns

We have learned that fitting an ARIMA(1,0,0) model to log returns is equivalent to fitting an ARIMA(1,1,0) model to the log prices. Here we will fit both models to the GE daily price data. Figure 4.8 shows the forecasts of the log returns up to 24 days ahead. The forecasts are given in red and 95% confidence limits on the forecasts are shown in blue. The observed time series is plotted in black.

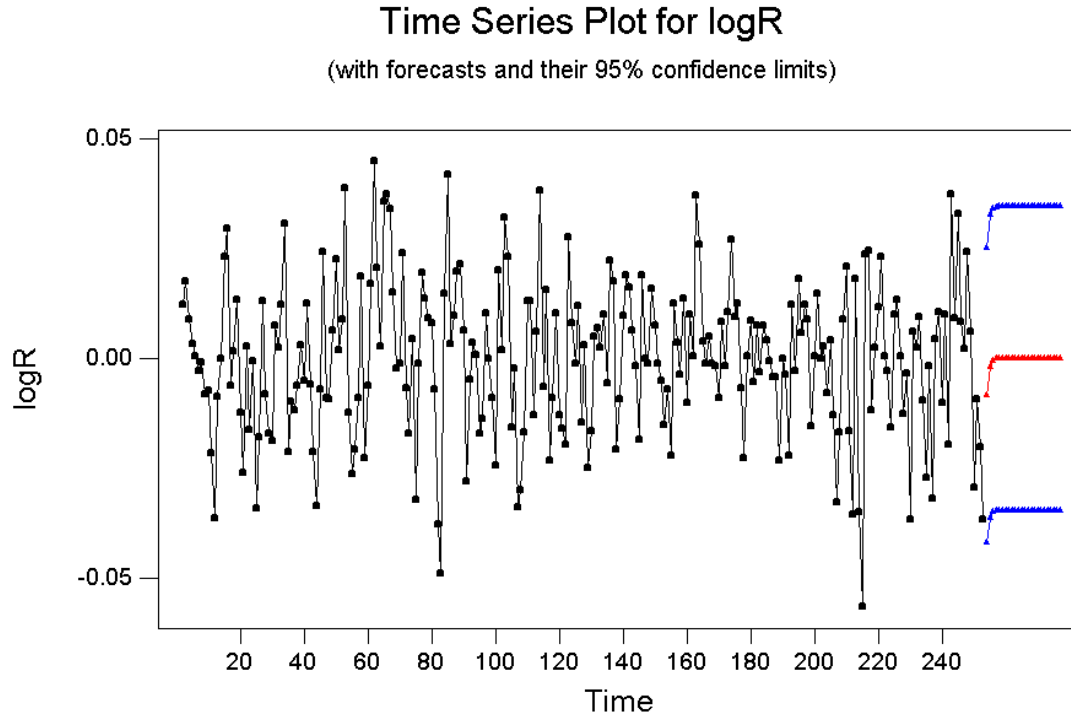


Figure 4.8: *Time series plot of the daily GE log returns with forecasts from an AR(1) model.*

Next we fit an ARIMA(1,1,0) model to the log prices. Although this model is equivalent to the last model, it generates forecasts of the log prices, not the log returns. (MINITAB always forecasts the input series.) The forecasts are given in Figure 4.9. Notice that the forecasts predict that the price of GE will stay constant, but the confidence limits on the forecasts get wider as we forecast further ahead. This is exactly the type of behavior we would expect from a random walk [ARIMA(0,1,0)] model. The ARIMA(1,1,0) model for the log prices isn't quite a random walk model, but it is similar to a random walk model with zero drift ( $\mu = 0$ ) since  $\hat{\phi}$  is close to 0

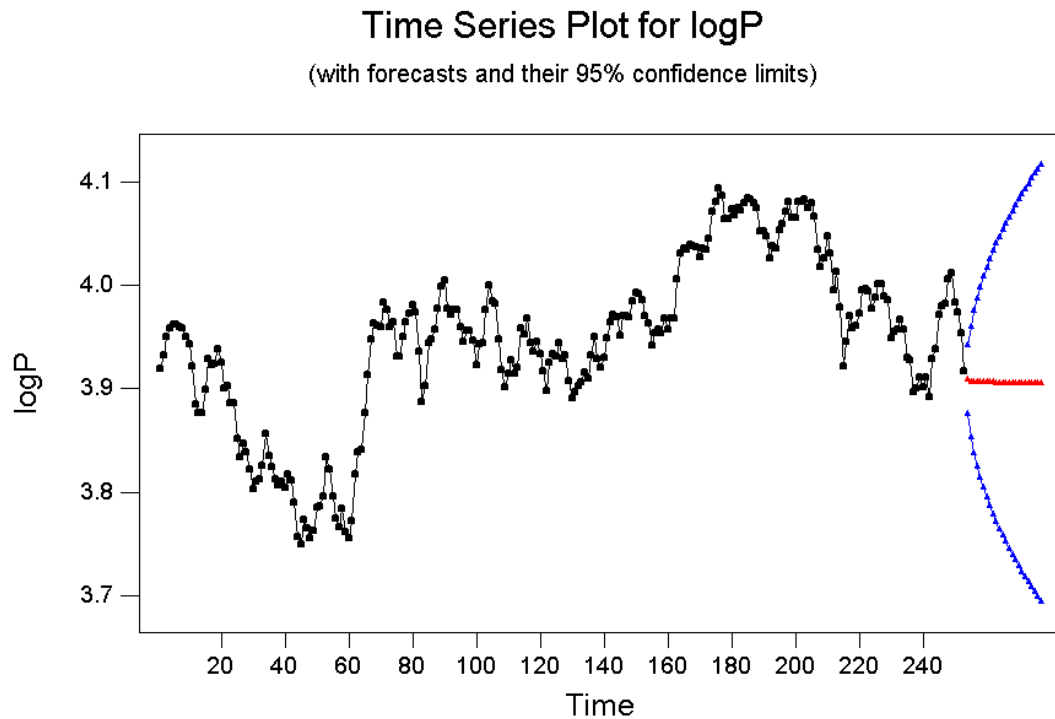


Figure 4.9: Time series plot of the daily GE log prices with forecasts from an  $AR(1)$  model.

and  $\hat{\mu}$  is extremely close to 0.

The forecast limits suggest that accurately forecasting future GE stock prices is pretty hopeless. For practical purposes the log prices behave like a random walk so that the prices behave like a geometric random walk.

# Chapter 5

## Portfolio Selection: 3/12/01

### 5.1 Trading off expected return and risk

How should we invest our wealth? There are two principles:

- we want to maximize the expected return
- we want to minimize the risk = variance of return

These goals are somewhat at odds. Nonetheless, there are optimal compromises between expected return and risk. In this chapter we will see how to maximize expected return subject to an upper bound on the risk, or to minimize the risk subject to a lower bound on the expected return.

The key concept that we will discuss is reduction of risk by diversifying the portfolio of assets held. Diversification was not always considered as favorably as it is now.

#### **The investment philosophy of Keynes**

The famous economist, John Maynard Keynes, did not believe in diversifying a portfolio. He wrote:

... the management of stock exchange investment of any kind is a low pursuit ... from which it is a good thing for most members of society to be free

I am in favor of having as large a unit as market conditions will allow ... to suppose that safety-first consists in having a small gamble in a large number of different [companies] where I have no information to reach a good judgement, as compared with a substantial stake in a company where ones's information is adequate, strikes me as a travesty of investment policy

This quote is taken from Bernstein, *Capital Ideas: The Improbable Origins of Modern Wall Street*.

Keynes is advocating stock picking or "fundamental analysis." But the semi-strong version of the EMH says that fundamental analysis does not lead to economic profit. Of course, Keynes lived well before the EMH and one wonders what Keynes with think about diversification if he were alive now. Modern portfolio theory takes a very different viewpoint than Keynes. This is not to say that Keynes was wrong. Keynes was investing on a long time horizon, and fundamental analysis, if done well, might be very successful in the long run. However, portfolio managers are judged on short-term successes. Also, using fundamental analysis to find bargains is probably more difficult now than in Keynes's time.

## 5.2 One risky asset and one risk-free asset

We will start with a simple example where we have

- one risky asset, which could be a portfolio, e.g., a mutual fund
  - expected return is .15
  - standard deviation of the return is .25
- one risk-free asset, e.g., a 30-day T-bill
  - expected value of the return is .06
  - standard deviation of the return is 0 by definition of "risk-free."



We are faced with the problem of constructing an investment portfolio that we will hold for one time period which could be an hour, a day, a month, a quarter, a year, ten years, etc. At the end of the time period we might want to readjust the portfolio, so for now we are only looking at returns over one time period. Suppose that

- a fraction  $w$  of our wealth is invested in the risky asset
- the remaining fraction  $1 - w$  is invested in the risk-free asset
- then the expected return is  $E(R) = w(.15) + (1 - w)(.06) = .06 + .09w$ .
- the variance of the return is

$$\sigma_R^2 = w^2 (.25)^2 + (1 - w)^2 (0)^2 = w^2 (.25)^2.$$

$$\text{or } \sigma_R = .25 w.$$

Would  $w > 1$  make any sense?

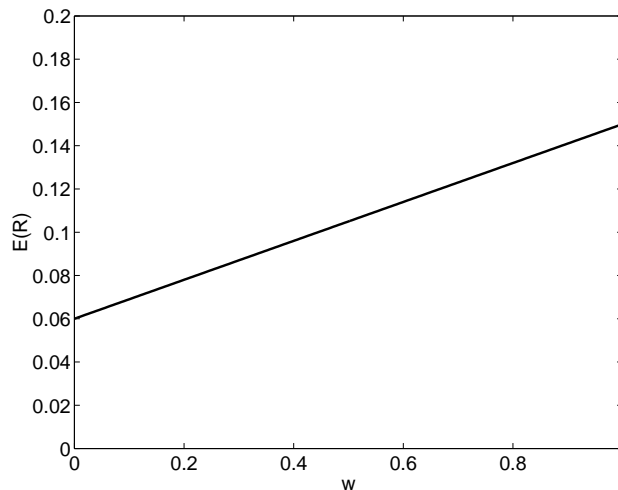


Figure 5.1: Expected return for a portfolio with allocation  $w$  to the risky asset with expected return 0.15 and allocation  $1 - w$  to the risk-free return with return 0.06.

**Question:** Suppose you want an expected return of .10? What should  $w$  be? [answer: 4/9]

**Question:** Suppose you want  $\sigma_R = .05$ . What should  $w$  be? [answer: 0.2]

More generally, if the expected returns on the risky and risk-free assets are  $\mu_1$  and  $\mu_f$  and if the standard deviation of the risky asset is  $\sigma_1$ , then the expected return on the portfolio is  $w\mu_1 + (1 - w)\mu_f$  while the standard deviation of the portfolio's return is  $w\sigma_1$ .

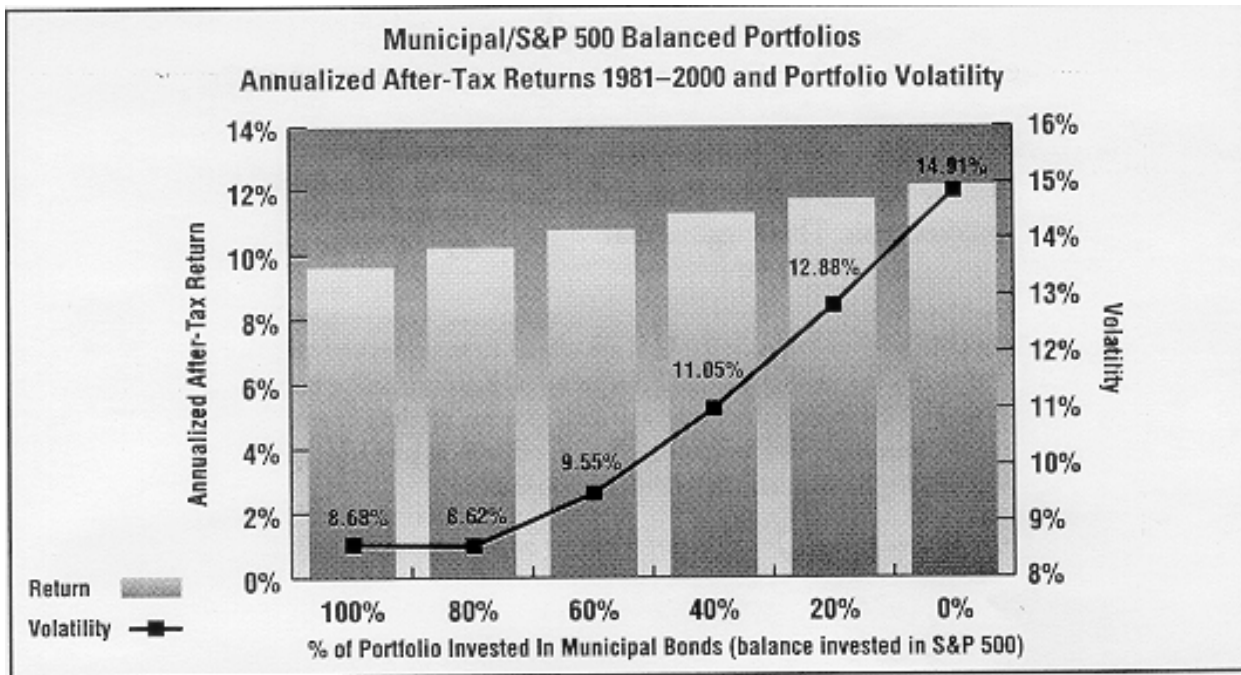
This model is simple but not as useless as it might seem at first. Finding an optimal portfolio can be achieved in two steps.

1. finding the "optimal" portfolio of risky assets, called the "tangency portfolio"
2. finding the appropriate mix of the risk-free asset and the tangency portfolio from step one

So we now know how to do the second step. What we need to learn is how to mix optimally a number of risky assets; we will do that in the next sections. First, we look at a related example.

### 5.2.1 Example

In the February 2001 issue of PaineWebber's *Investment Intelligence: A Report for Our Clients*, the advantages of holding municipal bonds are touted. PaineWebber says "The chart at the right shows that a 20% municipal/80% S&P 500 mix sacrificed only 0.42% annual after-tax return relative to a 100% S&P 500 portfolio, while reducing risk by 13.6% from 14.91% to 12.88%. The chart is show here as Figure 5.2. Although PaineWebber's point is correct, the chart is cleverly designed to over-emphasize the reduction in volatility; how?



Source: Nuveen Investments, "Two Is Greater Than One," January 2001.  
Past performance is no guarantee of future results.

Figure 5.2: Chart from PaineWebber newsletter showing reduction in volatility by mixing municipal bonds with the S&P 500 index.

### 5.2.2 Estimating $E(R)$ and $\sigma_R$

The risk-free rate,  $\mu_f$ , will be known; Treasury bill rates are published in most newspapers.

What should we use as the values of  $E(R)$  and  $\sigma_R$ ? If returns on the asset are assumed to be stationary, then we can take a time series of past returns and use the sample mean and standard deviation. Whether the stationarity assumption is realistic or not is always debatable. If we think that  $E(R)$  and  $\sigma_R$  will be different than in the past, we could subjectively adjust these estimates upward or downward according to our opinions, but we must live with the consequences if our opinions prove to be incorrect.

Another question is how long a time series to use, that is how far back in time when should gather data. A long series, say 10 or 20 years, will give much less variable estimates. However, if the series is not stationary but rather has slowly drifting parameters, then a shorter series (maybe 1 or 2 years) will be more representative of the future.

## 5.3 Two risky assets

The mathematics of mixing risky assets is most easily understood when there are only two risky assets. This is where we will start.

Suppose the two risky assets have returns  $R_1$  and  $R_2$  and that we mix them in proportions  $w$  and  $1 - w$ , respectively. The return is  $R = wR_1 + (1 - w)R_2$ . The expected return on the portfolio is  $E(R) = w\mu_1 + (1 - w)\mu_2$ . Let  $\rho_{12}$  be the correlation between the returns on the two risky assets. The variance of the return on the portfolio is

$$\sigma_R^2 = w^2\sigma_1^2 + (1 - w)^2\sigma_2^2 + 2w(1 - w)\rho_{12}\sigma_1\sigma_2.$$

[Note:  $\sigma_{R_1, R_2} = \rho_{12}\sigma_1\sigma_2$ .]

**Example:**

If  $\mu_1 = .14$ ,  $\mu_2 = .08$ ,  $\sigma_1 = .2$ ,  $\sigma_2 = .15$ , and  $\rho_{12} = 0$ , then

$$E(R) = .08 + .06w.$$

Also, because  $\rho_{12} = 0$

$$\sigma_R^2 = (.2)^2 w^2 + (.15)^2 (1 - w)^2.$$

Using differential calculus, one can easily show that the portfolio with the minimum risk is  $w = .045/.125 = .36$ . For this portfolio  $E(R) = .08 + (.06)(.36) = .1016$  and  $\sigma_R = \sqrt{(.2)^2(.36)^2 + (.15)^2(.64)^2} = .12$ . Here are values of  $E(R)$  and  $\sigma_R$  for some other values of  $w$ :

$w$	$E(R)$	$\sigma_R$
0	.080	.150
1/4	.095	.123
1/2	.110	.125
3/4	.125	.155
1	.140	.200

The somewhat parabolic curve in Figure 5.3 is the locus of values of  $(\sigma_R, E(R))$  when  $0 \leq w \leq 1$ . The points labeled  $R_1$  and  $R_2$  corresponds to  $w = 1$  and  $w = 0$ , respectively. The other features of this figure will be explained in the next section.

### 5.3.1 Estimating means, standard deviations, and covariances

Estimates of  $\mu_1$  and  $\sigma_1$  can be obtained from a univariate times series of past returns on the first risky asset; denote this time series by  $R_{1,1}, \dots, R_{1,n}$  where the first subscript indicates the asset and the second subscript is for time. Let  $\bar{R}_1$  and  $s_{R_1}$  be the sample mean and standard deviation of this series. Similarly,  $\mu_2$  and  $\sigma_2$  can be estimated from a time series of past returns on the second risky asset. The covariance  $\sigma_{12}$  can be estimated by sample covariance

$$\hat{\sigma}_{12} = n^{-1} \sum_{t=1}^n (R_{1,t} - \bar{R}_1)(R_{2,t} - \bar{R}_2).$$

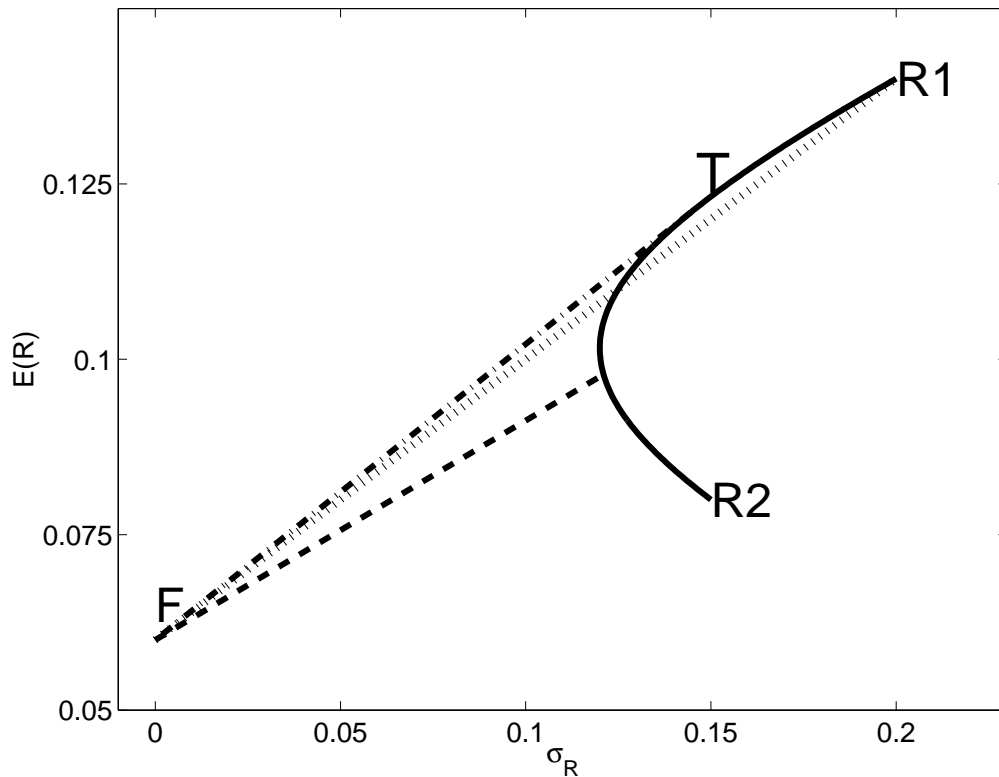


Figure 5.3: *Expected return versus risk. The parabola is the locus of portfolios combining the two risky assets. The lines are the locus of portfolios of two risky assets and the risk-free asset. F = risk-free asset. T = tangency portfolio. R<sub>1</sub> is the first risky asset. R<sub>2</sub> is the second risky asset.*

The correlation  $\rho_{12}$  can be estimated by the sample correlation

$$\hat{\rho}_{12} = \frac{\hat{\sigma}_{12}}{s_1 s_2}.$$

$\hat{\rho}_{12}$ , sometimes denoted by  $r_{12}$ , is called the cross-correlation coefficient between  $R_1$  and  $R_2$  at lag 0, since we are correlating the return on the first risky asset with the return on the second during the same time periods. Cross-correlations at other lags can be defined but are not needed here. In fact, we can define a cross-correlation function, which is a function of lag. The cross-correlation function plays an important role in the analysis of

multivariate time series.

Sample correlations and covariances can be computed on MINITAB. Go to “Stat,” then “Basic statistics,” and then “Correlation” or “Covariance.”

## 5.4 Combining two risky assets with a risk-free asset

As mentioned at the end of the last section, each point on the parabola in Figure 5.3 is  $(\sigma_R, E(R))$  for some value of  $w$  between 0 and 1. If we fix  $w$ , then we have a fixed portfolio of the two risky assets. Now let us mix that portfolio of risky assets with the risk-free asset. The point F in Figure 5.3 gives  $(\sigma_R, E(R))$  for the risk-free asset; of course  $\sigma_R = 0$  at F. The possible values of  $(\sigma_R, E(R))$  for a portfolio consisting of the fixed portfolio of two risky assets and the risk-free asset is a line connecting the point F with a point on the parabola, e.g., the dashed line. The dotted line connecting F with  $R_1$  mixes the risk-free asset with the first risky asset.

Notice that the dotted line lies above the dashed line. This means that for any value of  $\sigma_R$ , the dotted line gives a higher expected return than the dashed line. The slope of any line is called the “Sharpe ratio” of the line; it is named after William Sharpe whom we have met before in Section 3.8 and will meet again in Chapter 6. Sharpe’s ratio can be thought of as a “reward-to-risk” ratio. It is the ratio of the “excess expected return” to the risk as measure by the standard deviation.

Clearly, the bigger the Sharpe ratio the better. Why? The point T on the parabola represents the portfolio with the highest Sharpe ratio. It is the optimal portfolio for the purpose of mixing with the risk-free asset. This portfolio is called the “tangency portfolio” since its line is tangent to the parabola.

**Key result:** The optimal or “efficient” portfolios mix the tangency portfolio of two risky assets with the risk-free asset. Each efficient portfolio has

two properties:

- it has a higher expected return than any other portfolio with the same (or smaller) risk
- it has a smaller risk than any other portfolio with the same (or smaller) expected return.

Thus we can only improve (reduce) the risk of an efficient portfolio by accepted a worse (smaller) expected return, and we can only improve (increase) the expected return of an efficient portfolio by accepting worse (higher) risk.

Note that all efficient portfolios use the same mix of the two risky assets, namely the tangency portfolio. Only the proportion allocated to the tangency portfolio and the proportion allocated to the risk-free asset vary.

### 5.4.1 Tangency portfolio with two risky assets

Given the importance of the tangency portfolio, you may be wondering “how do we find it?”

Again let  $\mu_1$ ,  $\mu_2$ , and  $\mu_f$  be the expected returns on the two risky assets and the return on the risk-free asset. Let  $\sigma_1$  and  $\sigma_2$  be the standard deviations of the returns on the two risky assets and let  $\rho_{12}$  be the correlation between the returns on the risky assets.

Define  $V_1 = \mu_1 - \mu_f$  and  $V_2 = \mu_2 - \mu_f$ ;  $V_1$  and  $V_2$  are called the “excess returns.” Then the tangency portfolio uses weight

$$w_T = \frac{V_1\sigma_2^2 - V_2\rho_{12}\sigma_1\sigma_2}{V_1\sigma_2^2 + V_2\sigma_1^2 - (V_1 + V_2)\rho_{12}\sigma_1\sigma_2}. \quad (5.1)$$

This formula will be derived in Section 5.6.5.

The tangency portfolio allocates a fraction  $w_T$  of the investment to the first risky asset and  $(1 - w_T)$  to the second risky asset.



#### 5.4. COMBINING TWO RISKY ASSETS WITH A RISK-FREE ASSET 105

Let  $R_T$ ,  $E(R_T)$ , and  $\sigma_T$  be the return, expected return, and standard deviation of the return on the tangency portfolio.

**Example:** Suppose as before that  $\mu_1 = .14$ ,  $\mu_2 = .08$ ,  $\sigma_1 = .2$ ,  $\sigma_2 = .15$ , and  $\rho_{12} = 0$ . Suppose as well that  $\mu_f = .06$ . Then  $V_1 = .14 - .06 = .08$  and  $V_2 = .08 - .06 = .02$ . Using (5.1) we get  $w_T = .693$ . Therefore,

$$E(R_T) = (.693)(.14) + (.307)(.08) = .122,$$

and

$$\sigma_T = \sqrt{(.693)^2(.2)^2 + (.307)^2(.15)^2} = .146.$$

Let  $R$  be the return on the portfolio that allocates a fraction  $\omega$  of the investment to the tangency portfolio and  $1 - \omega$  to the risk-free asset.

Then  $R = \omega R_T + (1 - \omega)\mu_f = \mu_f + \omega(R_T - \mu_f)$  so that

$$E(R) = \mu_f + \omega\{E(R_T) - \mu_f\} \quad \text{and} \quad \sigma_R = \omega\sigma_T.$$

**Continuation of previous example:** What is the optimal investment with  $\sigma_R = .05$ ?

**answer:** The maximum expected return with  $\sigma_R = .05$  mixes the tangency portfolio and the risk-free asset such that  $\sigma_R = .05$ . Since  $\sigma_T = .146$ , we have that  $.05 = \sigma_R = \omega\sigma_T = .146\omega$ , so that  $\omega = .05/.146 = .343$  and  $1 - \omega = .657$ .

So 65.7% of the portfolio should be in the risk-free asset. 34.3% should be in the tangency portfolio. Thus  $(.343)(69.3\%) = 23.7\%$  should be in the first risky asset and  $(.343)(30.7\%) = 10.5$  should be in the second risky asset.

In summary

Asset	Allocation
risk-free	65.7%
risky 1	23.7%
risky 2	10.5%
Total	99.9%

The total is not quite 100% because of rounding errors.

Now suppose that you want a 10% expected return. Compare

- The best portfolio of only risky assets
- the best portfolio of the risky assets and the risk-free asset

**Answer:**

- (best portfolio of risky assets)
  - $.1 = w(.14) + (1 - w)(.08)$  implies that  $w = 1/3$ .
  - This is the *only* portfolio of risky assets with  $E(R) = .1$ , so by default it is best.
  - Then
 
$$\sigma_R = \sqrt{w^2(.2)^2 + (1 - w)^2(.15)^2} = \sqrt{(1/9)(.2)^2 + 4/9(.15)^2} = .120.$$
- (best portfolio of the two risky assets and the risk-free asset)
  - $.1 = E(R) = .06 + .062\omega = .06 + .425\sigma_R$ , since  $\sigma_R = \omega\sigma_T$  or  $\omega = \sigma_R/\sigma_T = \sigma_R/.146$ .
  - This implies that  $\sigma_R = .04/.425 = .094$  and  $\omega = .04/.062 = .645$ .

So combining the risk-free asset with the two risky assets reduces  $\sigma_R$  from .120 to .094 while maintaining  $E(R)$  at .1. The reduction in risk is  $(.120 - .094)/.094 = 28\%$ .

**More on the example:** What is the best we can do combining the risk-free asset with only one risky asset? Assume that we still want to have  $E(R) = .1$

5.4. COMBINING TWO RISKY ASSETS WITH A RISK-FREE ASSET 107

- Second risky asset with the risk-free
  - Since  $\mu_f = .06 < .1$  and  $\mu_2 = .08 < .1$ , no portfolio with only the second risky asset and the risk-free asset will have an expected return of .1.
- First risky asset with the risk-free
  - $.1 = \omega(.14) + (1-\omega)(.06) = .06 + \omega(.08)$  implies that  $\omega = .04/.08 = 1/2$ .
  - Then  $\sigma_R = \omega(.20) = .10$  which is greater than .094, the smallest risk with two risky assets and the risk-free asset such that  $E(R) = .1$ .

The minimum value of  $\sigma_R$  under various combinations of available assets are given in Table 5.4.1.

Available Assets	Minimum $\sigma_R$
1st risky, risk-free	0.1
2nd risky, risk-free	-
Both riskies	0.12
All three	0.094

Table 5.1: Minimum value of  $\sigma_R$  as a function of the available assets.

### 5.4.2 Effect of $\rho_{12}$

Positive correlation between the two risky assets is bad. With positive correlation, then two assets tend to move together which increases the volatility of the portfolio. Conversely, negative correlation is good. If the assets are negatively correlated, a negative return of one tends to occur with a positive return of the other so the volatility of the portfolio decreases. Figure 5.4 shows the efficient frontier and tangency portfolio when  $\mu_1 = .14$ ,  $\mu_2 = .09$ ,  $\sigma_1 = .2$ ,  $\sigma_2 = .15$ , and  $\mu_f = .03$ . The value of  $\rho_{12}$  is varied

from .7 to  $-.7$ . Notice that the standard deviation of the portfolio returns decreases as  $\rho_{12}$  decreases.

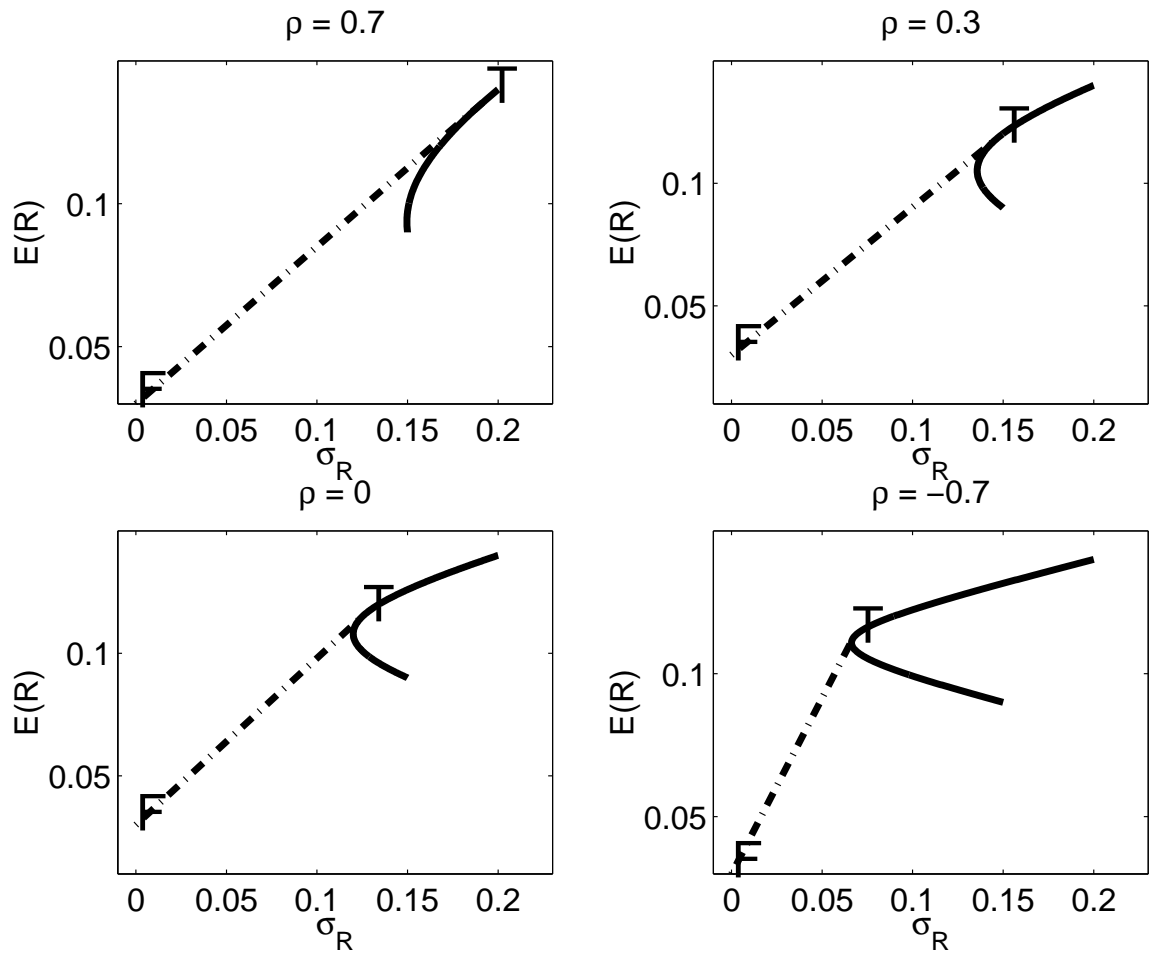


Figure 5.4: *Efficient frontier and tangency portfolio when  $\mu_1 = .14$ ,  $\mu_2 = .09$ ,  $\sigma_1 = .2$ ,  $\sigma_2 = .15$ , and  $\mu_f = .03$ . The value of  $\rho_{12}$  is varied from .7 to  $-.7$ .*

## 5.5 Harry Markowitz

Chapter Two of *Capital Ideas: The Improbable Origins of Modern Wall Street* by Peter Bernstein is titled “Fourteen Pages to Fame.” The title refers to the paper “Portfolio Selection” by Harry Markowitz that was published in the *Journal of Finance* in 1952. This article is indeed only fourteen pages though it was later expanded to the book *Portfolio Selection: Efficient Diversification of Investments* that was published by Markowitz in 1959.

Markowitz was not primarily interested in the stock market or investing. Rather, he was drawn to the more general issue of how people make trade-offs. Investors are faced with a trade-off between risk and expected return. The maxim “nothing ventured, nothing gained” isn’t quite true, but risk-free rates of return can be smaller than many investors find acceptable. Markowitz’s solution to the problem of risk also can be expressed as a maxim, “don’t put all your eggs in one basket.” (Keynes, would have agreed with Mark Twain who said, “put all your eggs in one basket — and then, **watch that basket!**”

Markowitz was born in 1927 and grew up in Chicago. His high school grades were not impressive, but he was intellectually curious and read a great deal on his own. At fourteen, he read Darwin’s *Origin of Species* and later his hero was the philosopher David Hume. The knowledge he acquired on his own got him into the University of Chicago and even exempted him from the required science courses there. This self-study may have been ideal preparation for the highly original work that came later.

After graduation, Markowitz became a research associate at the Cowles Commission and a graduate student at his Alma Mater. While waiting outside his advisor’s office one day, he began a conversation with a stock broker who suggested that he write his thesis on the stock market. Markowitz was somewhat surprised when later his advisor was enthusiastic over this idea.

Markowitz started to read what he could about investing. In the 1937 book

*The Theory of Investment Value* by John Burr Williams, he found Williams's prescription for selecting stocks: one estimated the "intrinsic value" of a stock by forecasting all future dividends and calculating the "present value" all future dividends, that is, the discounted sum of all future dividends. William then recommends that one put all one's capital in the stock with the highest intrinsic value.

Markowitz had enough knowledge of the world to realize that this is not how investors actually operated. He had the key insight that humans are risk-averse, and he began to explore the relationship between diversification and risk.

Interestingly, Markowitz did not recommend that expected returns be estimated from past data but rather from Williams's Dividend Discounted Model.

## 5.6 Risk-efficient portfolios with $N$ risky assets

### 5.6.1 Efficient-set mathematics

Efficient-set mathematics generalizes our previous analysis with two risky assets to the more realistic case of many risky assets. This material is taken from Section 5.2 of Campbell, Lo, and MacKinlay.

Assume that we have  $N$  risky assets and that the return on the  $i$ th risky asset is  $\mu_i$ . Define

$$\mathbf{R} = \begin{pmatrix} R_1 \\ \vdots \\ R_N \end{pmatrix}$$

to be the random vector of returns. Then

$$E(\mathbf{R}) = \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_N \end{pmatrix}.$$

Let  $\Omega_{ij}$  be the covariance between  $R_i$  and  $R_j$ . Also, let  $\sigma_i = \sqrt{\Omega_{ii}}$  be the standard deviation of  $R_i$ . Define  $\rho_{ij} = \Omega_{ij}/(\sigma_i\sigma_j)$  as the correlation between  $R_i$  and  $R_j$ . Finally, let  $\mathbf{\Omega}$  be the covariance matrix of  $\mathbf{R}$ , i.e.,

$$\mathbf{\Omega} = \text{COV}(\mathbf{R}),$$

so that the  $i, j$ th element of  $\mathbf{\Omega}$  is  $\Omega_{ij}$ .

Let

$$\boldsymbol{\omega} = \begin{pmatrix} \omega_1 \\ \vdots \\ \omega_N \end{pmatrix}$$

be a matrix of portfolio weights and let

$$\mathbf{1} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

be a column of  $N$  ones. We assume that  $\omega_1 + \cdots + \omega_N = \mathbf{1}^T \boldsymbol{\omega} = 1$ . The expected return on a portfolio with weights  $\boldsymbol{\omega}$  is  $\sum_{i=1}^N \omega_i \mu_i = \boldsymbol{\omega}^T \boldsymbol{\mu}$ .

When  $N = 2$ ,  $\omega_2 = 1 - \omega_1$ . Suppose there is a target value,  $\mu_P$ , of the expected return on the portfolio. We assume that

$$\min_{i=1,\dots,N} \mu_i \leq \mu_P \leq \max_{i=1,\dots,N} \mu_i,$$

since no portfolio can have an expected return higher than the individual asset with the highest expected return or smaller than the individual asset with the lowest expected return. When  $N = 2$  the target,  $\mu_P$ , is achieved by only one portfolio and its  $\omega_1$  value solves

$$\mu_P = \omega_1 \mu_1 + \omega_2 \mu_2 = \mu_2 + \omega_1 (\mu_1 - \mu_2).$$

For  $N \geq 3$ , there will be an infinite number of portfolios achieving the target,  $\mu_P$ . The one with the smallest variance is called the "efficient" portfolio. Our goal is to find the efficient portfolio.

By equation (2.2), the variance of the return on the portfolio with weights  $\boldsymbol{\omega}$  is

$$\sum_{i=1}^N \sum_{j=1}^N \omega_i \omega_j \Omega_{ij} = \boldsymbol{\omega}^T \boldsymbol{\Omega} \boldsymbol{\omega}. \quad (5.2)$$

Thus, given a target  $\mu_P$ , the efficient portfolio minimizes (5.2) subject to

$$\boldsymbol{\omega}^T \boldsymbol{\mu} = \mu_P \quad (5.3)$$

and

$$\boldsymbol{\omega}^T \mathbf{1} = 1. \quad (5.4)$$

We will denote the weights of the efficient portfolio by  $\boldsymbol{\omega}_{\mu_P}$ . To find  $\boldsymbol{\omega}_{\mu_P}$ , form the Lagrangian

$$L = \boldsymbol{\omega}^T \boldsymbol{\Omega} \boldsymbol{\omega} + \delta_1 (\mu_P - \boldsymbol{\omega}_{\mu_P}^T \boldsymbol{\mu}) + \delta_2 (1 - \boldsymbol{\omega}^T \mathbf{1}).$$

Then solve

$$0 = \frac{\partial}{\partial \boldsymbol{\omega}} L = 2\boldsymbol{\Omega} \boldsymbol{\omega}_{\mu_P} - \delta_1 \boldsymbol{\mu} - \delta_2 \mathbf{1}. \quad (5.5)$$

**Definition:** Here

$$\frac{\partial}{\partial \boldsymbol{\omega}} L$$

means the gradient of  $L$  with respect to  $\boldsymbol{\omega}$  with the other variables in  $L$  held fixed.

**Fact:** For an  $n \times n$  matrix  $\mathbf{A}$  and an  $n$ -dimensional vector  $\mathbf{x}$ ,

$$\frac{\partial}{\partial \mathbf{x}} \mathbf{x}^T \mathbf{A} \mathbf{x} = (\mathbf{A} + \mathbf{A}^T) \mathbf{x}$$

The solution to (5.5) is

$$\boldsymbol{\omega}_{\mu_P} = \frac{1}{2} \boldsymbol{\Omega}^{-1} (\delta_1 \boldsymbol{\mu} + \delta_2 \mathbf{1}) = \boldsymbol{\Omega}^{-1} (\lambda_1 \boldsymbol{\mu} + \lambda_2 \mathbf{1})$$



where  $\lambda_1$  and  $\lambda_2$  are new Lagrange multipliers:

$$\lambda_1 = \frac{1}{2}\delta_1 \quad \text{and} \quad \lambda_2 = \frac{1}{2}\delta_2.$$

Thus,

$$\boldsymbol{\omega}_{\mu_P} = \lambda_1 \boldsymbol{\Omega}^{-1} \boldsymbol{\mu} + \lambda_2 \boldsymbol{\Omega}^{-1} \mathbf{1},$$

where  $\lambda_1$  and  $\lambda_2$  are yet to be determined scalar quantities. We need to use the constraints to determine  $\lambda_1$  and  $\lambda_2$ . Therefore,

$$\mu_P = \boldsymbol{\mu}^\top \boldsymbol{\omega}_{\mu_P} = \lambda_1 \boldsymbol{\mu}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\mu} + \lambda_2 \boldsymbol{\mu}^\top \boldsymbol{\Omega}^{-1} \mathbf{1},$$

and

$$1 = \mathbf{1}^\top \boldsymbol{\omega}_{\mu_P} = \lambda_1 \mathbf{1}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\mu} + \lambda_2 \mathbf{1}^\top \boldsymbol{\Omega}^{-1} \mathbf{1}.$$

Define

$$\begin{aligned} A &= \boldsymbol{\mu}^\top \boldsymbol{\Omega}^{-1} \mathbf{1} = \mathbf{1}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\mu} \\ B &= \boldsymbol{\mu}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\mu}, \\ C &= \mathbf{1}^\top \boldsymbol{\Omega}^{-1} \mathbf{1}, \\ D &= BC - A^2. \end{aligned}$$

Then

$$\begin{aligned} \mu_P &= B\lambda_1 + A\lambda_2 \\ 1 &= A\lambda_1 + C\lambda_2. \end{aligned}$$

These are equations in  $\lambda_1$  and  $\lambda_2$ ;  $A$ ,  $B$ , and  $C$  are known quantities. The solution is

$$\lambda_1 = \frac{-A + C\mu_P}{D} \quad \text{and} \quad \lambda_2 = \frac{B - A\mu_P}{D}.$$

It follows after some algebra that

$$\boldsymbol{\omega}_{\mu_P} = \mathbf{g} + \mathbf{h} \mu_P, \tag{5.6}$$

where

$$\mathbf{g} = \frac{B \boldsymbol{\Omega}^{-1} \mathbf{1} - A \boldsymbol{\Omega}^{-1} \boldsymbol{\mu}}{D}, \tag{5.7}$$

and

$$\mathbf{h} = \frac{C \boldsymbol{\Omega}^{-1} \boldsymbol{\mu} - A \boldsymbol{\Omega}^{-1} \mathbf{1}}{D}. \quad (5.8)$$

Notice that  $\mathbf{g}$  and  $\mathbf{h}$  are fixed vectors, since they depend on the fixed vector  $\boldsymbol{\mu}$  and the fixed matrix  $\boldsymbol{\Omega}$ . Also, the scalars  $A$ ,  $C$ , and  $D$  are functions of  $\boldsymbol{\mu}$  and  $\boldsymbol{\Omega}$  so they are also fixed. The target expected return,  $\mu_P$ , can be varied over the range

$$\min_{i=1,\dots,N} \mu_i \leq \mu_P \leq \max_{i=1,\dots,N} \mu_i.$$

As  $\mu_P$  varies over this range, we get a locus  $\boldsymbol{\omega}_{\mu_P}$  of efficient portfolios called the "efficient frontier." We can illustrate the efficient frontier by the following algorithm:

1. Vary  $\mu_P$  along a grid. For each value of  $\mu_P$  on this grid, compute  $\sigma_{\mu_P}$  by:
  - (a) computing  $\boldsymbol{\omega}_{\mu_P} = \mathbf{g} + \mathbf{h} \mu_P$
  - (b) then computing  $\sigma_{\mu_P} = \sqrt{\boldsymbol{\omega}_{\mu_P} \boldsymbol{\Omega} \boldsymbol{\omega}_{\mu_P}}$
2. Plot the values  $(\mu_P, \sigma_{\mu_P})$

This algorithm is implemented in the MATLAB program “portfolio02.m” on the course’s web site and listed below:

```
% Input mean vector and covariance matrix of returns here

bmu = [.08;.03;.05] ;
bOmega = [ .3 .02 .01 ;
           .02 .15 .03 ;
           .01 .03 .18 ] ;

bone = ones(length(bmu),1) ;

short = 1 ; % short = 1 implies extensive short selling

% short = 0 reduces the short selling, but
% does not eliminate short sell

ngrid = 200 ;
if short == 1 ;
muP = linspace(-.02,.2,ngrid) ;
w = linspace(-5,7,ngrid) ;

else ;
muP = linspace(min(bmu),max(bmu),ngrid) ;
w = linspace(0,1,ngrid) ;
end ;

sigmaP = zeros(1,ngrid) ;
omegaP = zeros(3,ngrid) ;

mu12 = zeros(1,ngrid) ;
sigma12 = mu12 ;
mu13 = zeros(1,ngrid) ;
sigma13 = mu12 ;
mu23 = zeros(1,ngrid) ;
sigma23 = mu12 ;
```

```

ibOmega = inv(bOmega) ;
A = bone'*ibOmega*bmu ;
B = bmu'*ibOmega*bmu ;
C = bone'*ibOmega*bone ;
D = B*C - A^2 ;
bg = (B*ibOmega*bone - A*ibOmega*bmu)/D ;
bh = (C*ibOmega*bmu - A*ibOmega*bone)/D ;

for i=1:ngrid ;
omegaP(:,i) = bg + muP(i)*bh ;
sigmaP(i) = sqrt(omegaP(:,i)'*bOmega*omegaP(:,i)) ;

mul2(i) = w(i)*bmu(1) + (1-w(i))*bmu(2) ;
sigma12(i) = sqrt(w(i)^2*bOmega(1,1) + 2*w(i)*(1-w(i))*bOmega(1,2) ...
+ (1-w(i))^2*bOmega(2,2)) ;

mul3(i) = w(i)*bmu(1) + (1-w(i))*bmu(3) ;
sigma13(i) = sqrt(w(i)^2*bOmega(1,1) + 2*w(i)*(1-w(i))*bOmega(1,3) ...
+ (1-w(i))^2*bOmega(3,3)) ;
mu23(i) = w(i)*bmu(2) + (1-w(i))*bmu(3) ;
sigma23(i) = sqrt(w(i)^2*bOmega(2,2) + 2*w(i)*(1-w(i))*bOmega(2,3) ...
+ (1-w(i))^2*bOmega(3,3)) ;
end ;

fsize = 16 ;
figure(1)
p = plot(sigmaP,muP,sigma12,mul2,'--',sigma13,mul3,'-.',sigma23,mu23,':') ;
set(p,'linewidth',6) ;
xlabel('standard deviation of return (\sigma_P)','fontsize',fsize) ;
ylabel('expected return (\mu_P)','fontsize',fsize) ;
text(sqrt(bOmega(1,1)),bmu(1),'1','fontsize',24) ;
text(sqrt(bOmega(2,2)),bmu(2),'2','fontsize',24) ;
text(sqrt(bOmega(3,3)),bmu(3),'3','fontsize',24) ;
set(gca,'fontsize',fsize) ;
if short == 0 ;
set(gca,'ylim',[.025,.085]) ;
end ;
if short == 1 ;
set(gca,'ylim',[-.02,.2]) ;
set(gca,'xlim',[.2,2]) ;
end ;

```

```

grid ;

if short == 0 ;
print portfolio02.ps -depsec ;
!mv portfolio02.ps ~/public_html/or473/LectNotes/portfolio02.ps ;

else ;
print portfolio02SH.ps -depsec ;
!mv portfolio02SH.ps ~/public_html/or473/LectNotes/portfolio02SH.ps ;

end ;

figure(2)
p2 = plot(muP,omegaP(1,:),muP,omegaP(2,:), '--',muP,omegaP(3,:), '-.') ;
set(p2,'linewidth',6) ;
set(gca,'fontsize',fsize) ;
grid ;
xlabel('\mu_P','fontsize',fsize) ;
ylabel('weight','fontsize',fsize) ;
legend('w_1','w_2','w_3',0) ;

if short == 0
print portfolio02_wt.ps -depsec ;
!mv portfolio02_wt.ps ~/public_html/or473/LectNotes/portfolio02_wt.ps ;
else ;
print portfolio02_wtSH.ps -depsec ;
!mv portfolio02_wtSH.ps ~/public_html/or473/LectNotes/portfolio02_wtSH.ps ;

end ;

```

To use this program replace  $\mathbf{b}\mu$  and  $\mathbf{b}\Omega$  in the program by the vector of expected returns and covariance matrix of returns for the assets you wish to analyze. The parameter “short” should be set equal to 0 or 1. If “short” is 1, then there is extensive short selling, i.e., weights get quite negative. If “short” is 0, then the amount of short selling is small.

Figure 5.5 was produced by this program with “short” equal to 0.

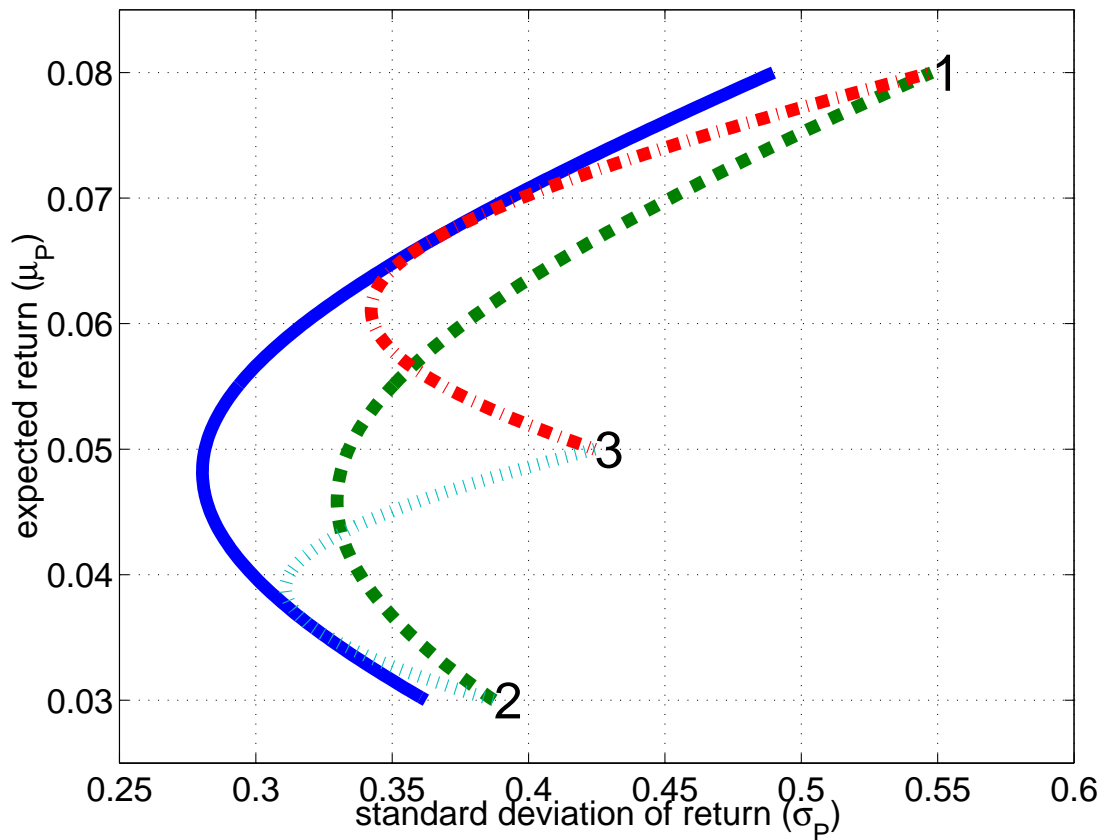


Figure 5.5: Efficient frontier (solid) plotted for  $N = 3$  assets by the program “portfolio02.m” with the parameter “short” equal to 0. “1,” “2,” and “3” are the three single assets. The efficient frontiers for just two assets are dashed (1 and 2), dashed-and-dotted (1 and 3), and dotted (2 and 3).

The portfolio weights as functions of  $\mu_P$  are plotted in Figure 5.6. The weights can be negative. Negative weights can be obtained by the technique of selling short which is described in Section 5.6.2.

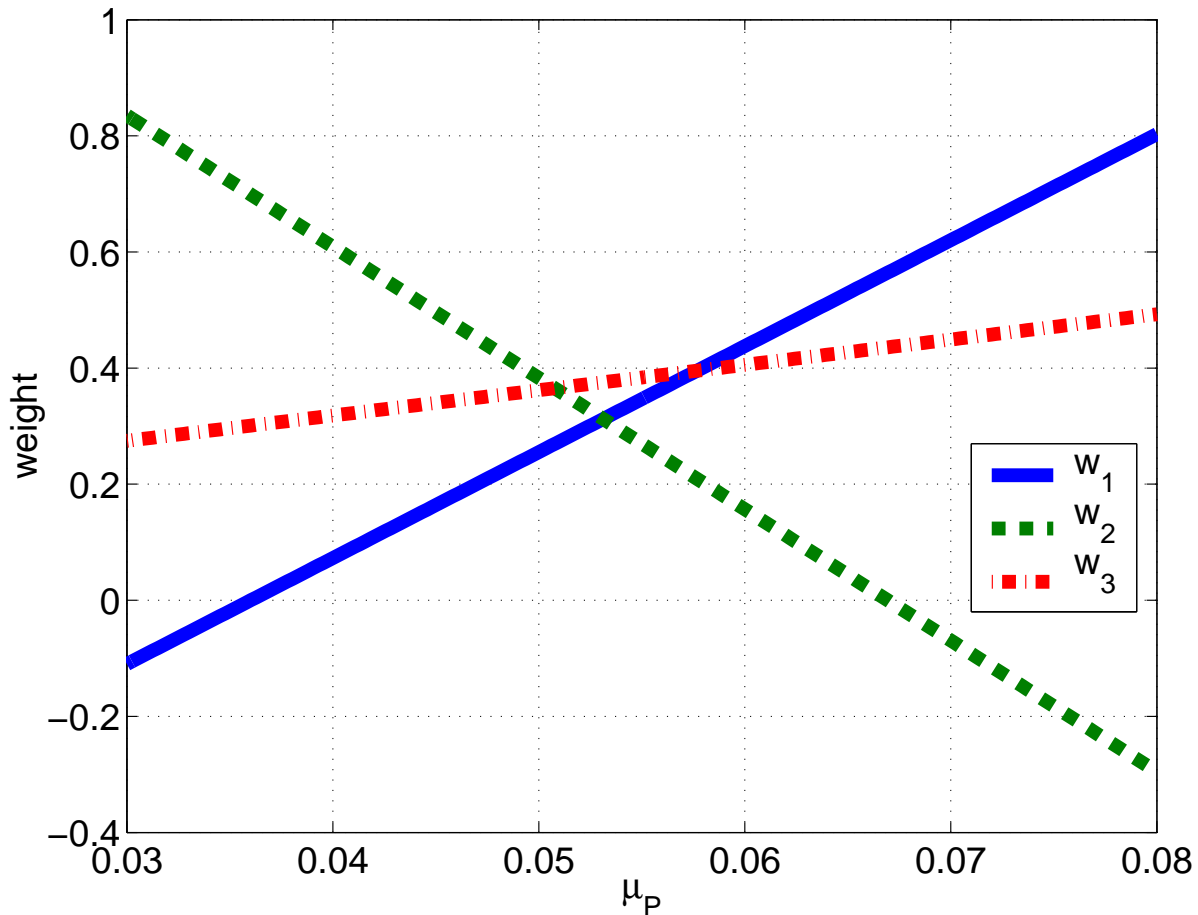


Figure 5.6: Weights for assets 1, 2, and 3 as functions of  $\mu_p$ . Note that the weights for assets 1 and 2 can be negative, so that short selling would be required.

If one wants to avoid short selling, then one must impose the additional constraints that  $w_i \geq 0$  for  $i = 1, \dots, N$ . Minimization of portfolio risk subject to  $\boldsymbol{\omega}^\top \boldsymbol{\mu} = \mu_P$ ,  $\boldsymbol{\omega}^\top \mathbf{1} = 1$ , and these additional nonnegativity constraints is a quadratic programming problem. (This minimization problem cannot be solved by the method of Lagrange multipliers because of the inequality constraints.) Quadratic programming algorithms are not hard to find. For example, the program “quadprog” in MATLAB’s Optimization Toolbox does quadratic programming.

Figure 5.7 and 5.8 were produced by the program “portfolio02QP.m” that uses “quadprog” in MATLAB. Quadratic programming in MATLAB and “portfolio02QP.m are discussed in Section 5.9.



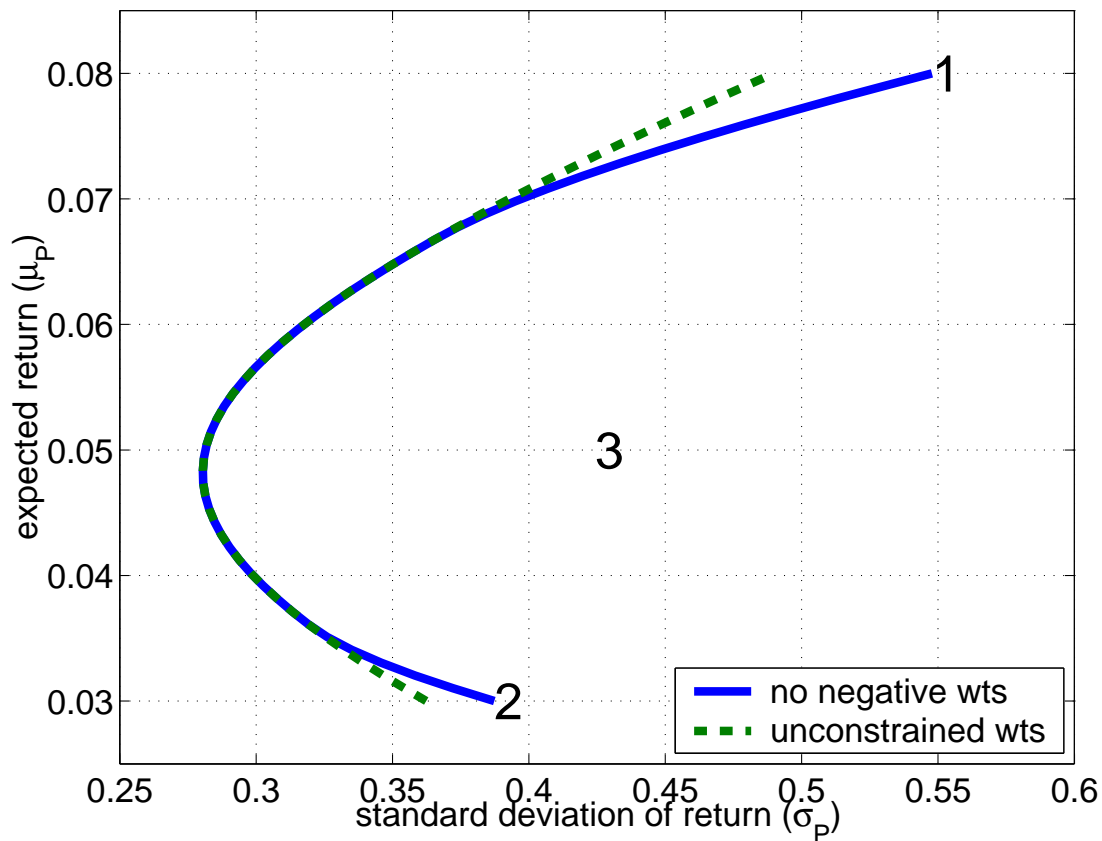


Figure 5.7: Efficient frontier plotted by the program “portfolio02QP.m” for  $N = 3$  assets. “1,” “2,” and “3” are the three single assets. The efficient frontiers are found with and without the constraint of no negative weights. The constrained efficient frontier is computed using MATLAB’s quadratic programming algorithm.

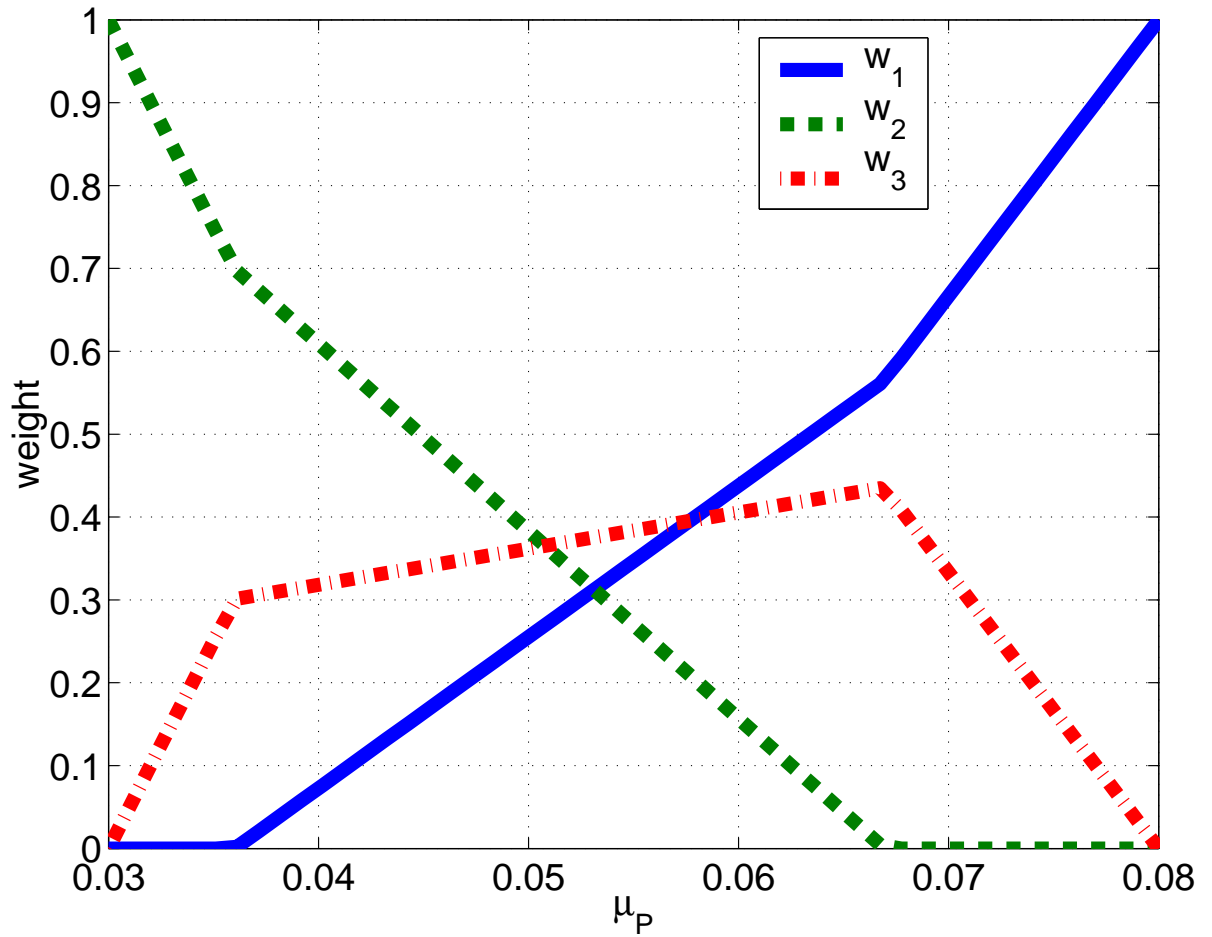


Figure 5.8: Weights for assets 1, 2, and 3 as functions of  $\mu_p$ . The weights for all three assets are constrained to be nonnegative.

The portfolio weights with the nonnegativity constraint are plotted as functions of  $\mu_P$  in Figure 5.8. That figure was made using the program “portfolio02QP.m” listed here:

```
% Input mean vector and covariance matrix of returns here

bmu = [.08;.03;.05] ;
bOmega = [ .3 .02 .01 ;
           .02 .15 .03 ;
           .01 .03 .18 ] ;
A = [ones(1,3);bmu'] ;
ngrid = 50 ;
muP = linspace(.03,.08,ngrid)' ;

icompute = 0 ;
if icompute == 1 ;
sigmaP = muP ;
sigmaP2 = sigmaP ;
omegaP = zeros(3,ngrid) ;
omegaP2 = omegaP ;

for i = 1:ngrid ;
omegaP(:,i) = quadprog(bOmega,zeros(3,1),-eye(3),zeros(3,1),A,[1;muP(i)]) ;
omegaP2(:,i) = quadprog(bOmega,zeros(3,1),zeros(1,3),0,A,[1;muP(i)]) ;

sigmaP(i) = sqrt(omegaP(:,i)'*bOmega*omegaP(:,i)) ;
sigmaP2(i) = sqrt(omegaP2(:,i)'*bOmega*omegaP2(:,i)) ;

end ;
end ;

fsize = 16 ;
figure(1)
clf
p = plot(sigmaP,muP,sigmaP2,muP,'--') ;
l=legend('no negative wts','unconstrained wts',4) ;
set(gca,'fontsize',fsize) ;
set(l,'fontsize',fsize) ;

xlabel('standard deviation of return (\sigma_P)','fontsize',fsize) ;
```

```
ylabel('expected return (\mu_P)', 'fontsize', fsize) ;
text(sqrt(bOmega(1,1)), bmu(1), '1', 'fontsize', 24) ;
text(sqrt(bOmega(2,2)), bmu(2), '2', 'fontsize', 24) ;
text(sqrt(bOmega(3,3)), bmu(3), '3', 'fontsize', 24) ;
set(gca, 'ylim', [.025, .085]) ;

set(p, 'linewidth', 4) ;
grid ;

print portfolio02QP.ps -depsc ;
!mv portfolio02QP.ps ~/public_html/or473/LectNotes/portfolio02QP.ps ;

figure(2)
p2 = plot(muP, omegaP(1,:), muP, omegaP(2,:), '--', muP, omegaP(3,:), '-.') ;
set(p2, 'linewidth', 6) ;
set(gca, 'fontsize', fsize) ;
grid ;
xlabel('\mu_P', 'fontsize', fsize) ;
ylabel('weight', 'fontsize', fsize) ;
legend('w_1', 'w_2', 'w_3', 0) ;

print portfolio02_wtQP.ps -depsc ;
!mv portfolio02_wtQP.ps ~/public_html/or473/LectNotes/portfolio02_wtQP.ps ;
```

Now suppose that we have a risk-free asset and we want to mix the risk-free asset with *some* efficient portfolio. One can see geometrically that there is a tangency portfolio; see Figure 5.9. **The optimal portfolio *always* is a mixture of the risk-free asset with the tangency portfolio.** This is a remarkable simplification.

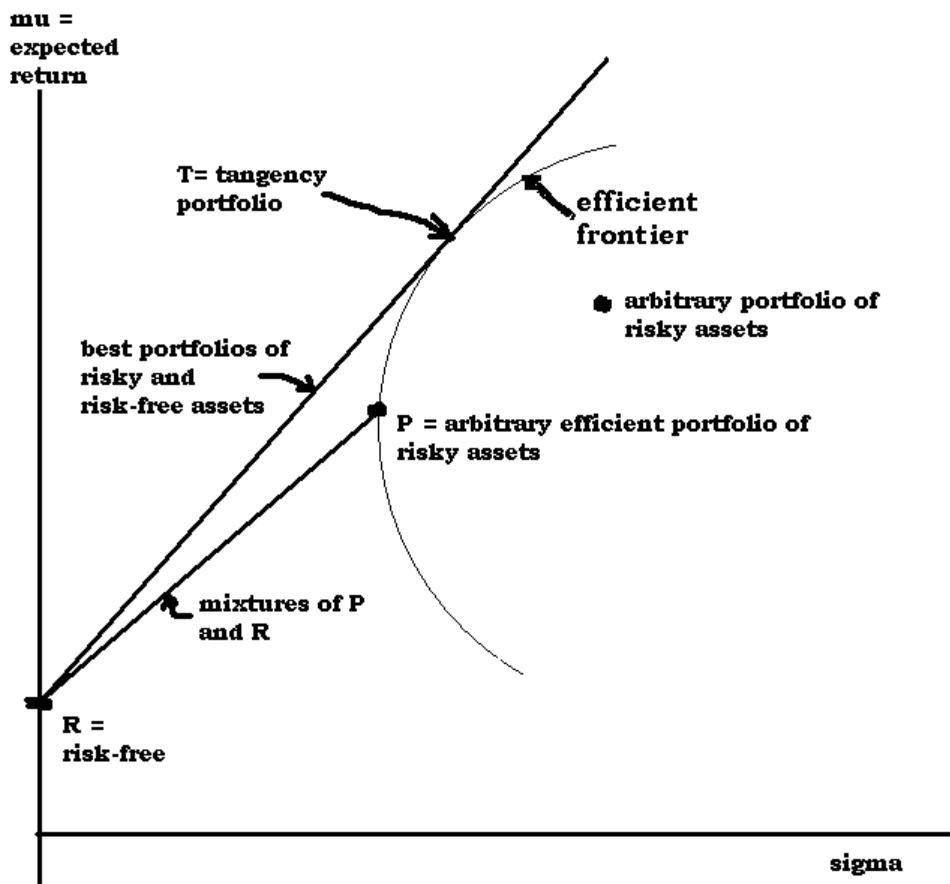


Figure 5.9: Finding the best portfolios that combine risky and risk-free assets.  $R$  is the risk-free asset.  $T$  is the tangency portfolio. The optimal portfolios are on the line connecting  $R$  and  $T$ . The efficient frontier gives the set of optimal portfolios of risky assets.

### 5.6.2 Selling short

Selling short is a way to profit if a stock price goes *down*. To sell a stock short, one sells the stock without owning it. The stock must be borrowed from a broker or another customer of the broker. At a later point in time, one buys the share and gives it back to the lender. This closes the short position.

Suppose a stock is selling at \$25/share and you sell 100 shares short. This gives you \$2,500. If the price goes down to \$17/share, you can buy the 100 shares for \$1,700 and close out your short position.

Suppose that you have \$100 and there are two risky assets. With your money you could buy \$150 worth of risky asset 1 and sell \$50 short of risky asset 2. The net cost would be exactly \$100. If  $R_1$  and  $R_2$  are the returns on risky assets 1 and 2, then our the return on your portfolio would be

$$\frac{3}{2}R_1 + \left(-\frac{1}{2}\right)R_2.$$

Your portfolio weights are  $w_1 = 3/2$  and  $w_2 = -1/2$ . Thus, you hope that risky asset 1 rises in price and risky asset 2 falls in price.

Here, as elsewhere, we have ignored transaction costs.

Figure 5.10 is the same as Figure 5.5 except that the range of values of  $\mu_P$  has been expanded. Values of  $\mu_P$  below  $\min(\mu_i)$  and above  $\max(\mu_i)$  are possible by using short selling. In principle, there is no upper limit to  $\mu_P$ , but in practice security exchanges place limits on the amount of stock one can sell short because selling short increases risk.

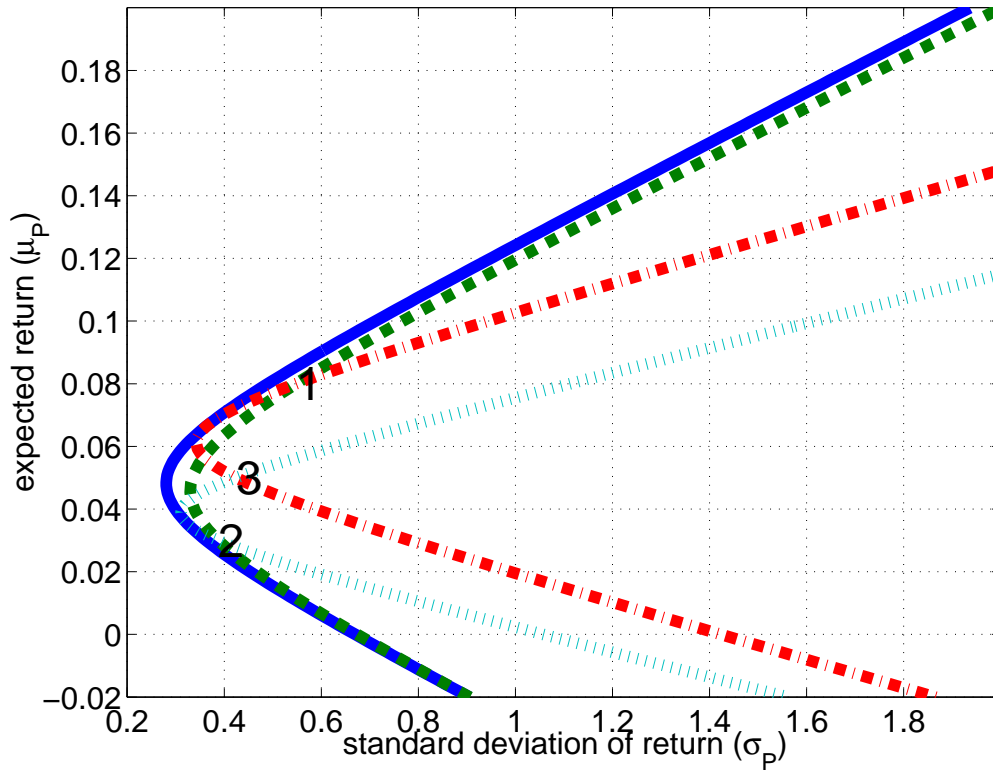


Figure 5.10: Efficient frontier (solid) plotted for  $N = 3$  assets by the program "portfolio02.m" with the parameter "short" equal to 1. "1," "2," and "3" are the three single assets. The efficient frontiers for just two assets are dashed (1 and 2), dashed-and-dotted (1 and 3), and dotted (2 and 3). This figure is the same as Figure 5.5 except that the range of  $\mu_p$  has been expanded.

### 5.6.3 The Interior decorator fallacy

It is often thought that a stock portfolio should be tailored to the financial circumstances of a client, as an interior decorator furnishes your home to suit your tastes. For example, widows and orphans should hold conservative “income stocks,” or so it is said.

Bernstein, in his book *Capital Ideas*, calls this the “interior decorator fallacy.” Bernstein tells the story of a woman in her forties who came to him in 1961 for investment advice. She was married to a clergyman with a modest income. She had just inherited money which she wanted to invest. Bernstein recommended a portfolio that included stocks with good growth potential but low dividends, e.g., Georgia Pacific, IBM, and Gillette. The client was worried that these were too risky, but she eventually took Bernstein’s advice, which turned out to be sound. Bernstein reasoned that even someone with modest means should benefit from the long-term growth potential of the “hot” stocks.

In another case, Bernstein recommended electric utilities, a conservative choice, to a young business executive who wanted a more aggressive portfolio. Again, this recommendation was at odds with conventional wisdom.

A new view, based both on mathematical theory and experience, is that there is a best portfolio (the tangency portfolio) that is the same for everyone. An individual’s circumstances only determines the appropriate mix between risk-free assets and the tangency portfolio. The clergyman’s wife should invest a higher percentage of her money in risk-free assets than the young business executive. In 1961, Bernstein had the right intuition but he had not yet heard of the Efficient Frontier or the tangency portfolio.

### 5.6.4 Back to the math

Here’s the mathematics behind Figure 5.9. We now remove the assumption that  $\omega^T \mathbf{1} = 1$ . The quantity  $1 - \omega^T \mathbf{1}$  is invested in the risk-free asset.



(Does it make sense to have  $1 - \boldsymbol{\omega}^\top \mathbf{1} < 0$ ?). The expected return is

$$\boldsymbol{\omega}^\top \boldsymbol{\mu} + (1 - \boldsymbol{\omega}^\top \mathbf{1})\mu_f, \quad (5.9)$$

where  $\mu_f$  is the return on the risk-free asset. The constraint to be satisfied is that (5.9) is equal to  $\mu_P$ . Thus, the Lagrangian function is

$$L = \boldsymbol{\omega}^\top \boldsymbol{\Omega} \boldsymbol{\omega} + \delta \{ \mu_P - \boldsymbol{\omega}^\top \boldsymbol{\mu} - (1 - \boldsymbol{\omega}^\top \mathbf{1})\mu_f \}.$$

Here  $\delta$  is a Lagrange multiplier. Since

$$0 = \frac{\partial}{\partial \boldsymbol{\omega}} L = 2\boldsymbol{\Omega} \boldsymbol{\omega} + \delta(-\boldsymbol{\mu} + \mathbf{1}\mu_f),$$

the optimal weight vector, i.e., the vector of weights that minimizes risk subject to the constraint on the expected return, is

$$\boldsymbol{\omega}_{\mu_P} = \lambda \boldsymbol{\Omega}^{-1}(\boldsymbol{\mu} - \mu_f \mathbf{1}), \quad (5.10)$$

where  $\lambda = \delta/2$ . To find  $\lambda$ , we use our constraint:

$$\boldsymbol{\omega}_{\mu_P}^\top \boldsymbol{\mu} + (1 - \boldsymbol{\omega}_{\mu_P}^\top \mathbf{1})\mu_f = \mu_P. \quad (5.11)$$

Rearranging (5.11), we get

$$\boldsymbol{\omega}_{\mu_P}^\top (\boldsymbol{\mu} - \mu_f \mathbf{1}) = \mu_P - \mu_f. \quad (5.12)$$

Therefore, substituting (5.10) into (5.12) we have

$$\lambda (\boldsymbol{\mu} - \mu_f \mathbf{1})^\top \boldsymbol{\Omega}^{-1} (\boldsymbol{\mu} - \mu_f \mathbf{1}) = \mu_P - \mu_f,$$

or

$$\lambda = \frac{\mu_P - \mu_f}{(\boldsymbol{\mu} - \mu_f \mathbf{1})^\top \boldsymbol{\Omega}^{-1} (\boldsymbol{\mu} - \mu_f \mathbf{1})}. \quad (5.13)$$

Then substituting (5.13) into (5.10)

$$\boldsymbol{\omega}_{\mu_P} = c_P \bar{\boldsymbol{\omega}},$$

where

$$c_P = \frac{\mu_P - \mu_f}{(\boldsymbol{\mu} - \mu_f \mathbf{1})^\top \boldsymbol{\Omega}^{-1} (\boldsymbol{\mu} - \mu_f \mathbf{1})}$$

and

$$\bar{\omega} = \Omega^{-1}(\boldsymbol{\mu} - \mu_f \mathbf{1}). \quad (5.14)$$

Note that  $(\boldsymbol{\mu} - \mu_f \mathbf{1})$  is the vector of “excess returns,” that is, the amount by which the expected returns on the risky assets exceed the risk-free return. The excess returns measure how much the market buys for assuming risk.

$\bar{\omega}$  is not quite a portfolio because these weights do not necessarily sum to one. The tangency portfolio is a scalar multiple of  $\bar{\omega}$ :

$$\boldsymbol{\omega}_T = \frac{\bar{\omega}}{\mathbf{1}^T \bar{\omega}}. \quad (5.15)$$

$c_P$  tells us how much weight to put on  $\bar{\omega}$  and therefore on the tangency portfolio. The amount of weight to put of the risk-free asset is  $1 - \bar{\omega}^T \mathbf{1} = 1 - c_P (\bar{\omega}^T \mathbf{1})$ . The weight on the tangency portfolio is  $c_P (\bar{\omega}^T \mathbf{1})$ .

Note that  $\bar{\omega}$  and  $\boldsymbol{\omega}_T$  do not depend on  $\mu_p$ .

The MATLAB program “portfolio03.m” on the course web site is an extension of “portfolio02.m.” portfolio03.m, which is listed below, also plots of the tangency portfolio (T) and the line connecting the risk-free asset (F) with the tangency portfolio.

```
% portfolio03 - extension of portfolio02
% Input mean vector and covariance matrix of returns here

bmu = [.08;.03;.05] ;
bOmega = [ .3 .02 .01 ;
           .02 .15 .03 ;
           .01 .03 .18 ] ;

muf = .02 ;

bone = ones(length(bmu),1) ;
```

```

muP = linspace(min(bmu),max(bmu),50) ;
sigmaP = zeros(1,50) ;

ibOmega = inv(bOmega) ;
A = bone'*ibOmega*bmu ;
B = bmu'*ibOmega*bmu ;
C = bone'*ibOmega*bone ;
D = B*C - A^2 ;
bg = (B*ibOmega*bone - A*ibOmega*bmu)/D ;
bh = (C*ibOmega*bmu - A*ibOmega*bone)/D ;

for i=1:50 ;
omegaP = bg + muP(i)*bh ;
sigmaP(i) = sqrt(omegaP'*bOmega*omegaP) ;
end ;

bomegabar = ibOmega*(bmu - muf*bone) ;
bomegaT = bomegabar/(bone'*bomegabar) ;
sigmaT = sqrt(bomegaT'*bOmega*bomegaT) ;
muT = bmu'*bomegaT ;

fsize = 16 ;
fsize2 = 28 ;
bomegaP2 = [0;.3;.7] ;
sigmaP2 = sqrt(bomegaP2'*bOmega*bomegaP2) ;
muP2 = bmu'*bomegaP2 ;

clf ;
p1 = plot(sigmaP,muP) ;
l1 = line([0,sigmaT],[muf,muT]) ;
t1= text(sigmaP2,muP2,'* P','fontsize',fsize2) ;
t2= text(sigmaT,muT,'* T','fontsize',fsize2) ;
t3=text(.01,muf+.006,'F','fontsize',fsize2) ;
t3B= text(0,muf,'*', 'fontsize',fsize2) ;

```

```
set(p1,'linewidth',2) ;
set(l1,'linewidth',2) ;
set(l1,'linestyle','--') ;

xlabel('standard deviation of return','fontsize',fsize) ;
ylabel('expected return','fontsize',fsize) ;

print portfolio03.ps -deps ;
```

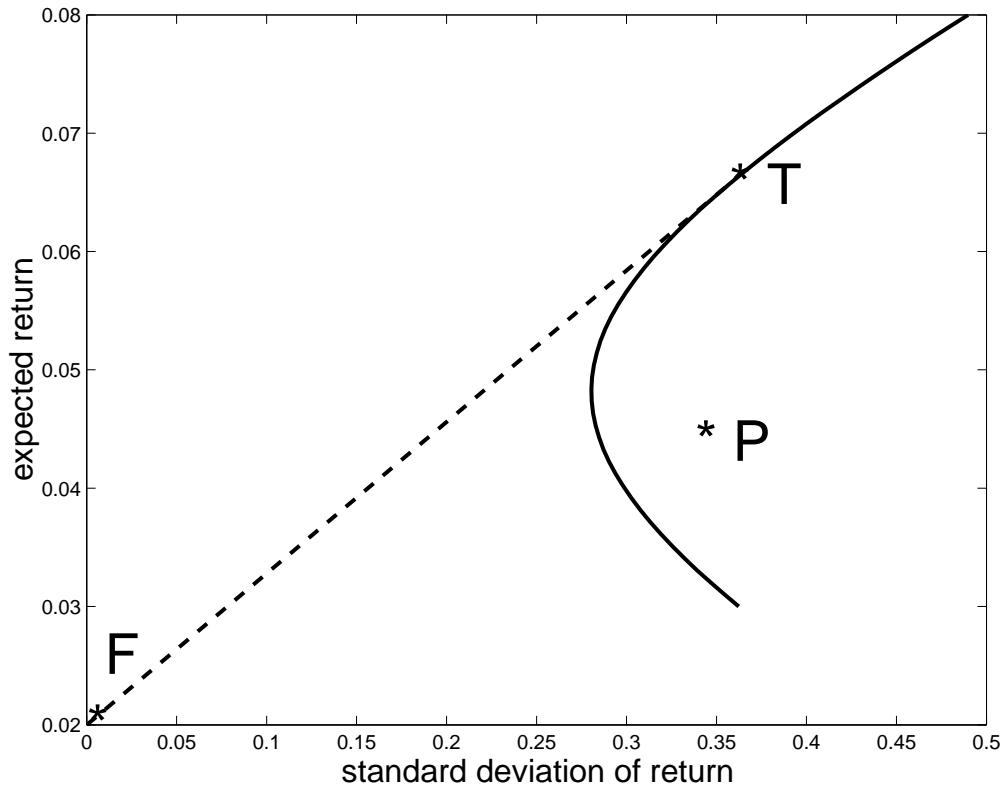


Figure 5.11: Efficient frontier and line of optimal combinations of risky and risk-free assets plotted by the program "portfolio03.m" for  $N = 3$  assets. "P" is the portfolio with weights (0 .3 .7) that is not on the efficient frontier. "T" is the tangency portfolio and "F" is the risk-free asset.

### 5.6.5 Example: $N = 2$

If  $N = 2$ , then

$$\mathbf{\Omega} = \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

You should check that

$$\frac{1}{1 - \rho_{12}^2} \mathbf{\Omega}^{-1} = \begin{pmatrix} \sigma_1^{-2} & -\rho_{12}\sigma_1^{-1}\sigma_2^{-1} \\ -\rho_{12}\sigma_1^{-1}\sigma_2^{-1} & \sigma_2^{-2} \end{pmatrix}.$$

Also,

$$\boldsymbol{\mu} - \mu_f \mathbf{1} = \begin{pmatrix} \mu_1 - \mu_f \\ \mu_2 - \mu_f \end{pmatrix}.$$

Therefore,

$$\bar{\boldsymbol{w}} = \mathbf{\Omega}^{-1}(\boldsymbol{\mu} - \mu_f \mathbf{1}) = \frac{1}{1 - \rho_{12}} \begin{pmatrix} \frac{\mu_1 - \mu_f}{\sigma_2^2} - \frac{\rho_{12}(\mu_2 - \mu_f)}{\sigma_1\sigma_2} \\ -\frac{\rho_{12}(\mu_1 - \mu_f)}{\sigma_1\sigma_2} + \frac{\mu_2 - \mu_f}{\sigma_2^2} \end{pmatrix}.$$

Next, let  $V_1 = \mu_1 - \mu_f$  and  $V_2 = \mu_2 - \mu_f$ . Then,

$$\mathbf{1}^\top \bar{\boldsymbol{w}} = \frac{V_1 \sigma_2^2 + V_2 \sigma_1^2 - (V_1 + V_2) \rho_{12} \sigma_1 \sigma_2}{\sigma_1^2 \sigma_2^2 (1 - \rho_{12})}.$$

It follows that

$$\frac{\bar{\boldsymbol{w}}}{\mathbf{1}^\top \bar{\boldsymbol{w}}} = \frac{1}{V_2 \sigma_2 + V_2 \sigma_1^2 - (V_1 + V_2) \rho_{12} \sigma_1 \sigma_2} \begin{pmatrix} V_1 \sigma_2^2 - V_2 \rho_{12} \sigma_1 \sigma_2 \\ V_2 \sigma_1^2 - V_1 \rho_{12} \sigma_1 \sigma_2 \end{pmatrix}.$$

Compare the first element of this vector with (5.1), the formula that gives the weight of the first of two risky assets in the tangency portfolio.

## 5.7 Is the theory useful?

This theory of portfolio selection could be used if  $N$  were small. We would need estimates of  $\boldsymbol{\mu}$  and  $\mathbf{\Omega}$ . These would be obtained from recent returns data. Of course, there is no guarantee that future returns will behave like returns in the past, but this is the working assumption.

The next section gives an example of using portfolio theory to allocate capital among the various international markets.

However, suppose that we were considering selecting a portfolio from all 500 stocks on the S&P index. Or, even worse, consider all 3000 stocks on the Russell index. Ugh! There would be  $(3000)(2999)/2 \approx 4.5$  million covariances to calculate. Moreover,  $\Omega$  would be 3000 by 3000 and its inverse is required. However, the most serious difficulty would not be the computations. It would be data collection.

Portfolio theory was an important theoretical development; Markowitz was awarded the Nobel Prize in economics for this work. However, a practical version of this theory awaited the work of Sharpe and Lintner. Sharpe, who was Markowitz's PhD student, shared the Nobel Prize with Markowitz.

Sharpe's CAPM asserts that the tangency portfolio is also the market portfolio. This is a tremendous simplification.

## 5.8 Example—Global Asset Allocation

This example is taken from *Efficient Asset Management* by Richard O. Michaud. The problem is to allocate capital to eight major classes of assets: U.S. stocks, U.S. government and corporate bonds, Euros, and the Canadian, French, German, Japanese, and U.K. equity market. The historic data used to estimate expected returns, variances, and covariances consisted of 216 months (Jan 1978 to Dec 1995) of index total returns in U.S. dollars for all eight asset classes and for U.S. 30-day T-bills.

The efficient frontier, with all weights constrained to be non-negative, was found by quadratic programming and is shown in Figure 5.12. There are three reference portfolios. Michaud states that

The index portfolio is roughly consistent with a capitalization weighted portfolio relative to a world equity benchmark for the

six equity markets. The current portfolio represent a typical U.S.-based investor's global portfolio asset allocation. ... An equal weighted portfolio is useful as a reference point.

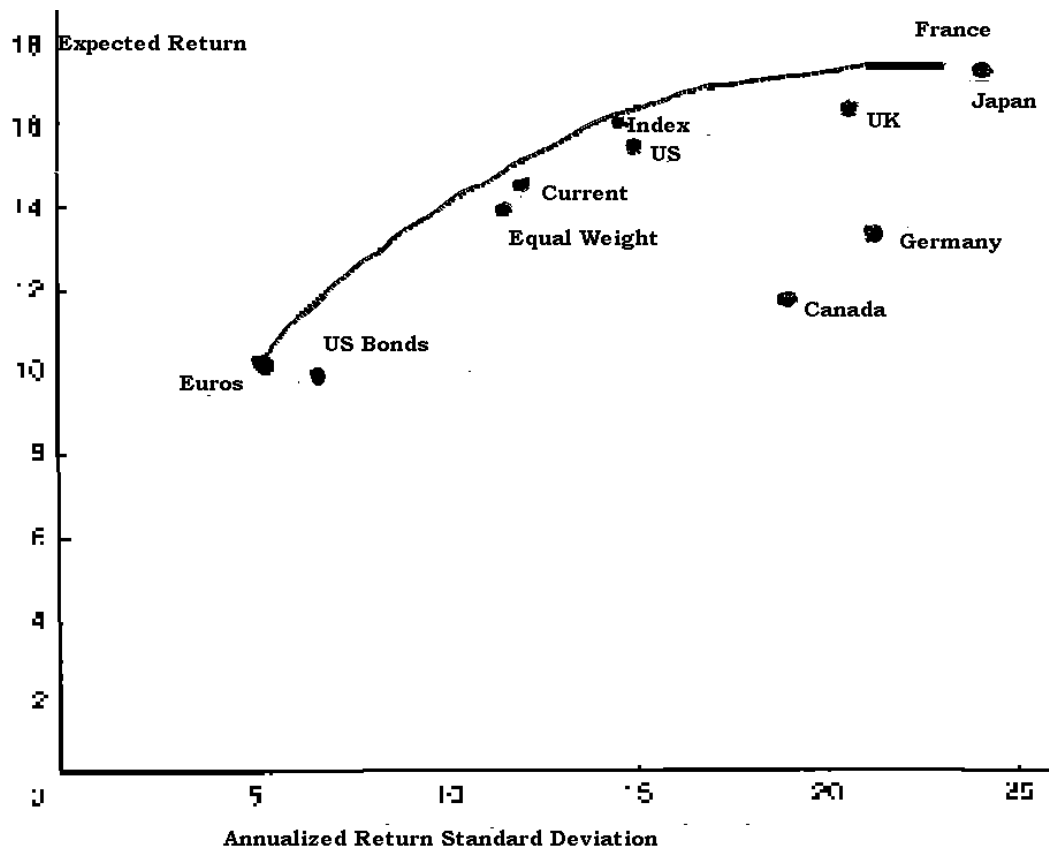


Figure 5.12: *Efficient frontier for the global asset allocation problem.*



## 5.9 Quadratic programming

Quadratic programming can be used to solve problems such as minimizing

$$\frac{1}{2} \mathbf{x}^\top \mathbf{H} \mathbf{x} + \mathbf{f}^\top \mathbf{x}$$

subject to

$$\mathbf{A} \mathbf{x} \leq \mathbf{b},$$

and

$$\mathbf{A}_{eq} \mathbf{x} = \mathbf{b}_{eq}.$$

Here for some  $N$ ,  $\mathbf{x}$  and  $\mathbf{f}$  are  $N \times 1$  vectors and  $\mathbf{H}$  is an  $N \times N$  matrix. Also,  $\mathbf{A}$  is  $m \times N$  and  $\mathbf{b}$  is  $m \times 1$  for some  $m$ , while  $\mathbf{A}_{eq}$  is  $n \times N$  and  $\mathbf{b}_{eq}$  is  $n \times 1$  for some  $n$ .

We can impose nonnegativity constraints on the weights of a portfolio by solving the minimization problem above with  $\mathbf{x} = \boldsymbol{\omega}$ ,  $\mathbf{H} = \boldsymbol{\Omega}$ ,  $\mathbf{f}$  equal to a  $N \times 1$  vector of zeros,  $\mathbf{A} = -\mathbf{I}$  (the  $N \times N$  identity matrix),  $\mathbf{b}$  equal to a  $N \times 1$  vector of zeros,

$$\mathbf{A}_{eq} = \begin{pmatrix} \mathbf{1}^\top \\ \boldsymbol{\mu}^\top \end{pmatrix},$$

and

$$\mathbf{b}_{eq} = \begin{pmatrix} 1 \\ \mu_P \end{pmatrix}.$$

Here is the documentation for MATLAB's "quadprog" illustrating several ways that this program can be used. In our applications, e.g., in the program "portfolio02QP.m," we call the program "quadprog" with a command of the type "X=QUADPROG(H,f,A,b,Aeq,beq)". This can be seen in the listing of "portfolio02QP.m" which is given later.

QUADPROG Quadratic programming. X=QUADPROG(H,f,A,b) solves the quadratic programming problem:

```
min 0.5*x'*H*x + f'*x    subject to:  A*x <= b
x
```

`X=QUADPROG(H,f,A,b,Aeq,beq)` solves the problem above while additionally satisfying the equality constraints  $Aeq*x = beq$ .

`X=QUADPROG(H,f,A,b,Aeq,beq,LB,UB)` defines a set of lower and upper bounds on the design variables,  $X$ , so that the solution is in the range  $LB \leq X \leq UB$ . Use empty matrices for  $LB$  and  $UB$  if no bounds exist. Set  $LB(i) = -Inf$  if  $X(i)$  is unbounded below; set  $UB(i) = Inf$  if  $X(i)$  is unbounded above.

`X=QUADPROG(H,f,A,b,Aeq,beq,LB,UB,X0)` sets the starting point to  $X0$ .

`X=QUADPROG(H,f,A,b,Aeq,beq,LB,UB,X0,OPTIONS)` minimizes with the default optimization parameters replaced by values in the structure `OPTIONS`, an argument created with the `OPTIMSET` function. See `OPTIMSET` for details. Used options are `Display`, `Diagnostics`, `TolX`, `TolFun`, `HessMult`, `LargeScale`, `MaxIter`, `PrecondBandWidth`, `TypicalX`, `TolPCG`, and `MaxPCGIter`. Currently, only 'final' and 'off' are valid values for the parameter `Display` ('iter' is not available).

`X=QUADPROG(Hinfo,f,A,b,Aeq,beq,LB,UB,X0,OPTIONS,P1,P2,...)` passes the problem-dependent parameters  $P1, P2, \dots$  directly to the `HMFUN` function when `OPTIMSET('HessMult',HMFUN)` is set. `HMFUN` is provided by the user. Pass empty matrices for  $A$ ,  $b$ ,  $Aeq$ ,  $beq$ ,  $LB$ ,  $UB$ ,  $X0$ , `OPTIONS`, to use the default values.

`[X,FVAL]=QUADPROG(H,f,A,b)` returns the value of the objective function at  $X$ :  $FVAL = 0.5*X'*H*X + f'*X$ .

`[X,FVAL,EXITFLAG] = QUADPROG(H,f,A,b)` returns a string `EXITFLAG` that describes the exit condition of `QUADPROG`. If `EXITFLAG` is:  $> 0$  then `QUADPROG` converged with a solution  $X$ .  $0$  then the maximum number of iterations was exceeded (only occurs with large-scale method).  $< 0$  then the problem is unbounded, infeasible, or `QUADPROG` failed to converge with a solution  $X$ .

`[X,FVAL,EXITFLAG,OUTPUT] = QUADPROG(H,f,A,b)` returns a structure `OUTPUT` with the number of iterations taken in `OUTPUT.iterations`, the type of algorithm used in `OUTPUT.algorithm`, the number of conjugate gradient iterations (if used) in `OUTPUT.cgiterations`, and a measure of first order optimality (if used) in `OUTPUT.firstorderopt`.

`[X,FVAL,EXITFLAG,OUTPUT,LAMBDA]=QUADPROG(H,f,A,b)` returns the set of Lagrangian multipliers `LAMBDA`, at the solution: `LAMBDA.ineqlin` for the linear inequalities  $A$ , `LAMBDA.eqlin` for the linear equalities  $Aeq$ , `LAMBDA.lower` for  $LB$ , and `LAMBDA.upper` for  $UB$ .

Here is the program "portfolio02QP.m":

```
% Input mean vector and covariance matrix of returns here

bmu = [.08;.03;.05] ;
bOmega = [ .3 .02 .01 ;
```

```

    .02 .15 .03 ;
    .01 .03 .18 ] ;
A = [ones(1,3);bmu'] ;
ngrid = 50 ;
muP = linspace(.03,.08,ngrid)' ;

icompute = 0 ;
if icompute == 1 ;
sigmaP = muP ;
sigmaP2 = sigmaP ;
omegaP = zeros(3,ngrid) ;
omegaP2 = omegaP ;

for i = 1:ngrid ;
omegaP(:,i) = quadprog(bOmega,zeros(3,1),-eye(3),...
zeros(3,1),A,[1;muP(i)]) ;
omegaP2(:,i) = quadprog(bOmega,zeros(3,1),zeros(1,3),...
0,A,[1;muP(i)]) ;

sigmaP(i) = sqrt(omegaP(:,i)'*bOmega*omegaP(:,i)) ;
sigmaP2(i) = sqrt(omegaP2(:,i)'*bOmega*omegaP2(:,i)) ;

end ;
end ;

fsize = 16 ;
figure(1)
clf
p = plot(sigmaP,muP,sigmaP2,muP,'--') ;
l=legend('no negative wts','unconstrained wts',4) ;
set(gca,'fontsize',fsize) ;
set(l,'fontsize',fsize) ;

xlabel('standard deviation of return (\sigma_P)',...

```

```
'fontsize',fsize) ;
ylabel('expected return (\mu_P)', 'fontsize',fsize) ;
text(sqrt(bOmega(1,1)),bmu(1),'1', 'fontsize',24) ;
text(sqrt(bOmega(2,2)),bmu(2),'2', 'fontsize',24) ;
text(sqrt(bOmega(3,3)),bmu(3),'3', 'fontsize',24) ;
set(gca,'ylim',[.025,.085]) ;

set(p,'linewidth',4) ;
grid ;

print portfolio02QP.ps -depsc ;
!mv portfolio02QP.ps ...
~/public_html/or473/LectNotes/portfolio02QP.ps ;

figure(2)
p2 = plot(muP,omegaP(1,:),muP,omegaP(2,:), '--',muP,...
omegaP(3,:), '-.') ;
set(p2,'linewidth',6) ;
set(gca,'fontsize',fsize) ;
grid ;
xlabel('\mu_P', 'fontsize',fsize) ;
ylabel('weight', 'fontsize',fsize) ;
legend('w_1', 'w_2', 'w_3',0) ;

print portfolio02_wtQP.ps -depsc ;
!mv portfolio02_wtQP.ps ...
~/public_html/or473/LectNotes/portfolio02_wtQP.ps ;
```

# Chapter 6

## The Capital Asset Pricing Model:

3/26/01

### 6.1 Introduction to CAPM

The CAPM (capital asset pricing model) has a variety of uses:

- It provides a theoretical justification for the widespread practice of “passive” investing known as *indexing*.
  - Indexing means holding a diversified portfolio in which securities are held in the same relative proportions as in a broad market index such as the S&P 500. Individual investors can do this easily by holding shares in an *index fund*.
- CAPM can provide estimates of expected rates of return on individual investments
- CAPM can establish “fair” rates of return on invested capital in regulated firms or in firms working on a cost-plus basis — what should the “plus” be?
- CAPM starts with the question, what would be the risk premiums on securities if the following assumptions were true?

- The market prices are “in equilibrium.”
  - \* In particular, for each asset, supply equals demand.
- Everyone has the same forecasts of expected returns and risks.
- All investors chose portfolios optimally according to the principles of efficient diversification discussed in Chapter 5.
  - \* This implies that everyone holds the tangency portfolio of risky assets.
- The market rewards people for assuming unavoidable risk, but there is no reward for needless risks due to inefficient portfolio selection.
  - \* Therefore, the risk-premium on a single security is not due to its “stand alone” risk, but rather to its contribution to the risk of the tangency portfolio.
    - The various components of risk will be discussed in Section 6.4.

As in Chapter 5, “return” can either refer to one-period net returns or one-period log returns.

Suppose that there are exactly three assets with a total market value of \$100 billion.

- Stock A: \$60 billion
- Stock B: \$30 billion
- risk-free: \$10 billion

The market portfolio of Stock A to Stock B is 2:1. CAPM says that under equilibrium, all investors will hold Stock A to Stock B in a 2:1 ratio. Therefore, the tangency portfolio puts weight  $2/3$  on Stock A and  $1/3$  on Stock B and *all* investors will have two-thirds of their allocation to risky assets in Stock A and one-third in Stock B.

Suppose there was too little of Stock A and too much of Stock B for everyone to have a 2:1 allocation. For example, suppose that there were one million shares of each stock and the price per share was \$60 for Stock A and \$40 for Stock B. Then the market portfolio must hold Stock A to Stock B in a 3:2 ratio, not 2:1. Not everyone could hold the tangency portfolio, though everyone would want to. Thus, prices would be in disequilibrium and would change. The price of Stock A would go up since the supply of Stock A is less than the demand. Similarly the price of Stock B would go down. As these prices changed, so would expected returns and the tangency portfolio would change. These changes in prices and expected returns would stop when the market portfolio was equal to the tangency portfolio, so that prices were in equilibrium. At least, this adjustment to equilibrium would happen under the ideal conditions of economic theory. The real world would be a little messier. The underlying message from theory, is however, correct. Prices adjust as all investors look for an efficient portfolio and supply and demand converge to each other.

The market portfolio is 9:1 risky to risk-free. In total, investors must hold risky to risk-free in a 9:1 ratio — they *are* the market.

For an individual investor, the risky:risk-free ratio will depend on that investor's risk aversion.

- At one extreme, a portfolio of all risk-free has a standard deviation of returns equal to 0
- At the other extreme, all risky assets, the standard deviation is maximized. (This assumes no margin. If we allow negative positions in the risk-free, then there is no limit to the risk)

At equilibrium, returns on risky and risk-free assets are such that aggregate demand for risk-free assets equal supply.

## 6.2 The capital market line (CML)

The capital market line (CML) relates the excess expected return on an efficient portfolio to its risk; “excess expected return” means the amount by which the expected return exceeds the risk-free rate of return. The CML is

$$\mu_R = \mu_f + \frac{\mu_M - \mu_f}{\sigma_M} \sigma_R, \quad (6.1)$$

where  $R$  is the return on a given efficient portfolio (mixture of the market portfolio and the risk-free asset),  $\mu_R = E(R_M)$ ,  $\mu_f$  is the rate of return on the risk-free asset,  $R_M$  is the return on the market portfolio,  $\sigma_M$  is the standard deviation of the return on the market portfolio, and  $\sigma_R$  is the standard deviation of return on the portfolio. The slope of the CML is, of course,

$$\frac{\mu_M - \mu_f}{\sigma_M}$$

which can be interpreted as the ratio of the “risk premium” to the standard deviation of the market portfolio. This is Sharpe’s “reward-to-risk ratio.” Equation (6.1) can be rewritten as

$$\frac{\mu_R - \mu_f}{\sigma_R} = \frac{\mu_M - \mu_f}{\sigma_M},$$

which says that the reward-to-risk ratio for any efficient portfolio equals that ratio for the market portfolio.

**Example:** Suppose that risk-free rate of interest is  $\mu_f = 0.06$ , that the expected return on the market portfolio is  $\mu_M = .15$ , and the risk of the market portfolio is  $\sigma_M = 0.22$ . Then the slope of the CML is  $(.15 - .06)/.22 = 9/22$ . The CML of this example is illustrated in Figure 6.1.

The CML is easy to derive. Consider an efficient portfolio that allocates a proportion  $w$  of its assets to the market portfolio and  $(1 - w)$  to the risk-free asset. Then

$$R = wR_M + (1 - w)\mu_f = \mu_f + w(R_M - \mu_f).$$



Therefore,

$$\mu_R = \mu_f + w(\mu_R - \mu_f). \quad (6.2)$$

Also,

$$\sigma_R = w\sigma_M,$$

or

$$w = \frac{\sigma_R}{\sigma_M}. \quad (6.3)$$

Substituting (6.3) into (6.2) gives the CML.

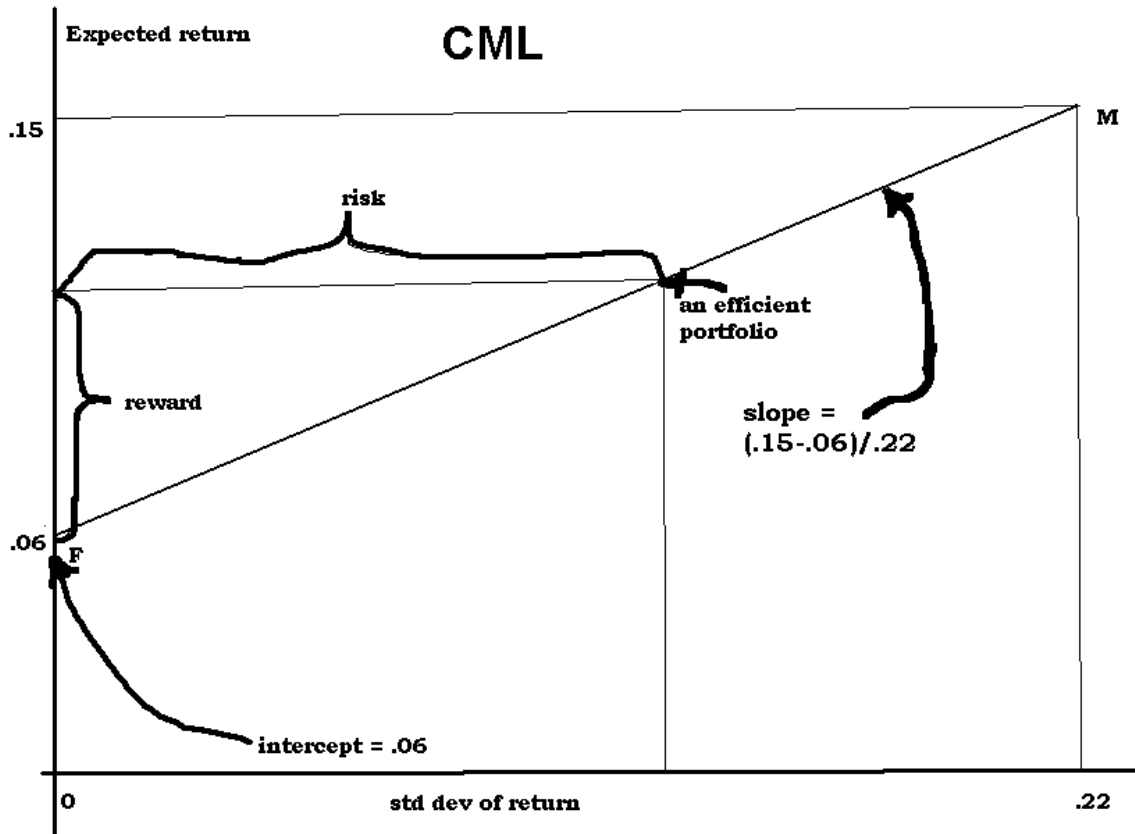


Figure 6.1: CML when  $\mu_f = 0.06$ ,  $\mu_M = 0.15$ , and  $\sigma_M = 0.22$ . All efficient portfolios are on the line connecting the risk-free asset (F) and the market portfolio (M). Therefore, the reward-to-risk ratio is the same for all efficient portfolios, including the market portfolio.

CAPM says that the optimal way to invest is to:

1. Decide on the risk  $\sigma_R$  that you can tolerate,  $0 \leq \sigma_R \leq \sigma_M$ . ( $\sigma_R > \sigma_M$  is possible by borrowing money to buy risky assets.)
2. Calculate  $w = \sigma_R / \sigma_M$ .
3. Invest  $w$  proportion of your investment in an index fund, i.e., a fund that tracks the index.
4. Invest  $1 - w$  proportion of your investment in risk-free treasury bills, or a money-market fund that invests in T-bills.

Alternatively,

1. Choose the reward  $\mu_R - \mu_f$  that you want.

2. Calculate

$$w = \frac{\mu_R - \mu_f}{\mu_M - \mu_f}.$$

3. Do steps 3 and 4 as above.

One can view  $w = \sigma_R / \sigma_M$  as is an index of the risk aversion of the investor. The smaller the value of  $w$  the more risk averse the investor. If an investor has  $w$  equal to 0, then that investor is 100% in risk-free assets. Similarly, an invest with  $w = 1$  is totally invested in the tangency portfolio of risky assets.

## 6.3 Betas and the Security Market Line

The Security Market Line (SML) relates the excess return on an asset to the slope of its regression on the market portfolio.

Suppose that there are many securities indexed by  $j$ . Define

$\sigma_{jM}$  = covariance between the  $j$ th security and the market portfolio.

Also, define

$$\beta_j = \frac{\sigma_{jM}}{\sigma_M^2}.$$

It follows from regression theory that  $\beta_j$  is the slope when the  $j$ th security's returns are regressed on the returns of the market portfolio. This fact follows from equation (2.1) for the slope of a best linear prediction equation. Another way to appreciate this fact is to suppose that we have a bivariate time series  $(R_{jt}, R_{Mt})_{t=1}^n$  of returns on the  $j$ th asset and the market portfolio. Then, the estimated slope of the regression of  $R_{jt}$  on  $R_{Mt}$  is

$$\hat{\beta}_j = \frac{\sum_{t=1}^n (R_{jt} - \bar{R}_j)(R_{Mt} - \bar{R}_M)}{\sum_{t=1}^n (R_{Mt} - \bar{R}_M)^2},$$

which is an estimate of  $\sigma_{jM}$  divided by an estimate of  $\sigma_M^2$ .

Let  $\mu_j$  be the expected return on the  $j$ th security. Then  $\mu_j - \mu_f$  is the "risk premium" (or "reward for risk" or "excess expected return") for that security. Using CAPM, it can be shown that

$$\mu_j - \mu_f = \beta_j(\mu_M - \mu_f). \quad (6.4)$$

This equation, which is called the security market line (SML), will be derived in Section 6.5.2. In (6.4)  $\beta_j$  is a variable in the linear equation, not the slope; more precisely,  $\mu_j$  is a linear function of  $\beta_j$  with slope  $\mu_M - \mu_f$ . This point is worth remembering. Otherwise, there could be some confusion since  $\beta_j$  was defined earlier as a slope of a regression model. In other words,  $\beta_j$  is a slope in one context but is the independent variable in the SML.

The SML says that the risk premium of the  $j$ th asset is the product of its beta ( $\beta_j$ ) and the risk premium of the market portfolio ( $\mu_M - \mu_f$ ).  $\beta_j$  measures both the riskiness of the  $j$ th asset and the reward for assuming that riskiness.  $\beta_j$  is, therefore, a measure of how "aggressive" the  $j$ th asset is. By definition, the beta for the market portfolio is 1, i.e.,  $\beta_M = 1$ . Therefore,

$$\begin{aligned} \beta_j > 1 &\Rightarrow \text{"aggressive"} \\ \beta_j = 1 &\Rightarrow \text{"average risk"} \\ \beta_j < 1 &\Rightarrow \text{"not aggressive"}. \end{aligned}$$

Figure 6.2 illustrated the SML and an asset, J, that is not on the SML. This asset contradicts the CAPM; according to CAPM no such asset exists.

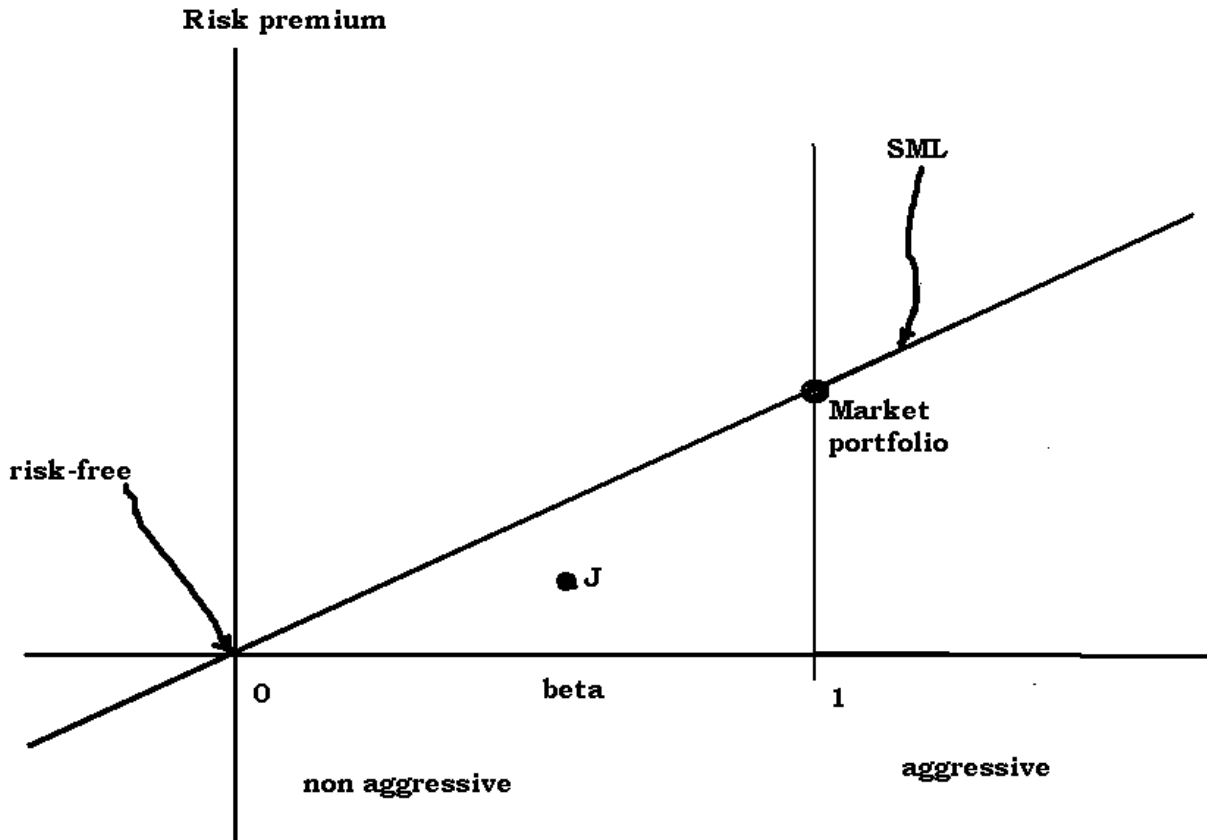


Figure 6.2: Security market line (SML) showing that the risk premium of an asset is a linear function of the asset's beta. J is a security not on the line and a contradiction to CAPM. Theory predicts that the price of J will decrease until J is on the SML.

Consider what would happen if an asset like J did exist. Investors would not want to buy it because its risk premium is too low. They would invest less in J and more in other securities. Therefore the price of J would

decline and its expected return would increase. After that increase, the asset J would be on the SML, or so the theory predicts. In other words, J is *mispriced* according to CAPM.

Stock (symbol)	Industry	Stock's $\beta$	Ind's $\beta$
Celanese (CZ)	Synthetics	0.13	0.86
General Mills (GIS)	Food - major diversif	0.29	0.39
Kellogg (K)	Food - major, diversif	0.30	0.39
Proctor & Gamble (PG)	Cleaning Prod	0.35	0.40
Exxon-Mobil (XOM)	Oil/gas	0.39	0.56
7-Eleven (SE)	Grocery stores	0.55	0.38
Merck (Mrk)	Major drug manuf	0.56	0.62
McDonalds (MCD)	Restaurants	0.71	0.63
McGraw-Hill (MHP)	Pub - books	0.87	0.77
Ford (F)	Auto	0.89	1.00
Aetna (AET)	Health care plans	1.11	0.98
General Motors (GM)	Major auto manuf	1.11	1.09
AT&T (T)	Long dist carrier	1.19	1.34
General Electric (GE)	Conglomerates	1.22	0.99
Genentech (DNA)	Biotech	1.43	0.69
Microsoft (MSFT)	Software applic.	1.77	1.72
Cree (Cree)	Semicond equip	2.16	2.30
Amazon (AMZN)	Net soft & serv	2.99	2.46
DoubleClick (Dclk)	Net soft & serv	4.06	2.46

### 6.3.1 Examples of betas

Netscape's home page has a link to stock quotes from Salomon Smith Barney. If you request a quote on a stock, you will be given menu for choosing further information about the company. Under "profile" you will find the five-year beta of the company, its industry, and the S&P 500. Table 6.3.1 has some "five-year betas" that I took from net on February 27 and March 5, 2001. The beta for the S&P 500 is given as 1.00; why?

### 6.3.2 Comparison of the CML with the SML

The CML applies only to the return  $R$  of an efficient portfolio. It can be arranged so as to relate the excess expect return of that portfolio to the

excess expected return of the market portfolio:

$$\mu_R - \mu_f = \left( \frac{\sigma_R}{\sigma_M} \right) (\mu_M - \mu_f). \quad (6.5)$$

The SML applies to *any* asset and like the CML relates its excess expected return to the excess expected return of the market portfolio:

$$\mu_j - \mu_f = \beta_j (\mu_M - \mu_f). \quad (6.6)$$

If we take an efficient portfolio and consider it as an asset, then  $\mu_R$  and  $\mu_j$  both denote the expected return on that portfolio/asset. Both (6.5) and (6.6) hold so that

$$\frac{\sigma_R}{\sigma_M} = \beta_R.$$

## 6.4 The security characteristic line

Let  $R_{jt}$  be the return at time  $t$  on the  $j$ th asset. Similarly, let  $R_{Mt}$  and  $\mu_{ft}$  be the return on the market portfolio and the risk-free return at time  $t$ . The security characteristic line (sometimes shortened to the characteristic line) is a regression model:

$$R_{jt} = \mu_{ft} + \beta_j (R_{Mt} - \mu_{ft}) + \epsilon_{jt}, \quad (6.7)$$

where  $\epsilon_{jt}$  is  $N(0, \sigma_{\epsilon_j}^2)$ . It is often assumed that the  $\epsilon_{jt}$ 's are uncorrelated across assets, that is, that  $\epsilon_{jt}$  is uncorrelated with  $\epsilon_{j't}$  for  $j \neq j'$ . This assumption has important ramifications for risk reduction by diversification; see Section 6.4.1.

Let  $\mu_j = E(R_{jt})$  and  $\mu_M = E(R_{Mt})$ . Taking expectations in (6.7) we get

$$\mu_j = \mu_f + \beta_j (\mu_M - \mu_f),$$

which is our friend the SML again. The SML gives us information about expected returns, but not about the variance of the returns. For the latter we need the characteristic line. The characteristic line is said to be a



“returning generating process” since it gives us a probability model of the returns, not just a model of their expected values.

An analogy to the distinction between the SML and characteristic line is this. The regression line  $E(Y|X) = \beta_0 + \beta_1 X$  gives of the expected value of  $Y$  given  $X$  but not the conditional probability distribution of  $Y$  given  $X$ . The regression model

$$Y_t = \beta_0 + \beta_1 X_t + \epsilon_t, \text{ and } \epsilon_t \sim N(0, \sigma^2)$$

does give us this conditional probability distribution.

The characteristic line implies that

$$\sigma_j^2 = \beta_j^2 \sigma_M^2 + \sigma_{\epsilon_j}^2,$$

$$\sigma_{jj'} = \beta_j \beta_{j'} \sigma_M^2$$

for  $j \neq j'$ , and that

$$\sigma_{Mj} = \beta_j \sigma_M^2.$$

The total risk of the  $j$ th asset is

$$\sqrt{\beta_j^2 \sigma_M^2 + \sigma_{\epsilon_j}^2}.$$

The risk has two components:  $\beta_j^2 \sigma_M^2$  is called the market or systematic component of risk and  $\sigma_{\epsilon_j}^2$  is called the unique, nonmarket, or unsystematic component of risk.

### 6.4.1 Reducing unique risk by diversification

The market component cannot be reduced by diversification, but the unique component can be reduced in this way.

Suppose that there are  $N$  assets with returns  $R_{1t}, \dots, R_{Nt}$  for holding period  $t$ . If we form a portfolio with weights  $w_1, \dots, w_N$  then the return of the portfolio is

$$R_{Pt} = w_1 R_{1t} + \dots + w_N R_{Nt}.$$

Let  $R_{Mt}$  be the return on the market portfolio. According to the characteristic line model  $R_{jt} = \mu_{ft} + \beta_j(R_{Mt} - \mu_{ft}) + \epsilon_{jt}$  so that

$$R_{Pt} = \mu_{ft} + \left( \sum_{j=1}^N \beta_j w_j \right) (R_{Mt} - \mu_{ft}) + \sum_{j=1}^N w_j \epsilon_{jt}.$$

Therefore, the portfolio beta is

$$\beta_P = \sum_{j=1}^N w_j \beta_j,$$

and the “epsilon” for the portfolio is

$$\epsilon_{Pt} = \sum_{j=1}^N w_j \epsilon_{jt}.$$

We will now assume that  $\epsilon_{1t}, \dots, \epsilon_{Nt}$  are uncorrelated. Therefore,

$$\sigma_{\epsilon_P}^2 = \sum_{j=1}^N w_j^2 \sigma_{\epsilon_j}^2.$$

### Example

Suppose that  $w_j = 1/N$  for all  $j$ . Then

$$\beta_P = \frac{\sum_{j=1}^N \beta_j}{N},$$

and

$$\sigma_{\epsilon_P}^2 = \frac{N^{-1} \sum_{j=1}^N \sigma_{\epsilon_j}^2}{N} = \frac{\bar{\sigma}_{\epsilon}^2}{N},$$

where  $\bar{\sigma}_{\epsilon}^2$  is the average of the  $\sigma_{\epsilon_j}^2$ . If  $\sigma_{\epsilon_j}^2$  is a constant, say  $\sigma_{\epsilon}^2$  for all  $j$ , then

$$\sigma_{\epsilon_P} = \frac{\sigma_{\epsilon}}{\sqrt{N}}.$$

For example, suppose that  $\sigma_{\epsilon}$  is 5%. If  $N = 20$ , then  $\sigma_{\epsilon_P}$  is 1.12%. If  $N = 100$ , then  $\sigma_{\epsilon_P}$  is 0.5%. There are approximately 1600 stocks on the NYSE; if  $N = 1600$ , then  $\sigma_{\epsilon_P} = 0.0125\%$ .

**Are the assumptions sensible?**

A key assumption that allows risk to be removed by diversification is that  $\epsilon_{1t}, \dots, \epsilon_{Nt}$  are uncorrelated. This assumption implies that *all* correlation among the cross-section<sup>1</sup> of asset returns is due to a single cause and that cause is measured by the market index. For this reason, the characteristic line is a “single factor” or “single index” model with  $R_{Mt}$  the “factor.”

This assumption of uncorrelated  $\epsilon_{jt}$  would not be valid if, say, two energy stocks are correlated over and beyond their correlation due to the market index. In this case, unique risk could not be eliminated by holding a large portfolio of all energy stocks.

**6.5 Some theory**

In this section we will show that  $\sigma_{jM}$  quantifies the contribution of the  $j$ th asset to the risk of the market portfolio. Also, we will derive the SML.

**6.5.1 Contributions to the market portfolio’s risk**

Suppose that we have  $N$  risky assets and that  $w_{1M}, \dots, w_{NM}$  are the weights of the market portfolio. Since

$$R_{Mt} = \sum_{i=1}^N w_{iM} R_{it},$$

the correlation between the return on the  $j$ th asset and the return on the market portfolio is

$$\sigma_{jM} = \text{Cov} \left( R_{jt}, \sum_{i=1}^N w_{iM} R_{it} \right) = \sum_{i=1}^N w_{iM} \sigma_{ij}.$$

Therefore,

$$\sigma_M^2 = \sum_{j=1}^N \sum_{i=1}^N w_{jM} w_{iM} \sigma_{ij} = \sum_{j=1}^N w_{jM} \sigma_{jM}. \quad (6.8)$$

---

<sup>1</sup>“Cross-section” returns means returns across assets within a *single* holding period.

Equation (6.8) shows that  $w_{jM}\sigma_{jM}$  is the contribution of the  $j$ th asset to the risk of the market portfolio.

### 6.5.2 Derivation of the SML

Consider a portfolio P with weight  $w_i$  given to the  $i$ th risky asset and weight  $(1 - w_i)$  given to the market portfolio. The return on this portfolio is

$$R_{Pt} = w_i R_{it} + (1 - w_i) R_{Mt}.$$

The expected return is

$$\mu_P = w_i \mu_i + (1 - w_i) \mu_M,$$

and the risk is

$$\sigma_P = \sqrt{w_i^2 \sigma_i^2 + (1 - w_i)^2 \sigma_M^2 + 2w_i(1 - w_i)\sigma_{iM}}.$$

As we vary  $w_i$  we get the locus of points on  $(\sigma, \mu)$  space that is shown as a blue curve in Figure 6.3.

**Key idea:** The derivative of this locus of points evaluated at the market portfolio is equal to the slope of the CML. We can calculate this derivative and equate it to the slope of the CML to see what we get. The result will be the SML.

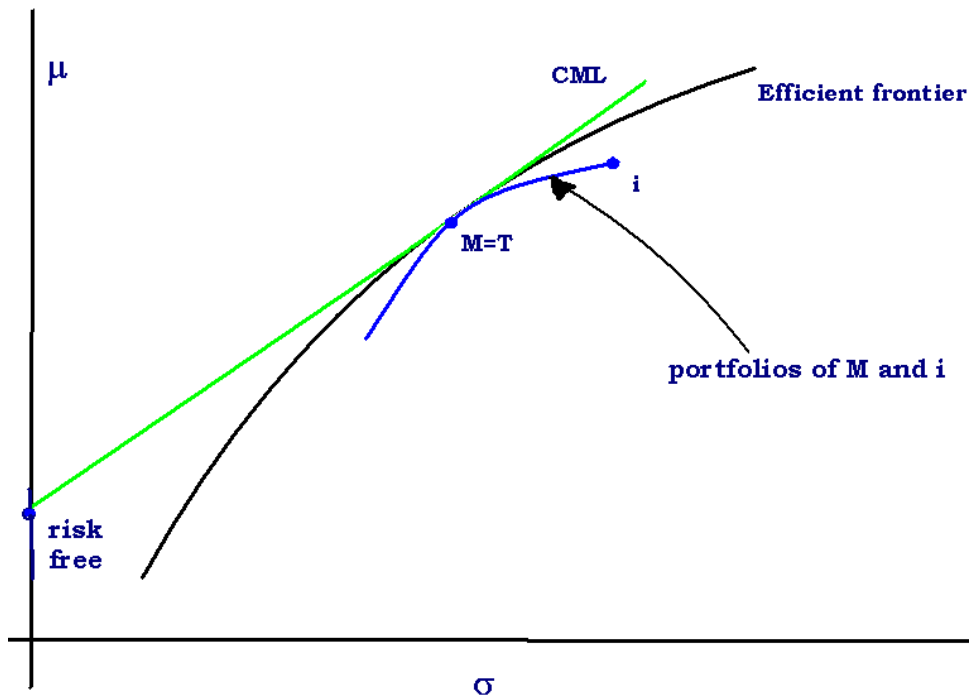


Figure 6.3: Derivation of the SML.  $M$  is the market portfolio and  $T$  is the tangency portfolio; they are equal according to the CAPM. The blue curve is the locus of portfolios combining asset  $i$  and the market portfolio. The derivative of this curve at  $M$  is equal to the slope of the CML, since this curve is tangent to the CML at  $M$ .

We have

$$\frac{d\mu_P}{dw_i} = \mu_i - \mu_M,$$

and

$$\frac{d\sigma_P}{dw_i} = \frac{1}{2}\sigma_P^{-1} \left\{ 2w_i\sigma_i^2 - 2(1-w_i)\sigma_M^2 + 2(1-2w_i)\sigma_{iM} \right\}.$$

Therefore,

$$\frac{d\mu_P}{d\sigma_P} = \frac{d\mu_P/dw_i}{d\sigma_P/dw_i} = \frac{(\mu_i - \mu_P)\sigma_P}{w_i\sigma_i^2 - \sigma_M^2 + w_i\sigma_M^2 + \sigma_{iM} - 2w_i\sigma_{iM}}.$$

Next,

$$\left. \frac{d\mu_P}{d\sigma_P} \right|_{w_i=0} = \frac{(\mu_i - \mu_M)\sigma_M}{\sigma_{iM} - \sigma_M^2}.$$

But

$$\left. \frac{d\mu_P}{d\sigma_P} \right|_{w_i=0}$$

must equal the slope of the CML which is  $(\mu_M - \mu_f)/\sigma_M$ . Therefore,

$$\frac{(\mu_i - \mu_M)\sigma_M}{\sigma_{iM} - \sigma_M^2} = \frac{\mu_M - \mu_f}{\sigma_M}.$$

which, after some algebra, gives us

$$\mu_i - \mu_f = \frac{\sigma_{iM}}{\sigma_M^2}(\mu_M - \mu_f) = \beta_j(\mu_M - \mu_f)$$

which is the SML given in equation (6.4).

## 6.6 Estimation of beta and testing the CAPM

Recall the security characteristic line

$$R_{jt} = \mu_{ft} + \beta_j(R_{Mt} - \mu_{ft}) + \epsilon_{jt}, \quad (6.9)$$

Let  $R_{jt}^* = R_{jt} - \mu_{ft}$  be the excess return on the  $j$ th security and let  $R_{Mt}^* = R_{Mt} - \mu_{ft}$  be the excess return on the market portfolio. Then (6.9) can be written as

$$R_{jt}^* = \beta_j R_{Mt}^* + \epsilon_{jt}. \quad (6.10)$$

Equation (6.10) is a regression model without an intercept and with  $\beta_j$  as the slope. A more elaborate model is

$$R_{jt}^* = \alpha_j + \beta_j R_{Mt}^* + \epsilon_{jt}. \quad (6.11)$$

which includes an intercept.

Given series  $R_{jt}$ ,  $R_{Mt}$ , and  $\mu_{ft}$  for  $t = 1, \dots, n$ , we can calculate  $R_{jt}^*$  and  $R_{Mt}^*$  and regress  $R_{jt}^*$  on  $R_{Mt}^*$  to estimate  $\alpha_j$ ,  $\beta_j$ , and  $\sigma_{\epsilon_j}^2$ .

By testing the null hypothesis that  $\alpha_j = 0$  we are testing whether the  $j$ th asset is mispriced according to the CAPM.

Here is an example done in MINITAB. The least squares line and the output from the regression command are shown below. The variable X:MS\_1 is the excess return on Microsoft. Five years of monthly data, March 1996 to February 2001, were used. The raw data are in the Excel file "Datstream01.xls" on the course home page. For Microsoft, we find that

$$\hat{\beta} = 1.44$$

and

$$\hat{\alpha} = .012.$$

Since the standard error of X:MS\_1 (i.e., of  $\hat{\beta}_j$ ) is 0.317, a 95% confidence interval for  $\beta_j$  is  $1.44 \pm (2)(.317)$  or  $(.81, 2.07)$ .<sup>2</sup> The p-value for X:MS\_1 is 0.000. This p-value is for testing the null hypothesis that  $\beta_j = 0$ , so it is not surprising that the null hypothesis is strongly rejected. We do not expect the beta of a stock to be zero.

The test that  $\alpha = 0$  has a p-value of 0.441 so we can accept the null hypothesis. This implies that the data are consistent with the CAPM. Moreover,

$$\hat{\sigma}_{\epsilon}^2 = 0.01381,$$

---

<sup>2</sup>Here "2" is used as an approximate t-value. The exact t-value is 2.0017. This value can be found in MINITAB. Go to the calc menu, then probability distributions, then "t." Use "inverse cumulative probability" with "noncentrality parameter" equal to 0 and "input constant" equal to .975 (for a 95% confidence interval).

the mean square residual error.

Here is the MINITAB plot.

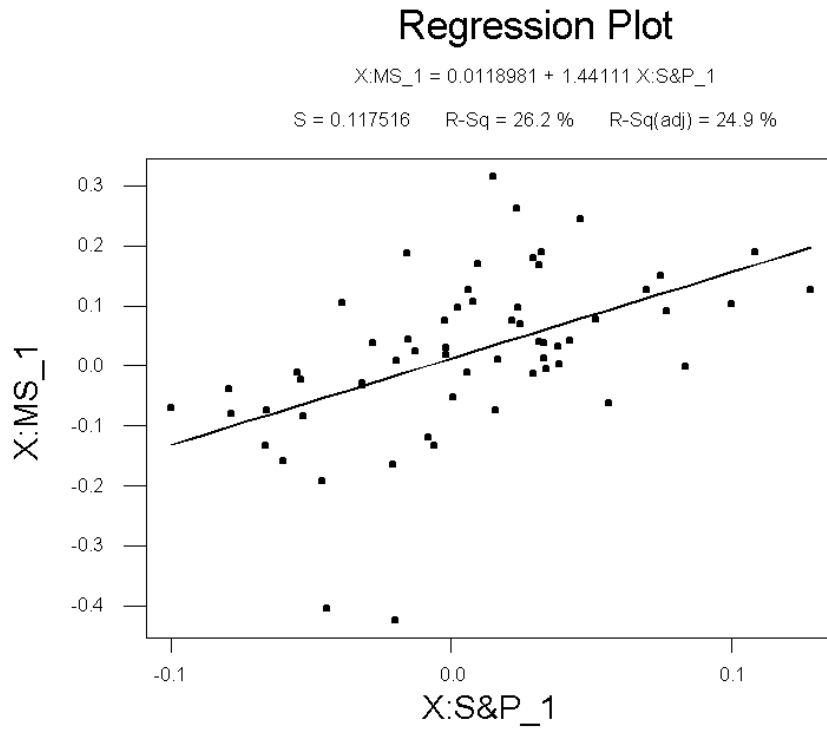


Figure 6.4: *Least squares line fit to the Microsoft data.*



Here is the MINITAB regression output.

### Model with intercept:

Regression Analysis: X:MS\_1 versus X:S&P\_1

The regression equation is  
 X:MS\_1 = 0.0119 + 1.44 X:S&P\_1

Predictor	Coef	SE Coef	T	P
Constant	0.01190	0.01532	0.78	0.441
X:S&P_1	1.4411	0.3174	4.54	0.000

S = 0.1175      R-Sq = 26.2%      R-Sq(adj) = 24.9%

### Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	0.28464	0.28464	20.61	0.000
Residual Error	58	0.80098	0.01381		
Total	59	1.08562			

### Unusual Observations

Obs	X:S&P_1	X:MS_1	Fit	SE Fit	Residual	St Resid
49	0.129	0.1258	0.1973	0.0416	-0.0715	-0.65 X
50	-0.044	-0.4040	-0.0518	0.0222	-0.3521	-3.05R
58	-0.020	-0.4249	-0.0166	0.0173	-0.4083	-3.51R
59	0.015	0.3164	0.0337	0.0154	0.2828	2.43R

R denotes an observation with a large standardized residual

X denotes an observation whose X value gives it large influence.

If we assume that  $\alpha = 0$ , then we can refit the model using a no intercept model. This is done with MINITAB's regression program by NOT choosing the "fit intercept" option; the default is to choose this option so you need to go to "options" and unchoose that option. Here is the MINITAB output when fitting a no intercept model.

Notice that the  $R^2$  (R-sq) value for the regression is 26.2%. The interpreta-

tion of  $R^2$  is the percent of the variance in the excess returns on Microsoft that is due to excess returns on the market. In other words, 26.2% of the risk is due to systematic or market risk ( $\beta_j^2 \sigma_M^2$ ). The remaining 73.8% is due to unique or nonmarket risk ( $\sigma_\epsilon^2$ ).

**Model without intercept:**

Regression Analysis: X:MS\_1 versus X:S&P\_1

The regression equation is

$$X:MS\_1 = 1.48 X:S\&P\_1$$

Predictor	Coef	SE Coef	T	P
Noconstant				
X:S&P_1	1.4755	0.3133	4.71	0.000

S = 0.1171

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	0.30432	0.30432	22.19	0.000
Residual Error	59	0.80931	0.01372		
Total	60	1.11362			

Unusual Observations

Obs	X:S&P_1	X:MS_1	Fit	SE Fit	Residual	St Resid
15	0.100	0.1040	0.1479	0.0314	-0.0439	-0.39 X
30	-0.100	-0.0705	-0.1473	0.0313	0.0767	0.68 X
33	0.109	0.1906	0.1606	0.0341	0.0300	0.27 X
45	0.084	-0.0017	0.1240	0.0263	-0.1257	-1.10 X
49	0.129	0.1258	0.1899	0.0403	-0.0640	-0.58 X
50	-0.044	-0.4040	-0.0653	0.0139	-0.3387	-2.91R
58	-0.020	-0.4249	-0.0292	0.0062	-0.3958	-3.38R
59	0.015	0.3164	0.0223	0.0047	0.2941	2.51R

R denotes an observation with a large standardized residual

X denotes an observation whose X value gives it large influence.

Now  $\hat{\beta} = 1.48$  and  $\hat{\sigma}_\epsilon^2 = 0.0137$  which are small changes from the value from the intercept model.

**Regression using returns instead of excess returns**

Often, as an alternative to regression using excess returns, the returns on the asset are regressed on the returns on the asset.

Figure 6.5 show the least squares line using returns instead of excess returns. The estimate of beta has changed from 1.44 to 1.38. The new value, 1.38, is well within the old confidence interval of (.81, 2.07) showing that there is little difference between using returns and using excess returns.

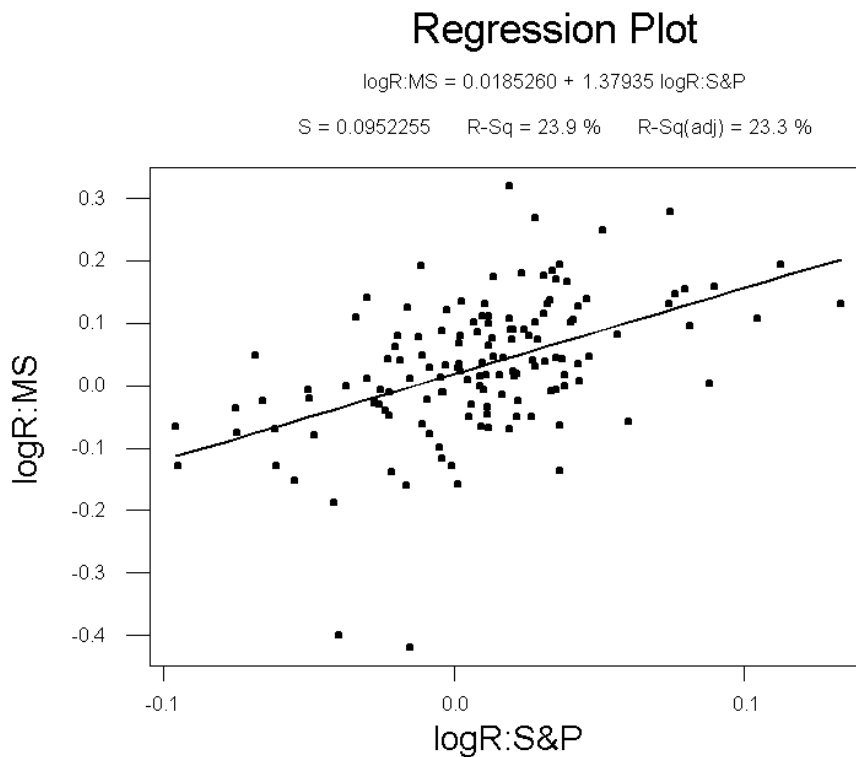


Figure 6.5:

### 6.6.1 Interpretation of alpha

If  $\alpha$  is nonzero then the security is mispriced, at least according to CAPM. If  $\alpha > 0$  then the security is underpriced; the returns are too large on average. This is an indication of an asset worth purchasing. Of course, one must be careful. If we reject the null hypothesis that  $\alpha = 0$ , all we have done is shown that the security was mispriced *in the past*.

Since for the Microsoft data we accepted the null hypothesis that  $\alpha$  is zero, there is no evidence that Microsoft was mispriced.

## 6.7 Summary

The CAPM assumes that prices are in equilibrium, that everyone has the same forecasts of returns, and that everyone uses the principles of portfolio selection introduced in Chapter 5.

The CAPM assumptions imply that everyone will hold risk efficient portfolios which mix the tangency portfolio and risk-free assets. This fact implies that the market portfolio will equal the tangency portfolio. A further consequence is that the Sharpe ratio for any efficient portfolio will equal the Sharpe ratio of the market portfolio:

$$\frac{\mu_R - \mu_f}{\sigma_R} = \frac{\mu_M - \mu_f}{\sigma_M}, \quad (6.12)$$

where R is any efficient portfolio. Equation (6.12) can be rearranged to give the CML which is

$$\mu_R = \mu_f + \frac{\mu_M - \mu_f}{\sigma_M} \sigma_R.$$

The CML applies only to efficient portfolios.

Another consequence of CAPM assumptions is the SML which applies to any security, say the  $j$ th, and is

$$\mu_j = \mu_f + (\mu_M - \mu_f)\beta_j.$$

Here  $\beta_j$  is the “independent variable” of this linear relationship and measures the riskiness of the security.  $\mu_j$  is the “dependent variable.”

The security characteristic line is a model for how actual returns are generated. (The SML only described expected returns.) The security characteristic line is

$$R_{jt} = \mu_f + \beta_j(R_{Mt} - \mu_f) + \epsilon_{jt}.$$

The variance of  $\epsilon_{jt}$  is  $\sigma_{\epsilon_j}^2$ . The security characteristic line implies that the risk of the  $j$ th asset can be decomposed into market and non market risks:

$$\sigma_j = \sqrt{\beta_j^2 \sigma_M^2 + \sigma_{\epsilon_j}^2}.$$

If one assumes that  $\epsilon_{jt}$  is uncorrelated with  $\epsilon_{j't}$  for  $j \neq j'$  (that is, for two different securities), then non market risk can be eliminated by portfolio diversification.

Since the security characteristic line is a regression model it can be used to estimate  $\beta_j$  and  $\sigma_{\epsilon_j}^2$ . The  $R^2$  value of the regression estimates the proportion of  $\sigma_j^2$  due to market risk, i.e., it estimates  $\beta_j^2 \sigma_M^2 / \sigma_j^2$ .



# Chapter 7

## Pricing Options: 4/12/01

### 7.1 Introduction

The European call options mentioned in Chapter 1 are one example of the many derivative now on the market. A *derivative* is a financial instrument whose value is derived from the value of some underlying instrument such an interest rate, foreign exchange rate, or stock price.

A call option gives one the right to buy a certain stock at the *exercise* or *strike* price, while a *put* option gives one the right to sell the stock at the exercise price. An option has an *exercise date*, which is also called the *strike date*, *maturity*, or *expiration date*. American options can be exercised at any time up to their exercise date, but European options can be exercised only at their exercise date. European options are easier to analyze than American options since one does not need to consider the possibility of early exercise.

In this chapter we will discuss the main ideas behind the pricing of options. We will not actually prove the Black-Scholes formula, since that derivation requires advanced mathematics. However, I will present a heuristic derivation of that formula to give an intuitive understanding of option pricing. For lack of time, we will only study European call options in detail. However, Black-Scholes type formulas exist for other derivatives as

well. We will give one example of a put option. The book *Introduction to Futures and Options Markets* by Hull is a nice overview of the options markets and the uses of options. *Financial Calculus* by Baxter and Rennie discuss the mathematics of many types of derivatives.

Why do companies purchase options and other derivatives? The answer is simple, to manage risk. In its 2000 Annual Report, the Coca Cola Company writes

Our company uses derivative financial instruments primarily to reduce our exposure to adverse fluctuations in interest rates and foreign exchange rates and, to a lesser extent, adverse fluctuations in commodity prices and other market risks. We do not enter into derivative financial instruments for trading purposes. As a matter of policy, all our derivative positions are used to reduce risk by hedging an underlying economic exposure. Because of the high correlation between the hedging instrument and the underlying exposure, fluctuations in the value of the instruments are generally offset by reciprocal changes in the value of the underlying exposure. The derivatives we use are straightforward instruments with liquid markets.

Derivatives can and have been used to speculate, but that is not their primary purpose. The intent of this quote is clear. The Company is assuring its stockholders that it is using derivatives to manage risk, not to gamble.

## 7.2 Call options

Suppose that you have purchased a European call option on 100 shares of Stock A with a exercise price of \$70. At the expiration date, suppose that Stock A is selling at \$73. The option allows you to purchase the 100 shares for \$70 and to immediately sell them for \$73, with a gain of \$300 on the 100 shares. Of course, the net profit for purchasing the option isn't \$300 since you had to pay for the option. If the option cost \$2/share, then you paid



\$200 for the option. Moreover, you paid the \$200 up front but only got the \$300 at the expiration date. Suppose that the expiration date was 3 months after the purchase date and the continuously compounded risk-free rate is 6% per annum or 1.5% for 3 months. Then the dollar value of your net profit is

$$\exp(-.015)300 - 200 = 95.53$$

at the time of purchase and is

$$300 - \exp(.015)200 = 96.98$$

at the exercise date.

We will use the notation  $(x)_+ = x$  if  $x > 0$  and  $= 0$  if  $x \leq 0$ . With this notation, the value of a call at exercise date is

$$(S_T - E)_+,$$

where  $E$  is the exercise data and  $S_T$  is stock's price on the exercise date,  $T$ .

A call is never exercised if the strike price is greater than the price of the stock, since exercising the option would amount to buying the stock for more than it would cost on the market. If a call is not exercised, then one loses the cost of the option.

One can lose money on an option even if it is exercised, because the amount gained by exercising the option might be less than the cost of the option. In the example above, if Stock A were selling for \$71 at the exercise date, then one would exercise the option and gain \$100. This would be less than the \$200 paid for the option. Even though exercising the option results in a loss, then loss is less than it would be if the option were not exercised. An option should always be exercised if  $(S_t - E)$  is positive.

### 7.3 The law of one price

The "law of one price" states that if two financial instruments have exactly the same payoffs, then they will have the same price. This principle is used

to price options. To value an option, one must find a portfolio or a self-financing<sup>1</sup> trading strategy with a *known price* and which has exactly the same payoffs as the option. The price of the option is then known; it must be the same as the price of the portfolio or self-financing trading strategy.

Here's a simple example of pricing by the law of one price. Suppose stock in company A sells at \$100/share. The risk-free rate of borrowing is 6% compounded annually. Consider a *futures contract* obliging one party to sell to the other party one share of Company A exactly one year from now at a price  $P$ . (No money changes hands now.) What is the fair market price, i.e., what should  $P$  be?

Note that this contract is not an option. The sale *must* take place. It would seem that  $P$  should depend on the expected price of company A stock one year from now. However, this is not the case. Consider the following strategy. The party that, one year from now, must sell the share of company A can borrow \$100 and buy one share now; this involves no capital since the share is purchased with borrowed money. A year from now that party sells the share for  $P$  dollars and pays back \$106 (principle plus interest) to the lender, who could be a third party. The profit is  $P - 106$ . The fair profit is 0 since no capital was used and there is no risk. Therefore,  $P$  should be \$106.

Consider what would happen if  $P$  were not \$106. You should be able to see that any other value of  $P$  besides \$106 would lead to unlimited risk-free profits. As investors rushed in to take advantage of this situation, the market would immediately correct the value of  $P$  to be \$106.

---

<sup>1</sup>A trading strategy is "self-financing" if it requires no investment other than the initial investment. After the initial investment, any further purchases of assets are financed by the sale of other assets or by borrowing.

### 7.3.1 Arbitrage

Arbitrage is the making of a guaranteed risk-free profit by trading in the market with no invested capital<sup>2</sup>. Speaking informally, arbitrage is a “free lunch.” The *arbitrage price* of a security is the price that guarantees no arbitrage opportunities. The law of one price is equivalent to stating that the market is free of arbitrage opportunities, i.e., that there are no free lunches. Arbitrage pricing is the same as pricing by the law of one price. The price of \$106 that we just derived in the example of the future contract is, therefore, the arbitrage price.

## 7.4 Time value of money and present value

“Time is money” is an old adage that is still true. A dollar a year from now is worth less to us than a dollar now. In finance it is essential that we be able to convert value in future payments to their present values, or vice versa. For example, we saw in Section 7.1 that the arbitrage enforced future price of a stock is simply the present price converted into a “future value” by multiplying by  $1 + r$ .

Let  $r$  be the risk-free annual interest rate. Then the “present value” of \$ $D$  dollars one year from now is  $\$D/(1 + r)$  without compounding or  $\$D \exp(-r)$  under continuous compounding. Another way of stating this is that \$ $D$  dollars now is worth  $\$(1+r)D$  dollars a year from now without compounding, or  $\$(\exp(r))D$  dollars a year from now under continuous compounding. When \$ $D$  is a future cash flow, then its present value is also called a *discounted* value and  $r$  is the discount rate.

The distinction between simple and compounding is not essential since an interest rate of  $r$  without compounding is equivalent to an interest rate of  $r'$  with continuously compounding where

$$1 + r = \exp(r')$$

---

<sup>2</sup>Investing in risk-free T-bills guarantees a positive net return but is *not* arbitrage since capital is invested.

so that

$$r = \exp(r') - 1 \text{ or } r' = \log(1 + r).$$

We will work with both simple and compound interest, whichever is most convenient.

### Examples

If  $r = 5\%$ , then  $r' = \log(1.05) = .0488$  or 4.88%. If  $r' = 4\%$ , then  $r = \exp(.04) - 1 = 1.0408 - 1$  or 4.08%. In general,  $r > r'$

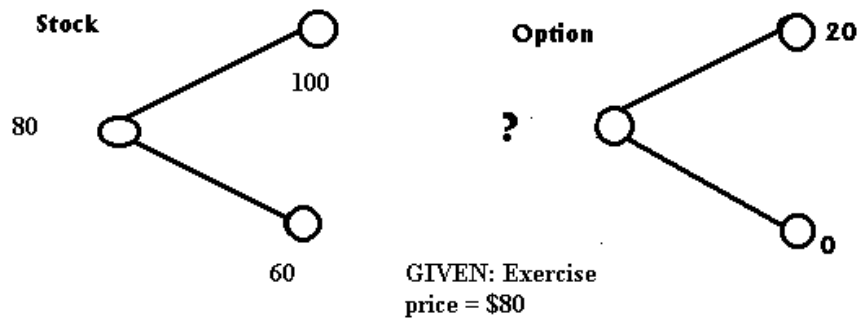
Occasionally, we will simplify life by making the unrealistic assumption that  $r = 0$  so that present and future values are equal. This simplifying assumption allows us to focus on other concepts besides discounting.

## 7.5 A simple binomial example

We will start our study of options with a very simple example. Suppose that a stock is currently selling for \$80. At the end of one time period it can either have increased to \$100 or decreased to \$60. What is the current value of a call option that allows one to purchase one share of the stock for \$80, the exercise price, after one time period?

At the end of the time period, the call option will be worth \$20 (\$100 – \$80) if the stock has gone up and worth \$0 dollars if the stock has gone down. See Figure 7.1. However, the question is “what is the option worth *now*?” This question is vital since the answer is, of course, the fair market price for the option at the current time.

One might think that the current value of the option depends on the probability that the stock will go up. However, this is not true. The current value of the option depends only on the rate of risk-free borrowing. For



**Hedge ratio =  $1/2$**   
**Buy  $1/2$  share**  
**Borrow: \$30**

**Initial value of portfolio =**  
 $1/2(80) - 30 = \$10$

**This must be the value of the call option.**

Figure 7.1: Example of one-step binomial option pricing. For simplicity, it is assumed that the interest rate is 0. The portfolio of  $1/2$  share of stock and  $-\$30$  of risk-free assets replicates the call option.

simplicity, we will assume that this rate is 0; later we will see how to value options when the rate of interest is positive. It turns out that the value of the option is \$10. How did I get this value?

Consider the following investment strategy. Borrow \$30 and buy one-half of a share of stock. The cost upfront is  $\$40 - \$30 = \$10$ , so the value now of the portfolio is \$10. If after one time period the stock goes up, then the portfolio is worth  $100/2 - 30 = 20$  dollars. If the stock goes down, then the portfolio is worth  $60/2 - 30 = 0$  dollars. Thus after one time period, the portfolio's value will be *exactly* the same as the value of the call option, no matter which way the stock moves. By the law of one price, the value of the call option *now* must be the same as the value *now* of the portfolio which is \$10.

Let's summarize what we have done. We have found a portfolio of the stock and the risk-free asset that *replicates* the call option. The current value of the portfolio is easy to calculate. Since the portfolio replicates the option, the option must have the same value as the portfolio.

Suppose we have just sold a call option. By purchasing this portfolio we have hedged the option. By hedging is meant that we have eliminated all risk, because the net return of selling the option and purchasing the portfolio is exactly 0 no matter what happens to the stock price.

How did I know that the portfolio should be 1/2 share of stock and  $-\$30$  in cash? I didn't use trial-and-error; that would have been tedious. Rather, I used the following logic. First, the volatility of the stock is  $\$100 - \$60 = \$40$  while the volatility of the option is  $\$20 - \$0 = \$20$ . The ratio of the volatility of the option to the volatility of the stock is 1/2; this is called the hedge ratio. If the portfolio is to exactly replicate the option, then the portfolio must have exactly the same volatility as the option; this means the portfolio must have one-half a share.

**Key point:** The number of shares in the portfolio must equal the hedge

ratio, where

$$\text{hedge ratio} = \frac{\text{volatility of option}}{\text{volatility of stock}}.$$

If the stock goes down, the portfolio is worth \$30 minus the amount borrowed. But we want the portfolio's value to equal that of the option, which is \$0. Thus, the amount borrowed is \$30.

**Key point:** We can determine the amount borrowed by equating the value of the portfolio when the stock goes down to the value of the option when the stock goes down. (Alternatively, we could equate the value of the portfolio to the value of the option when the stock goes up. This would tell us that \$50 minus the amount borrowed equals \$20, or that \$30 must be borrowed.)

Now suppose that the interest rate is 10%. Then, we borrow  $\$30/(1.1) = \$27.27$  so that the amount owed after one year is \$30. The cost of the portfolio is  $40 - 30/1.1 = \$12.7273$ . Thus, the value of the option is \$12.7273 if the risk-free rate of interest is 10%. This value is higher than the value of the option when the risk-free rate is 0, because the initial borrowing used by the self-financing strategy is more expensive when the interest rate is higher.

Here's how to value one-step binomial options for other values of the parameters. Suppose the current price is  $s_1$  and after one time period the stock either goes up to  $s_3$  or down to  $s_2$ . The exercise price is  $E$ . The risk-free rate of interest is  $r$ . It is assumed that  $s_2 < E < s_3$ , so the option is exercised if and only if the stock goes up.<sup>3</sup>

Then the hedge ratio is

$$\delta = \frac{s_3 - E}{s_3 - s_2}. \quad (7.1)$$

---

<sup>3</sup>If  $s_2 < s_3 \leq E$ , then the option will not be exercised under any circumstances. We are certain the option will be worthless, so its price must be 0. If  $E \leq s_2 < s_3$ , then the option will always be exercised; it really isn't an option, it is a futures contract. We have already seen how to value a futures contract—that was done in Section 7.3.

This is the number of shares of stock that are purchased; the cost is  $\delta s_1$ . The amount borrowed is

$$\frac{\delta s_2}{1+r}, \quad (7.2)$$

and the amount that will be paid back to the lender will be  $\delta S_2$ . Therefore, the price of the option is

$$\delta \left\{ s_1 - \frac{s_2}{1+r} \right\} = \frac{s_3 - E}{s_3 - s_2} \left\{ s_1 - \frac{s_2}{1+r} \right\}. \quad (7.3)$$

If the stock goes up, then the option is worth  $(s_3 - E)$  and the portfolio is also worth  $(s_3 - E)$ . If the stock goes down, both the option and the portfolio are worth 0. Thus, the portfolio does replicate the option.

### Example

In the example analyzed before,  $s_1 = 80$ ,  $s_3 = 100$ ,  $s_2 = 60$ , and  $E = 80$ . Therefore,

$$\delta = \frac{100 - 80}{100 - 60} = \frac{1}{2}.$$

The price of the option is

$$\frac{1}{2} \left\{ 80 - \frac{60}{1+r} \right\},$$

which is \$10 if  $r = 0$  and \$12.7273 if  $r = 0.1$ . The amount borrow is

$$\frac{\delta s_2}{1+r} = \frac{(1/2)60}{1+r} = \frac{30}{1+r},$$

which is \$30 if  $r = 0$  and \$27.27 if  $r = .1$ .

## 7.6 Two-step binomial option pricing

A one-step binomial model for a stock price may be realistic for very short maturities. For longer maturities, multiple-step binomial models are needed. A multiple-step model can be analyzed by analyzing the individual steps, going backwards in time.



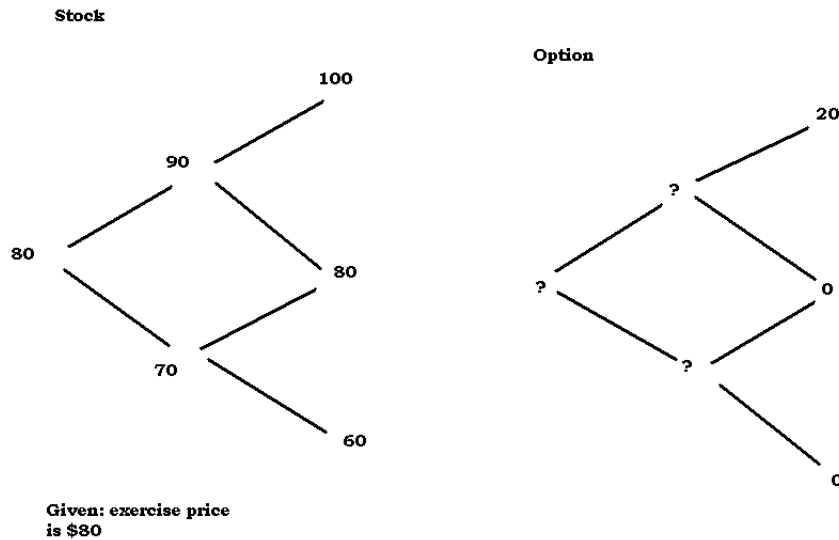


Figure 7.2: Two-step binomial model for option pricing.

To illustrate multi-step binomial pricing, consider the two-step model of a European call option in Figure 7.2. The option matures after the second step. The stock price can either go up \$10 or down \$10 on each step. Assume that  $r = 0$ .

Using the pricing principles just developed and working backwards, we can fill in the question marks in Figure 7.2. See Figure 7.3. For example, at node B, the hedge ratio is  $\delta = 1$  so we need to own one share which at this node is worth \$90. Also, we need to have borrowed  $\delta s_2 / (1 + r) = (1)(80) / (1 + 0) = \$80$  so that our portfolio has the same value at nodes E and F as the option, that is, the portfolio should be worth \$0 at node E and \$20 at node F. Since at node B we have stock worth \$90 and risk-free worth  $-\$80$ , the net value of our portfolio is \$10. By the same reasoning, at node C the hedge ratio is 0 and we should have no stock and no borrowing, so our portfolio is worth \$0.

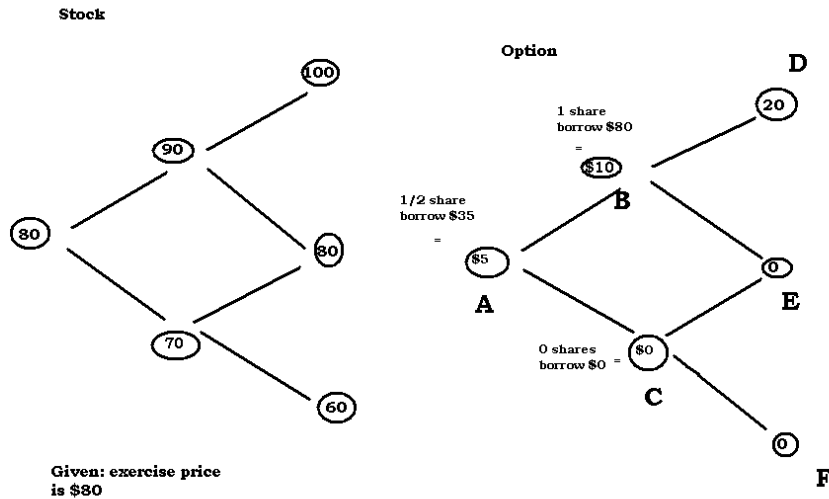


Figure 7.3: Pricing the option by backwards induction.

We can see in Figure 7.3 that at the end of the first step the option is worth \$10 if the stock is up (node B) and \$0 if it is down (node C). Applying one step pricing at node A, at the beginning of trading the hedge ratio is  $1/2$  and we should own  $1/2$  share of stock (worth \$40) and we should have borrowed \$35. Therefore, the portfolio is worth \$5 at node A, which proves that \$5 is the correct price of the option.

Note the need to work backwards. We could not apply one-step pricing at node A until we had already found the value of the portfolio (and of the option) at nodes B and C.

Let's show that our trading strategy is self-financing. To do this we need to show that we invest no money other than the initial \$5. Suppose that the stock is up on the first step, so we are at node B. Then our portfolio is worth  $\$90/2 - \$35$  or \$10. At this point we borrow \$45 and buy another half-share for \$45; this is self-financing. If the stock is down on the first

step, we sell the half share of stock for \$35 and buy off our debt; again the step is self-financing.

## 7.7 Arbitrage pricing by expectation

It was stated earlier that one prices an option by arbitrage, that is, the price is determined by the requirement that the market be arbitrage-free. The expected value of the option is *not* used to price the option. In fact, we do not even consider the probabilities that the stock moves up or down.

However, there is a remarkable result showing that arbitrage pricing *can* be done using expectations. More specifically, there exists probabilities of the stock moving up and down such that the arbitrage price of the option is equal to the expected value of the option according to these probabilities. Whether these are the “true” probabilities of the stock moving up or down is irrelevant. The fact is that these probabilities give the correct arbitrage price when they are used to calculate expectations.

Let “now” be time 0 and let “one step ahead” be time 1. Because of the time value of money, the present value of \$D dollars at time 1 is  $\$D/(1+r)$  where  $r$  is the interest rate. Let  $f(2) = 0$  and  $f(3) = s_3 - E$  be the values of the option if the stock moves up or down, respectively. We will now show that there is a value of  $q$  between 0 and 1, such that the present value of the option is

$$\frac{1}{1+r} \{qf(3) + (1-q)f(2)\}. \quad (7.4)$$

The quantity in (7.4) is the present value of the expectation of the option at time 1. To appreciate this, notice that the quantity in curly brackets is the value of the option if the stock goes up times  $q$ , which is the arbitrage determined “probability” that the stock goes up, plus the option’s value if the stock goes down times  $(1-q)$ . Thus the quantity in curly brackets is the expectation of the value’s option at the end of the holding period. Dividing by  $1+r$  converts this to a “present value.”

Okay, how do we find this magical value of  $q$ ? That's easy. We know that  $q$  must satisfy

$$\frac{1}{1+r}\{qf(3) + (1-q)f(2)\} = \frac{s_3 - E}{s_3 - s_2} \left\{ s_1 - \frac{s_2}{1+r} \right\}. \quad (7.5)$$

since the left hand side of this equation is (7.4) and the right-hand side is the value of the option according to (7.3). Substituting  $f(2) = 0$  and  $f(3) = s_3 - E$  into (7.5) we get an equation that can be solved for  $q$  to find that

$$q = \frac{(1+r)s_1 - s_2}{s_3 - s_2}. \quad (7.6)$$

We want  $q$  to be between 0 and 1 so that it can be interpreted as a probability. From (7.6) one can see that  $0 \leq q \leq 1$  if  $s_2 \leq (1+r)s_1 \leq s_3$ . Why should the latter hold? We will show that  $s_2 \leq (1+r)s_1 \leq s_3$  is required in order for the market to be arbitrage-free. If we invest  $s_1$  in a risk-free asset at time 0, then the value of our holdings at time 1 will be  $(1+r)s_1$ . If we invest  $s_1$  in the stock, then the value of our holdings at time 1 will be either  $s_2$  or  $s_3$ . If  $s_2 \leq (1+r)s_1 \leq s_3$  were not true, then there would be an arbitrage opportunity. For example, if  $(1+r)s_1 < s_2 \leq s_3$ , then could borrow at the risk-free rate and invest the borrowed money in the stock with a guaranteed profit; at time 1 we would pay back  $(1+r)s_1$  and receive *at least*  $s_2$  which is greater than  $(1+r)s_1$ .

**Exercise:** How would we make a guaranteed profit if  $s_2 \leq s_3 < (1+r)s_1$ ?

**Answer:** Sell the stock short and invest the  $s_1$  dollars in the risk-free asset. At the end of the holding period (maturity) receive  $(1+r)s_1$  from the risk-free investment and buy the stock for at most  $s_3 < (1+r)s_1$ .

Thus, the requirement that the market be arbitrage-free ensures that  $0 \leq q \leq 1$ .

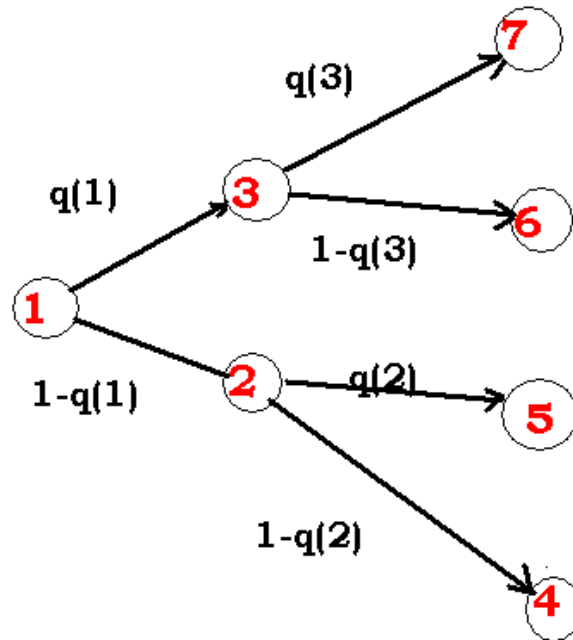


Figure 7.4: Two-step non-recombinant tree. The  $q(j)$  is the risk-neutral probability at node  $j$  of the stock moving upward.

## 7.8 A general binomial tree model

The material in this section follows Chapter 2 of *Financial Calculus* by Baxter and Rennie. Consider a possibly non-recombinant<sup>4</sup> tree as seen in Figure 7.4

Assume that:

- At the  $j$ th node the stock is worth  $s_j$  and the option is worth  $f(j)$ .

<sup>4</sup>The tree would be *recombinant* if the stock prices at nodes 5 and 6 were equal so that these two nodes could be combined.

- The  $j$ th node leads to either the  $2j + 1$ th node or the  $2j$ th node after one time “tick.”
- The actual time between ticks is  $\delta t$ .
- Interest is compounded continuously at a fixed rate  $r$  so that  $B_0$  dollars now is worth  $\exp(rn \delta t)B_0$  dollars after  $n$  time ticks. (Or,  $B_0$  dollars after  $n$  ticks is worth  $\exp(-rn\delta t)B_0$  dollars now.)

Then at node  $j$ :

- The value of the option is

$$f(j) = \exp(-r \delta t) \{q_j f(2j + 1) + (1 - q_j) f(2j)\}.$$

where

- The arbitrage determined  $q_j$  is

$$q_j = \frac{e^{r \delta t} s_j - s_{2j}}{s_{2j+1} - s_{2j}}. \quad (7.7)$$

- The number of shares of stock to be holding is

$$\phi_j = \frac{f(2j + 1) - f(2j)}{s_{2j+1} - s_{2j}} = \text{hedge ratio}.$$

- Denote the amount of capital to hold in the risk-free asset by  $\psi_j$ ; typically  $\psi_j$  is negative because money has been borrowed. Since the portfolio replicates the option, at node  $j$  the option's value, which is  $f(j)$ , must equal the portfolio's value which is  $s_j \phi_j + \psi_j$ . Therefore,

$$\psi_j = \{f(j) - \phi_j s_j\}. \quad (7.8)$$

( $\psi_j$  increases in value to  $e^{r \delta t} \{f(j) - \phi_j s_j\}$  after one more time tick).

Expectations for paths along the tree are computed using the  $q_j$ 's. The probability of any path is just the product of all the probabilities along the path.

**An example**

The tree for the example of Section 7.6 is shown in Figure 7.5. Because  $r = 0$  is assumed and because the stock moves either up or down the same amount (\$10), the  $q_j$  are all equal to  $1/2$ .<sup>5</sup>

The probability of each full path from node 1 to one of nodes 4, 5, 6, or 7 is  $1/4$ .

Given the values of the option at nodes 4, 5, 6, and 7, it is easy to compute the expectations of the option's value at other nodes. These expectations are shown in magenta in Figure 7.5.

The path probabilities are independent of the exercise price, since they depend only on the prices of stock at the nodes and on  $r$ . Therefore, it is easy to price options with other exercise prices.

**Exercise**

Assuming the same stock price process as in Figure 7.5, price the call option with an exercise price of \$70.

**Answer:** Given this exercise price, it is clear that the option is worth \$0, \$10, \$10, and \$30 dollars at nodes 4, 5, 6, and 7, respectively. Then we can use expectation to find that the option is worth \$5 and \$20 at nodes 2 and 3, respectively. Therefore, the option's value at node 1 is \$12.50; this is the price of the option.

**7.9 Martingales**

A *martingale* is a probability model for a fair game, that is, a game where the expected changes in one's fortune are always zero. More formally, a

---

<sup>5</sup>It follows from (7.7) that whenever  $r = 0$  and the up moves and down moves are of equal length, then  $q_j = 1/2$  for all  $j$ .

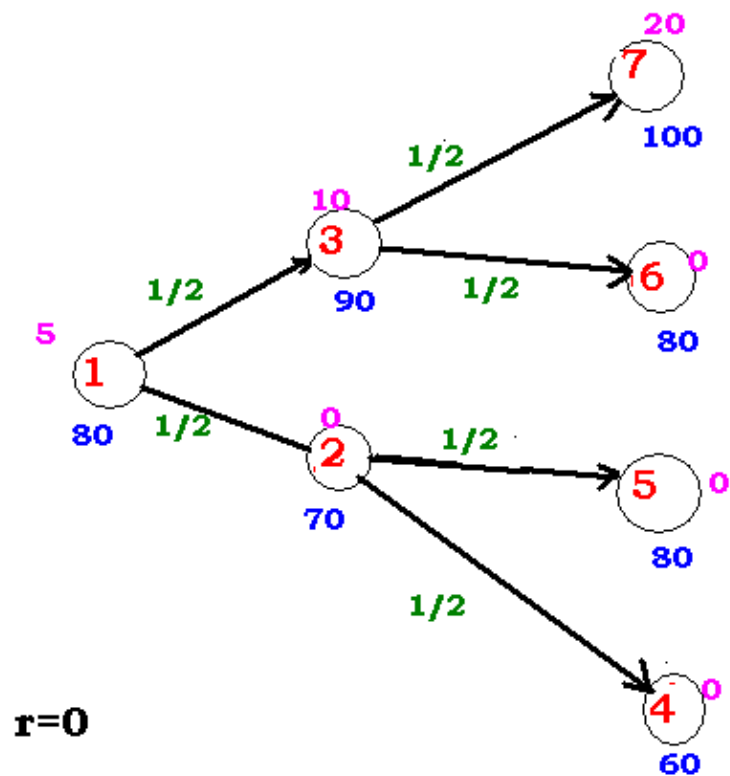


Figure 7.5: Two-step example with pricing by probabilities. Red is node number. Blue is value of the stock. Magenta is value of the option. Path probabilities are in dark green. The exercise price is \$80.



stochastic process  $Y_0, Y_1, Y_2, \dots$  is a martingale if

$$E(Y_{t+1}|Y_t) = Y_t$$

for all  $t$ .

Let  $P_t, t = 0, 1, \dots$  be the price of the stock at the end of the  $t$ th step in a binomial model. Then  $P_t^* := \exp(-rt \delta t)P_t$  is the discounted price process.

**Key fact:** Under the  $\{q_j\}$  probabilities, the discounted price process  $P_t^*$  is a martingale.

To see that  $P_t^*$  is a martingale, we calculate:

$$\begin{aligned} E(P_{t+1}|P_t = s_j) &= q_j s_{2j+1} + (1 - q_j) s_{2j} \\ &= s_{2j} + q_j (s_{2j+1} - s_{2j}) \\ &= s_{2j} + \{\exp(r \delta t) s_j - s_{2j}\} = \exp(r \delta t) s_j. \end{aligned}$$

This holds for all values of  $s_j$ . Therefore,

$$E(P_{t+1}|P_t) = \exp(r \delta t) P_t,$$

so that

$$E\{\exp(-r(t+1)\delta t)P_{t+1}|P_t\} = \exp(-rt\delta t)P_t,$$

or

$$E(P_{t+1}^*|P_t^*) = P_t^*.$$

This shows that  $P_t^*$  is a martingale.

Any set of path probabilities,  $\{p_j\}$ , is called a *measure* of the process. The measure  $\{q_j\}$  is called the *martingale measure* or the *risk-neutral measure*. We will also call  $\{q_j\}$  the risk-neutral path probabilities.

### 7.9.1 The risk-neutral world

If all investors were risk-neutral, that is, indifferent to risk, then there would be no risk premiums and all expected asset prices would rise at

the risk-free rate. Therefore, all discounted asset prices, with discounting at the risk-free rate, would be martingales.

We know that we do not live in such a risk-free world, but there is a general principle that expectations taken with respect to a risk-neutral model give correct, i.e., arbitrage-free, prices of options and other financial instruments.

### **Example**

In Section 7.3 it was argued that if a stock is selling at \$100/share and the risk-free interest rate is 6%. then the correct future delivery price of a share one year from now is \$106. We can now calculate this value using the risk-neutral measure—in the risk-neutral world, the expected stock price will increase to exactly \$106 one year from now.

## **7.10 Trees to random walks to Brownian motion**

### **7.10.1 Getting more realistic**

Binomial trees are useful because they illustrate several important concepts, in particular:

- arbitrage pricing
- self-financing trading strategies
- hedging
- computation of arbitrage prices by expectations with respect to an appropriate set of probabilities called the risk-neutral measure

However, binomial trees are not realistic, because stock prices are continuous, or at least approximately continuous. This lack of realism can be alleviated by increasing the number of step. In fact, one can increase the

number of steps without limit to derive the Black-Scholes model and formula. That is the goal of Section 7.11. The present section will get us closer to that goal.

### 7.10.2 A three-step binomial tree

Figure 7.6 is a three-step tree where at each step the stock price either goes up \$10 or down \$10. Assume that the risk-free rate is  $r = 0$ .

Now consider the price of the stock, call it  $P_t$  at time  $t$  where  $t = 0, 1, 2, 3$ . Using the risk-neutral path probabilities, which are each  $1/2$  in this example,  $P_t$  is a *stochastic process*, that is a process that evolves randomly in time. In fact, since  $P_{t+1}$  equals  $P_t \pm \$10$ , this process is a random walk. We have

$$P_t = P_0 + (\$10)\{2(W_1 + \cdots + W_t) - t\} \quad (7.9)$$

where  $W_1, \dots, W_3$  are independent and  $W_t$  equal 0 or 1, each with probability  $1/2$ . If  $W_t$  is 1, then  $2W_t - 1 = 1$  and the price jumps up \$10 on the  $t$ th step. If  $W_t$  is 0, then  $2W_t - 1 = -1$  the price jumps down \$10.

The random sum  $W_1 + \cdots + W_t$  is Binomial( $t, 1/2$ ) distributed and so has a mean of  $t/2$  and variance equal to  $t/4$ .

The value of the call option is

$$E\{(P_3 - E)_+\} \quad (7.10)$$

where  $x_+$  equals  $x$  if  $x \geq 0$  and equals 0 otherwise. The expectation in (7.10) is with respect to the risk-neutral probabilities. Since  $W_1 + W_2 + W_3$  is Binomial(3,  $1/2$ ), it equals 0, 1, 2, or 3 with probabilities  $1/8, 3/8, 3/8,$  and  $1/8$ , respectively. Therefore,

$$\begin{aligned} E\{(P_3 - E)_+\} &= \frac{1}{8}[\{P_0 - 30 - E + (20)(0)\}_+ + 3\{P_0 - 30 - E + (20)(1)\}_+ \\ &\quad + 3\{P_0 - 30 - E + (20)(2)\}_+ + \{P_0 - 30 - E + (20)(3)\}_+]. \end{aligned}$$

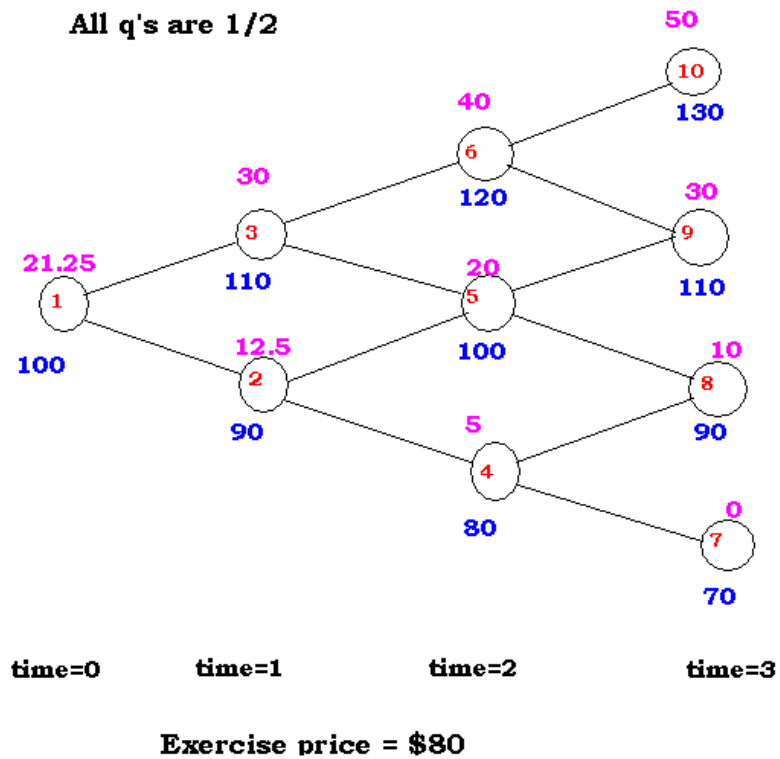


Figure 7.6: Three-step example of pricing a European call option by probabilities. Red is node number. Blue is value of the stock. Magenta is value of the option. Risk-neutral path probabilities are not shown, but they are all equal to  $1/2$ . The exercise price is \$80. The risk-free rate is  $r = 0$ .

**Examples**

If  $P_0 = 100$  and  $E = 80$ , then  $P_0 - 30 - E = -10$  and

$$\begin{aligned} E\{(P_3 - E)_+\} &= \frac{1}{8}\{(-10 + 0)_+ + 3(-10 + 20)_+ \\ &\quad + 3(-10 + 40)_+ + (-10 + 60)_+\} \\ &= \frac{1}{8}(0 + 30 + 90 + 50) = \frac{170}{8} = 21.25 \end{aligned}$$

as seen in Figure 7.6.

Similarly, if  $P_0 = 100$  and  $E = 100$ , then  $P_0 - 30 - E = -30$  and

$$\begin{aligned} E\{(P_3 - E)_+\} &= \frac{1}{8}\{(-30 + 0)_+ + 3(-30 + 20)_+ \\ &\quad + 3(-30 + 40)_+ + (-30 + 60)_+\} \\ &= \frac{1}{8}(0 + 0 + 30 + 30) = \frac{60}{8} = 7.5 \end{aligned}$$

**7.10.3 More time steps**

Let's consider a call option with maturity data equal to 1. Take the time interval  $[0, 1]$  and divide it into  $n$  steps, each of length  $1/n$ . Suppose that the stock price goes up or down  $\sigma/\sqrt{n}$  at each step. Then the price after  $m$  steps ( $0 \leq m \leq n$ ) when  $t = m/n$  is

$$P_{m/n} = P_0 + \frac{\sigma}{\sqrt{n}}\{2(W_1 + \cdots + W_m) - m\}. \quad (7.11)$$

Since,  $W_1 + \cdots + W_m$  is Binomial( $m, 1/2$ ) distributed, it follows that

$$E(P_t|P_0) = P_0. \quad (7.12)$$

and

$$\text{Var}(P_t|P_0) = \frac{4\sigma^2}{n} \frac{m}{4} = \frac{m}{n}\sigma^2 = t\sigma^2, \quad (7.13)$$

and, in particular,

$$\text{Var}(P_1|P_0) = \sigma^2. \quad (7.14)$$

Moreover, by the central limit theorem, as  $n \rightarrow \infty$ ,  $P_1$  converges to a  $N(P_0, \sigma^2)$  random variable.

Let  $E$  be the exercise price. Remember that the value of an option is the expectation with respect to the risk-neutral measure of the present value of the option at expiration. Therefore, in the limit, as the number of steps goes to  $\infty$ , the price of the option converges to

$$E\{(P_0 + \sigma Z - E)_+\} \quad (7.15)$$

where  $Z$  is  $N(0, 1)$  so that  $P_1 = P_0 + \sigma Z$  is  $N(P_0, \sigma^2)$ .

For a fixed value of  $n$ ,  $P_t$  is a discrete time stochastic process since  $t = 0, 1/n, 2/n, \dots, (n-1)/n, 1$ . In fact, as we saw before, for any finite value of  $n$ ,  $P_t$  is a random walk. However, in the limit as  $n \rightarrow \infty$ ,  $P_t$  becomes a continuous time stochastic process. This limit process is called *Brownian motion*. In other words, the continuous time limit of random walks is Brownian motion.

#### 7.10.4 Properties of Brownian motion

We have seen that Brownian motion is a continuous-time stochastic process that is the limit of discrete-time random walk processes. A Brownian motion process,  $B_t$ , starting at 0, i.e., with  $B_0 = 0$ , has the following mathematical properties:

1.  $E(B_t) = 0$  for all  $t$ .
2.  $\text{Var}(B_t) = t\sigma^2$  for all  $t$ . Here  $\sigma^2$  is the volatility of  $B_t$ .
3. Changes over non-overlapping increments are independent. More precisely, if  $t_1 < t_2 < t_3 < t_4$  then  $B_{t_2} - B_{t_1}$  and  $B_{t_4} - B_{t_3}$  are independent.
4.  $B_t$  is normally distributed for any  $t$ .

If  $B_0$  is not zero, then each of these properties holds for the process  $B_t - B_0$ , which is the change in  $B_t$  from times 0 to  $t$ . All of these properties but the last are shared to random walks with mean-zero steps.

## 7.11 Geometric Brownian motion

Random walks are not realistic models for stock prices, since a random walk can go negative. Therefore, (7.15) is close to but not quite the correct price of the option. To get the correct price we need to make our model more realistic. We saw in Chapter 3 that geometric random walks are much better than random walks as models for stock prices since geometric random walks are always non-negative.

We will now introduce a binomial tree model that is geometric random walk. We do this by making the steps proportional to the current stock price. Thus, if  $s$  is the stock price at the current node, then price at the next node is

$$s \exp(\mu/n \pm \sigma/\sqrt{n}) = (s_{\text{up}}, s_{\text{down}}).$$

Notice that the log of the stock price is a random walk since

$$(\log(s_{\text{up}}), \log(s_{\text{down}})) = \log(s) + \frac{\mu}{n} \pm \frac{\sigma}{\sqrt{n}}.$$

Therefore, the stock price process is a geometric random walk. There is a drift if  $\mu \neq 0$ , but we will see that the amount of drift is irrelevant. We could have set the drift equal to 0 but we didn't to show later that the drift does NOT affect the option's price.

The risk neutral probability of an up jump is

$$\begin{aligned} q &= \frac{s \exp(r/n) - s_{\text{down}}}{s_{\text{up}} - s_{\text{down}}} \\ &= \frac{\exp(r/n) - \exp(\mu/n - \sigma/\sqrt{n})}{\exp(\mu/n + \sigma/\sqrt{n}) - \exp(\mu/n - \sigma/\sqrt{n})} \\ &\approx \frac{1}{2} \left( 1 - \frac{\mu - r + \sigma^2/2}{\sigma\sqrt{n}} \right). \end{aligned}$$

Then

$$P_t = P_{m/n} = P_0 \exp \left\{ \mu t + \frac{\sigma}{\sqrt{n}} \sum_{i=1}^m (2W_i - 1) \right\}.$$

where as before  $W_i$  is either 0 or 1 (so  $2W_i - 1 = \pm 1$ ).

Using risk-neutral probabilities, we have

$$\begin{aligned} E \left\{ \frac{\sigma}{\sqrt{n}} \sum_{i=1}^m 2W_i - 1 \right\} &= \frac{\sigma m(2q-1)}{\sqrt{n}} \approx \frac{\sigma}{\sqrt{n}} m \left( \frac{r - \mu - \sigma^2/2}{\sigma\sqrt{n}} \right) \\ &= \left( r - \mu - \frac{\sigma^2}{2} \right) \frac{m}{n} = t(r - \mu - \sigma^2/2). \end{aligned}$$

and

$$\text{Var} \left\{ \frac{\sigma}{\sqrt{n}} \sum_{i=1}^m 2W_i - 1 \right\} = \frac{4\sigma^2 m q(1-q)}{n} \approx t\sigma^2,$$

since  $q \rightarrow 1/2$  as  $n \rightarrow \infty$ .

Therefore, in the risk-neutral world

$$P_t \approx P_0 \exp\{(r - \sigma^2/2)t + \sigma B_t\}, \quad (7.16)$$

where  $B_t$  is Brownian motion and  $0 \leq t \leq 1$ . Time could be easily extended beyond 1 by adding more steps. We will assume that this has been done.

Notice that (7.16) does NOT depend on  $\mu$ , only on  $\sigma$ . The reason is that in the risk-neutral world, the expectation of all assets increase at rate  $r$ . The rate of increase in the real world is  $\mu$  but this is irrelevant for risk-neutral calculations. Remember that risk-neutral expectations DO give the correct option price in the real world even if they do not correctly describe real world probability distributions.

If  $E$  is the exercise or strike price and  $T$  is the expiration date of a European call option, then the value of the option at maturity is

$$\left[ P_0 \exp\{(r - \sigma^2/2)T + \sigma B_T\} - E \right]_+. \quad (7.17)$$

Since  $B_T \sim N(0, T)$ , we can write  $B_T = \sqrt{T}Z$  where  $Z \sim N(0, 1)$ . The discounted value of (7.17) is

$$\left[ P_0 \exp \left\{ -\frac{\sigma^2 T}{2} + \sigma \sqrt{T} Z \right\} - \exp(-rT) E \right]_+. \quad (7.18)$$



We will again use the principle that the price of an option is the risk-neutral expectation of the option's discounted value at expiration. By this principle, the call's price at time  $t = 0$  is the expectation of (7.18). Therefore,

$$C = \int \left[ P_0 \exp \left\{ -\frac{\sigma^2 T}{2} + \sigma \sqrt{T} z \right\} - \exp(-rT) E \right]_+ \phi(z) dz, \quad (7.19)$$

where  $\phi$  is the  $N(0, 1)$  pdf (probability density function).

Computing this integral is not easy, but it can be done. The result is the famous Black-Scholes formula: Let  $S_0$  be the current stock price (we have switched notation from  $P_0$ ), let  $E$  be the exercise price, let  $r$  to be continuously compounded interest rate, let  $\sigma$  be the volatility, and let  $T$  be the expiration date of a call option. Then by evaluating the integral in (7.19) it can be shown that

$$C = \Phi(d_1) S_0 - \Phi(d_2) E \exp(-rT)$$

where  $\Phi$  is the standard normal CDF,

$$d_1 = \frac{\log(S_0/E) + (r + \sigma^2/2)T}{\sigma \sqrt{T}}, \quad \text{and} \quad d_2 = d_1 - \sigma \sqrt{T}.$$

### Example

Here's a numerical example. Suppose that  $S_0 = 100$ ,  $E = 90$ ,  $\sigma = .4$ ,  $r = .1$ , and  $T = .25$ . Then

$$d_1 = \frac{\log(100/90) + \{.1 + (.4)^2/2\}(.25)}{.4\sqrt{.25}} = 0.7518$$

and

$$d_2 = d_1 - .4\sqrt{.25} = .5518.$$

Then  $\Phi(d_1) = .7739$  and  $\Phi(d_2) = .7095$ . Also,  $\exp(-rT) = \exp\{(.1)(.25)\} = .9753$ . Therefore,

$$C = (100)(.7739) - (90)(.9753)(.7095) = 15.1.$$

## 7.12 Using the Black-Scholes formula

### 7.12.1 How does the option price depend on the inputs?

Figure 7.7 shows the variation in the price of a call option as the parameters change. The baseline values of the parameters are  $S_0 = 100$ ,  $E = 100 \exp(rT)$ ,  $T = .25$ ,  $r = .06$ , and  $\sigma = .1$ . The exercise price  $E$  and initial price  $S_0$  have been chosen so that if invested at the risk-free rate,  $S_0$  would increase to  $E$  at expiration time. In each of the subplots in Figure 7.7, one of the parameters is varied while the others are held at baseline.

One sees that the price of the call increases with  $\sigma$ . This makes sense since  $E = S_0 \exp(rT) = E(S_T)$  in this example ( $E(S_T)$  is the risk-neutral expectation of  $S_T$ ). The expected value of  $E(S_T)$  is at the money. Thus,  $S_T$  is equally likely to be in the money or out of the money. As  $\sigma$  increases, the likelihood that  $S_T$  is considerably larger than  $E$  also increases. As an extreme case, suppose that  $\sigma = 0$ . Then in the risk-neutral world  $S_T = \exp(rT)S_0 = E$  and the option at expiration is at the money so its value is 0.

The value at maturity is  $(S_T - E)_+$  so we expect that the price of the call will increase as  $S_0$  increases and decrease as  $E$  increases. This is exactly the behavior seen in Figure 7.7. Also, note that the price of the call increases as either  $r$  or  $T$  increases.

### 7.12.2 An example — GE

Table 7.12.2 gives the exercise price  $E$ , month of expiration, and the price of call options on GE on February 13, 2001. This information was taken from *The Wall Street Journal*, February 14. Traded options are generally American rather than European and that is true of the options in Table 7.12.2. However, under the Black-Scholes theory it can be proved that the price

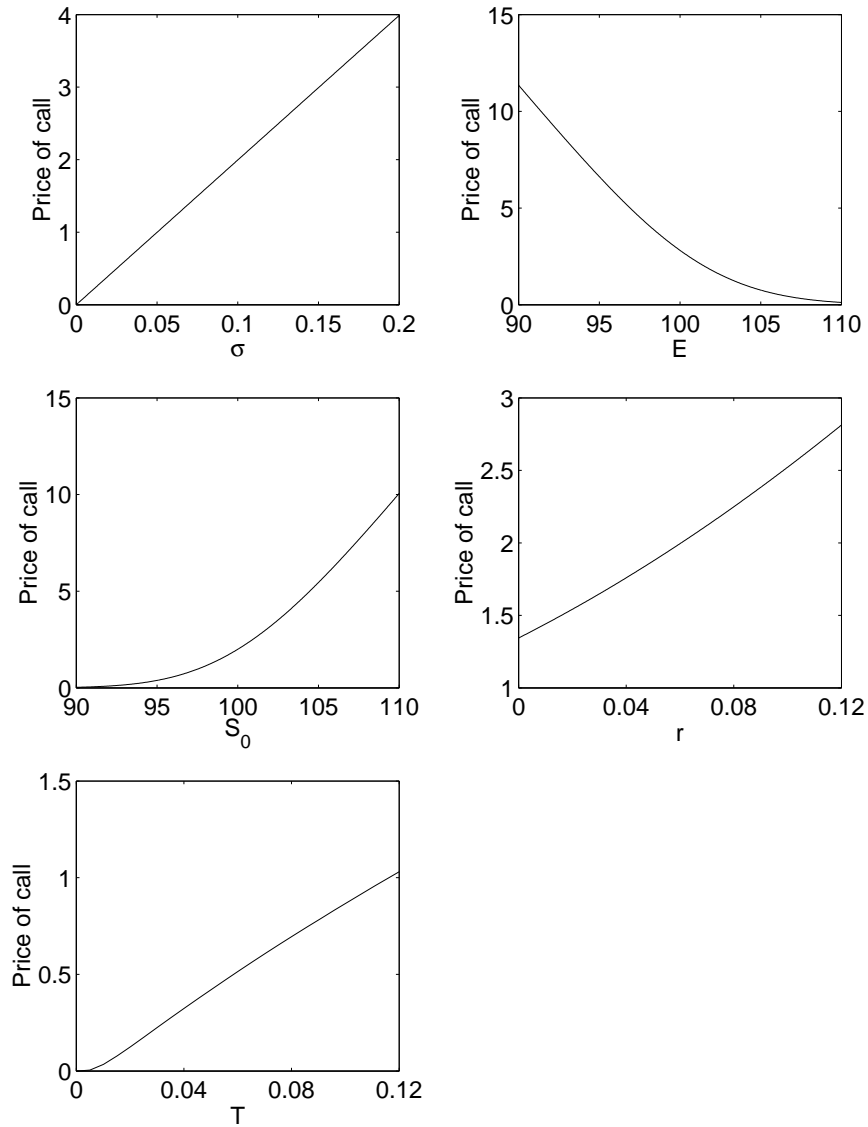


Figure 7.7: Price of a call as a function of volatility ( $\sigma$ ), exercise price ( $E$ ), initial price ( $S_0$ ), risk-free rate ( $r$ ), and expiration date ( $T$ ). Baseline values of the parameters are  $S_0 = 100$ ,  $E = 100 \exp(rT)$ ,  $T = .25$ ,  $r = .06$ , and  $\sigma = .1$ . In each subplot, all parameters except the one on the horizontal axis are fixed at baseline.

of an American call option is identical to the price of a European call option.<sup>6</sup> See Section 7.12.3 for discussion of this point. Since an American call has the same price as a European call, we can use the Black-Scholes formula for European call options to price the options in Table 7.12.2. We will compare the Black-Scholes prices with the actual market prices.

Only the month of maturity is listed in a newspaper. However, maturities (days until expiration) can be determined as follows. An option expires on 10:59pm Central Time of the Saturday after the third Friday in the month of expiration (Hull, 1995, page 180). February 16, 2001 was the third Friday of its month, so that on February 13, an option with a February expiration date had three trading days (and four calendar days) until expiration. Since there are returns on stocks only on trading days,  $T = 3$  for options expiring in February. Similarly, on February 13 an option expiring in March had  $T = 23$  trading days until expiration. Since there are 253 trading days/year, there are  $253/12 \approx 21$  trading days per month. For June, I used  $T = 23 + (21)(3)$  and for September I used  $T = 23 + (6)(21)$ .

GE closed at \$47.16 on February 13, so we use  $S_0 = 47.16$ .

On February 13, the 3-month T-bill rate was 4.91%. Thus, the daily rate of return on T-bills would be  $r = 0.0491/253 = .00019470$ , assuming that a T-bill only has a return on the 253 trading days per year; see Section 7.12.4.

I used two values of  $\sigma$ . The first, 0.0176, was based on daily GE return from December 1999 to December 2000. The second, 0.025, was chosen to give prices somewhat similar to the actual market prices.

### 7.12.3 Early exercise of calls is never optimal

It can be proved that early exercise of an American call option is never optimal. The reason is that at any time before the expiration date, the price of the option will be higher than the value of the option if exercised.

---

<sup>6</sup>However, American and European put options will in general have different prices.

$E$	Month of Expiration	$T$ (in days)	Actual Price	B&S calculated price		Implied Volatility
				$\sigma = .0176$	$\sigma = .025$	
35	Sep	149	14.90	13.40	14.03	.0320
40	Sep	149	10.80	9.22	10.37	.0275
42.50	Mar	23	5.30	5.03	5.38	.0235
45	Feb	3	2.40	2.22	2.32	.0290
45	Mar	23	3.40	3.00	3.57	.0228
50	Feb	3	0.10	0.016	0.09	.0258
50	Mar	23	0.90	0.64	1.23	.0209
50	Sep	149	4.70	3.42	5.12	.0232
55	Mar	23	0.20	0.06	0.28	.0223
55	Jun	86	1.30	0.92	2.00	.0204

Table 7.1: Actual prices and prices determined by the Black-Scholes formula for options on February 13, 2001.  $E$  is the exercise price.  $T$  is the maturity.

Therefore, it is always better to sell the option rather than to exercise it early.

To see empirical evidence of this principle, consider the first option in Table 7.12.2. The strike price is 35 and the closing price of GE was 47.16. Thus, if the option had been exercised at the closing of the market, the option holder would have gained  $\$(47.16 - 35) = \$12.16$ . However, the option was selling on the market for \$14.90 that day. Thus, one would gain  $\$(14.90 - 12.16) = \$2.74$  more by selling the option rather than exercising it.

Similarly, the other options in Table 7.12.2 are worth more if sold than if exercised. The second option is worth  $\$(47.16 - 40) = \$7.16$  if exercised but \$10.80 if sold. The third option is worth  $\$(47.16 - 42.5) = \$4.66$  if exercised but \$5.30 if exercised.

Since it is never optimal to exercise an American call option early, the abil-

ity to exercise an American call early is not worth anything. This is why American calls are equal in value to European calls with the same exercise price and expiration date.

#### **7.12.4 Are there returns on non-trading days?**

We have assumed that there are no returns on non-trading days. For T-bills, this assumption is justified by the way we calculated the daily interest rate. We took the daily rate to be the annual rate divided by 253 on trading days and 0 on non-trading days. If instead we took the daily rate to be the annual rate divided by 365 on every calendar day, then the interest on T-Bills over a year, or a quarter, would be the same.

A stock price is unchanged over a non-trading day. However, the efficient market theory says that stock prices change due to new information. Thus, we might expect that there is a return on a stock over a weekend or holiday but it is realized until the market reopens. If this were true, then returns from Friday to Monday would be more volatile than returns over a single trading day. Empirical evidence fails to find such an effect.

A reason why returns over weekends are not overly volatile might be that there is little business news over a weekend. However, this does not seem to be the explanation why there is not excess volatility over a weekend. In 1968, the NYSE was closed for a series of Wednesday. Of course, other businesses were open on these Wednesdays so there was the usual amount of business news during the Wednesdays when the NYSE was closed. For this reason, one would expect increased volatility for Tuesday to Thursday price changes on weeks with a Wednesday closing compared to, say, Tuesday to Wednesday price changes on weeks without a Wednesday market closing. However, no such effect has been detected (French and Roll, 1986).

Trading appears to generate volatility by itself. Traders react to each other. Stock prices react to both trading “noise” and to new information. Short term volatility might be mostly due to noise trading.

### 7.12.5 Implied volatility

Given the exercise price, current price, and maturity of an option and given the risk-free rate, there is some value of  $\sigma$  that makes the price determined by the Black-Scholes formula equal to the current market price. This value of  $\sigma$  is called the *implied volatility*. One might think of implied volatility as the amount of volatility the market believes to exist currently.

How does one determine the implied volatility? The Black-Scholes formula gives price as a function of  $\sigma$  with all other parameters held fixed. What we need is the inverse of this function, that is,  $\sigma$  as a function of the option price. Unfortunately, there is no formula for the inverse function. The function exists, of course, but there is no explicit formula for it. However, using interpolation one can invert the Black-Scholes formula to get  $\sigma$  as a function of price. Figure 7.8 shows how this could be done for the third option in Table 7.12.2. The implied volatility in Figure 7.8 is 0.0235 and was determined by MATLAB's interpolation function, `interp1.m`. The implied volatilities of the other options in Table 7.12.2 were determined in the same manner.

Notice that the implied volatilities are substantially higher than 0.0176, the average volatility over the previous year. However, there is evidence that volatility of GE was increasing at the end of last year; see the estimated volatility in Figure 3.3. In that figure, volatility is estimated from December 15, 1999 to December 15, 2000. Volatility is highest at the end of this period and shows some sign of continuing to increase. The estimated volatility on December 15, 2000 was 0.023, which is similar to the implied volatilities in Table 7.12.2. It would be worthwhile to re-estimate volatility with data from December 15, 2000 to February 13, 2001. It may be that the implied volatilities in Table 7.12.2 are similar to the observed volatility in early 2001.

The implied volatilities also vary somewhat among themselves. One reason for this variation is that the option prices and closing price of GE stock

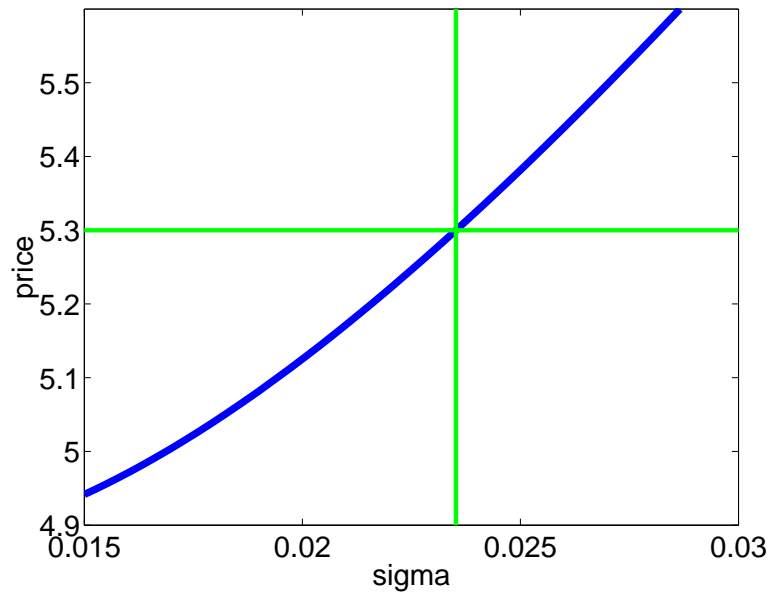


Figure 7.8: Calculating the volatility implied by the option with an exercise price of \$42.50 expiring in March 2001. The price was \$5.30 on February 13, 2001. The blue curve is the price given by the Black-Scholes formula as a function of  $\sigma$ . The horizontal line is drawn where price is \$5.30. This line intersects the curve at  $\sigma = .0242$ . This value of  $\sigma$  is the volatility implied by the option's price.



are not concurrent. Rather, each price is for the last trade of the day for that option or for the stock. This lack of concurrence introduces some error into pricing by the Black-Scholes formula and therefore into the implied volatilities. Another problem with these prices is that the Black-Scholes formula assumes that the stock pays no dividends, but GE does pay dividends.<sup>7</sup>

## 7.13 Puts

Recall that a put option gives one the right to sell a certain number of shares of a certain stock at the exercise price. The pricing of puts is similar to the pricing of calls, but as we will see in this section, there are some differences.

### 7.13.1 Pricing puts by binomial trees

Put options can be priced by binomial trees in the same way that call options are priced. Figure 7.9 shows a two-step binomial tree where the stock price starts at \$100 and increases or decreases by 20% at each step. Assume that the interest rate is 5% compounded continuously and that the strike price of the put is \$110.

In this example, European and American puts do NOT have the same price at all nodes. We will start with a European put and then see how an American put differs.

At each step,

$$q = \frac{\exp(.05) - .8}{1.2 - .8} = .6282.$$

The value of a put after two steps is  $(110 - S)_+$  where  $S$  is the price of the stock after two steps. Thus the put is worth \$46, \$14, and \$0 at nodes 4, 5,

---

<sup>7</sup>Modifications of the formula to accommodate dividend payments are possible, but we will not pursue that topic here.

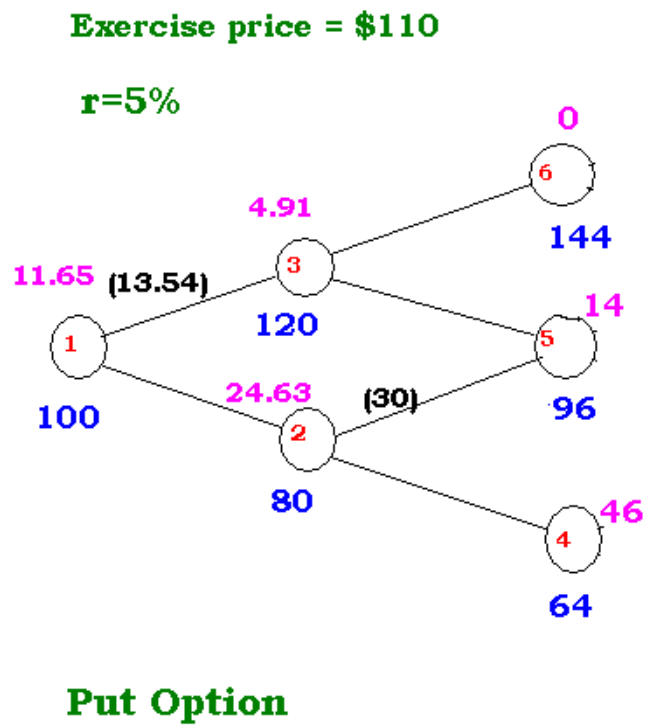


Figure 7.9: Pricing a put option. The stock price is in blue and the price of a European put option is in magenta. The price of an American put option is shown in black with parentheses when it differs from the price of a European put.

and 6 respectively. Therefore, the price of the option at node 3 is

$$e^{-.05}\{(q)(0) + (1 - q)(14)\} = e^{.05}\{(.6282)(0) + (.3718)(14)\} = 4.91.$$

The price of the option at node 2 is

$$e^{-.05}\{(q)(14) + (1 - q)(46)\} = 24.63.$$

Finally the price of the put at node 1 is

$$e^{-.05}\{(q)(4.91) + (1 - q)(24.63)\} = 11.65.$$

Now consider an American option. At nodes 4, 5, and 6 we have reached the expiration time so that the American option has the same value as the European option.

At node 3 the European option is worth \$4.91. At this node, should we exercise the American option early? Clearly not, since the strike price (\$110) is less than the stock price (\$120). Since early exercise is suboptimal at node 3, the American option is equivalent to the European option at this node and both options are worth \$4.91.

At node 2 the European option is worth \$24.63. The American option can be exercised to earn  $(\$110 - \$80) = \$30$ . Therefore, the American option should be exercised early since early exercise earns \$30 while holding the option is worth only \$25.89. Thus, at node 2 the European option is worth \$24.63 but the American option is worth \$30.

At node 1, the American option is worth

$$e^{-.05}\{(q)(4.91) + (1 - q)(30)\} = 13.65,$$

which is more than \$11.65, the value of the European option at node 1. The American option should NOT be exercised early at node 1 since that would earn only \$10. However, the American option is worth more than the European option at node 1 because the American can be exercised early at node 2 should the stock move down at node 1.

### 7.13.2 Why are puts different than calls?

We saw that in the Black-Scholes model where changes in price are proportional to the current price, it is never optimal to exercise an American call early. Puts are different. In the Black-Scholes model, early exercise of a call may be optimal. In the binomial model example just given, prices changes are proportional to current prices as in the Black-Scholes model, and in the binomial model early exercise of a put is again optimal at some nodes.

So why are puts different than calls? The basic idea is this. A put increases in value as the stock price *decreases*. As the stock price decreases, the size of further price changes also decreases. At some point we are in the range of diminishing returns. We expect further decreases in the stock price to be so small that the put will increase in value at less than the risk-free rate. Therefore, it is better to exercise the option and invest the profits in a risk-free asset.

With calls, everything is reversed. A call increases in value as the stock price increases. As the stock price increases, so does the size of future price changes. The expected returns on the call (expectations are with respect to the risk-neutral measure, of course) are greater than the risk-free rate of return.

### 7.13.3 Put-call parity

It is possible, of course, to derive the Black-Scholes formula for a European put option by the same reasoning used to price a call. However, this work can be avoided since there is a simple formula relating the price of a European put to that of a call:

$$P = C + e^{-rT}E - S_0, \quad (7.20)$$

where  $P$  and  $C$  are the prices of a put and of a call, both with expiration date  $T$  and exercise price  $E$ . Here, the stock price is  $S_0$  and  $r$  is the contin-

uously compounded risk-free rate. Thus, the price of a put is simply the price of the call plus  $e^{-rT}E - S_0$ .

Equation (7.20) is derived by a simple arbitrage argument. Consider two portfolios. The first portfolio holds one call and  $Ee^{-rT}$  dollars in the risk-free asset. Its payoff at time  $T$  is  $E$ , the value of the risk-free asset, plus the value of the call, which is  $(S_T - E)_+$ . Therefore, its payoff is  $E$  if  $S_T < E$  and  $S_T$  if  $S_T > E$ . In other words, the payoff is either  $E$  or  $S_T$ , whichever is larger.

The second portfolio holds a put and one share of stock. Its payoff at time  $T$  is  $S_T$  if  $S_T \geq E$  so that the put is not exercised. If  $S_T < E$ , then the put is exercised and the stock is sold for a payoff of  $E$ . Thus, the payoff is  $E$  or  $S_T$ , whichever is larger, which is the same payoff as the first portfolio.

Since the two portfolios have the same payoff for all values of  $S_T$ , their initial values at time 0 must be equal to avoid arbitrage. Thus,

$$C + e^{-rT}E = P + S_0,$$

which can be rearranged to yield equation (7.20).<sup>8</sup>

Relationship (7.20) holds only for European options. European calls have the same price as European calls so that the right hand side of (7.20) is the same for European and American options. American puts are worth more than European puts, so the left hand side of (7.20) is larger for American than for European puts. Thus, (7.20) becomes

$$P > C + e^{-rT}E - S_0, \tag{7.21}$$

for American options, and clearly (7.21) does not tell us the price of an American put.

---

<sup>8</sup>As usual in these notes, we are assuming that the stock pays no dividend, at least not during the lifetime of the two options. If there are dividends, then a simple adjustment of formula (7.20) is needed. The reason the adjustment is needed is that the two portfolios will no longer have exactly the same payoff. One can see that the first portfolio which holds the stock will receive a dividend and so receive a higher payoff than the second portfolio which will not receive the dividend.

## 7.14 The evolution of option prices

As time passes the price of an option changes with the changing stock price and the decreasing amount of time until the expiration date. We will assume that  $r$  and  $\sigma$  are constant, though in the real financial world these could change too. The Black-Scholes formula remains in effect and can be used to update the price of an option. Suppose that  $t = 0$  is when the option was written and  $t = T$  is the expiration date. Consider a time point  $t$  such that  $0 < t < T$ . Then the Black-Scholes formula can be used with  $S_0$  in the formula set equal to  $S_t$  and  $T$  in the formula set equal to  $T - t$ .

Figure 7.10 illustrates the evolution of option prices for two simulations of the geometric Brownian motion process of the stock price. Here  $T = 1$ ,  $\sigma = .1$ ,  $r = .06$ ,  $S_0 = 100$ , and  $E = 100$  for both the put and the call. In one case the call was in the money at expiration, while in the second case it was the put that was in the money.

Notice that around  $t = .18$  the stock price is around 110 in the red simulation but the put is still worth something, since there is still plenty of time for the price to go down. Around  $t = 1$  the stock price of the blue simulation is around 110 but the value of the put is essentially 0; now there is too little time for the put to go in the money (the risk-neutral probability is not 0, but almost 0).

## 7.15 Intrinsic value and time value

The intrinsic value of a call is  $(S_0 - E)_+$ , the payoff one would obtain for immediate exercise of the option (which would be possible only for an American option). The intrinsic value is always less than the price, so immediate exercise is never optimal. The difference between the intrinsic value and the price is called the time value of the option. Time value has two components. The first is a volatility component. The stock price could drop between now and the expiration date; by waiting until the last

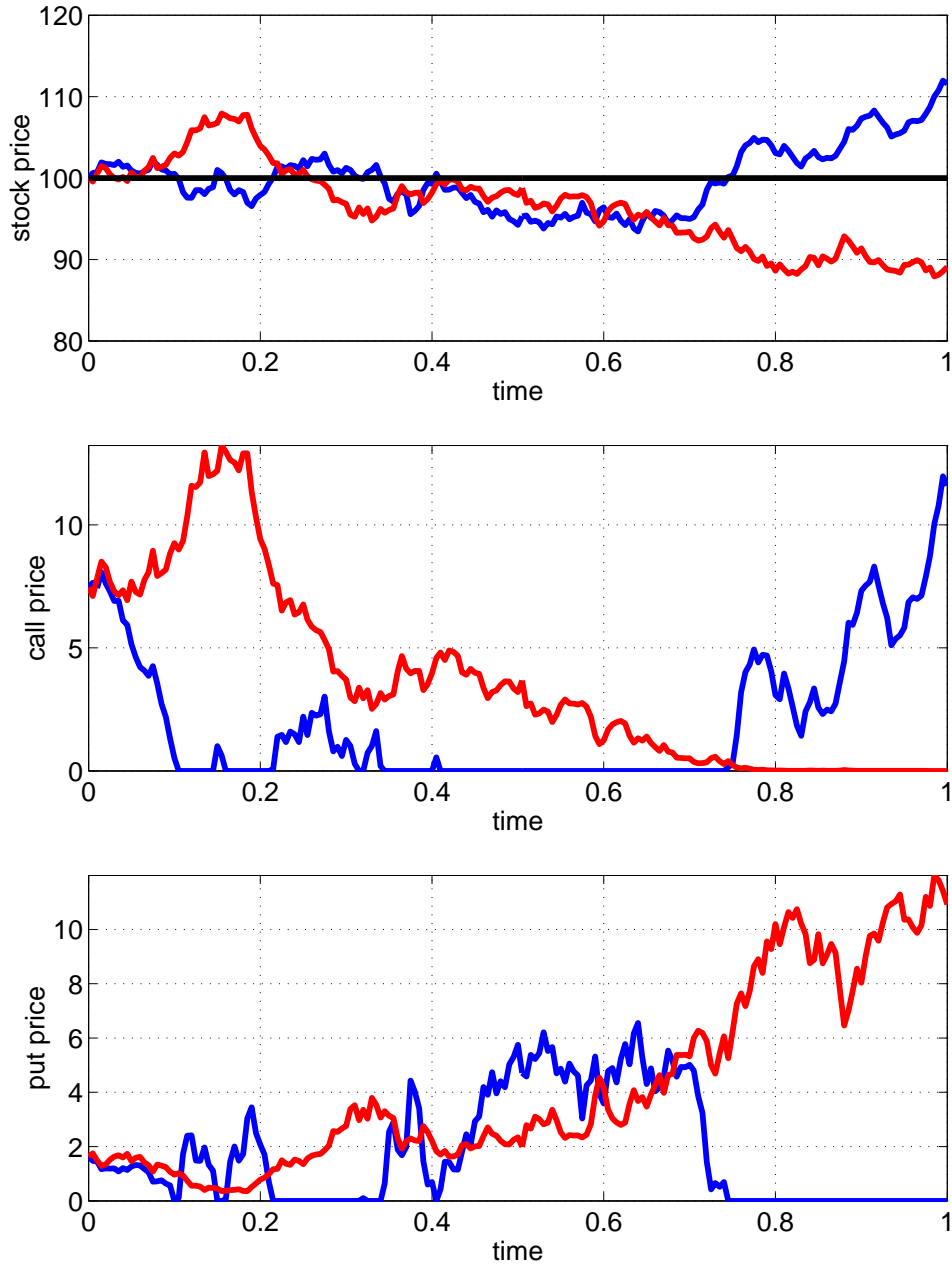


Figure 7.10: *Evolution of option prices. The stock price is a geometric Brownian motion. Two independent simulations of the stock price are shown and color coded. Here  $T = 1$ ,  $\sigma = .1$ ,  $r = .06$ ,  $S_0 = 100$ , and  $E = 100$  for both the put and the call. In the blue and red simulations the call, respectively, put are in the money at the expiration date.*

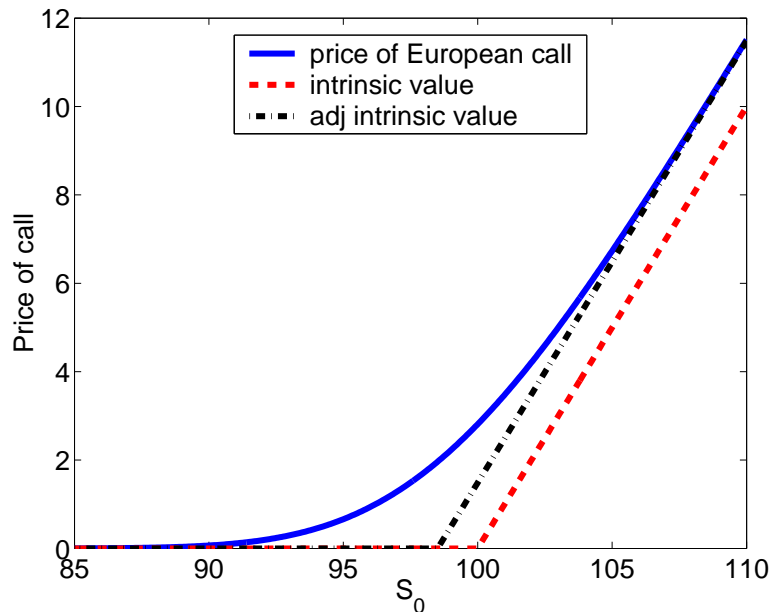


Figure 7.11: Price (for European or American option), intrinsic value, and adjusted intrinsic value of a call option. The intrinsic value is the payoff if one exercises early. Here  $E = 100$ ,  $T = .25$ ,  $r = 0.06$ , and  $\sigma = 0.1$ .

moment, one can avoid exercising the option when  $S_T < E$ . The second component is the time value of *money*. If you do exercise the option, it is best to wait until time  $T$  so that you delay payment of the exercise price.

The adjusted intrinsic value is  $(S_0 - e^{-rT}E)_+$ . The difference between the price and the adjusted intrinsic value is the volatility component of the time value of the option. As  $S_0 \rightarrow \infty$ , the price converges to the adjusted intrinsic value and the volatility component converges to 0. The reason this happens is that as  $S_0 \rightarrow \infty$  you become sure that the option will be in the money at the expiration date.

Figure 7.11 shows the price, intrinsic value, and adjusted intrinsic value of a call option when  $S_0 = 100$ ,  $E = 100$ ,  $T = .25$ ,  $r = 0.06$ , and  $\sigma = 0.1$



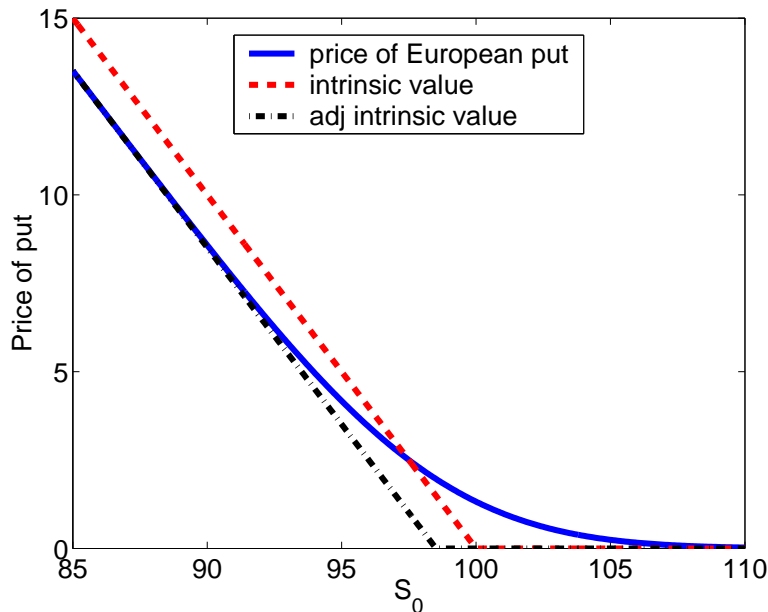


Figure 7.12: Price (for European option), intrinsic value, and adjusted intrinsic value of a put option. The intrinsic value is the payoff if one exercises early. The price of an American put would be either the price of the European put or the intrinsic value, whichever is larger. Here  $S_0 = 100$ ,  $E = 100$ ,  $T = .25$ ,  $r = 0.06$ , and  $\sigma = 0.1$ .

The intrinsic value of put is  $(E - S_0)_+$ , which again is the the payoff one would obtain for immediate exercise of the option, if that is possible (American option). The intrinsic value is sometimes greater than the price, in which case immediate exercise is optimal.

The adjusted intrinsic value is  $(e^{-rT}E - S_0)_+$ . As  $S_0 \rightarrow 0$ , the likelihood that the option will be in the money at the expiration date increase to 1 and the price converges to the adjusted intrinsic value.

Figure 7.12 shows the price, intrinsic value, and adjusted intrinsic value of a put option when  $S_0 = 100$ ,  $E = 100$ ,  $T = .25$ ,  $r = 0.06$ , and  $\sigma = 0.1$

## 7.16 Black, Scholes, and Merton

This section is based on chapter 11 of Bernstein's (1992) book *Capital Ideas*. Fischer Black graduated in 1959 from Harvard with a degree in physics. In 1964 he received a PhD in applied mathematics from Harvard where he studied operations research, computer design, and artificial intelligence. He never took a course in either finance or economics.

Finding his doctoral studies a bit too abstract, he went to work at Arthur D. Little where he became acquainted with the CAPM. He found this subject so fascinating that he moved into finance. At ADL, Black tried to apply the CAPM to the pricing of warrants, which are much like options. Bernstein (1992) quotes Black as recalling

I applied the Capital Asset Pricing Model to every moment in a warrant's life, for every possible stock price and warrant value ... . I stared at the differential equation for many, many months. I made hundreds of silly mistakes that led me down blind alleys. Nothing worked ...

[The calculations revealed that] the warrant value did not depend on the stock's expected return, or on any other asset's expected return. That fascinate me. ... Then Myron Scholes and I started working together.

Scholes received a bachelor's degree from McMaster's University in Ontario in 1962, earned a doctorate in finance from Chicago, and then took a teaching job at MIT. When Scholes meet Black he too was working intensely on warrant pricing by the CAPM. Realizing that they were working on the same problem, they began a collaboration that proved to be very fruitful.

Black and Scholes came to understand that the expected return on a stock or option had no effect of what the current price of the option should be.

With this insight and building on the CAPM, they arrived at the option equation and derived the formula for the option price.

In 1970, Scholes described his work with Black on options pricing to Robert C. Merton. Merton had studied engineering mathematics at Columbia and then Cal Tech. He developed an interest in economics and planned to study that subject in graduate school. His lack of formal training in economics put off many graduate schools, but MIT offered him a fellowship where he worked under the direction of Paul Samuelson.

Merton developed the “intertemporal capital asset pricing model” that converted the CAPM from a static model describing the market for a single discrete holding period to a model for finance in continuous time. Merton realized that Ito’s stochastic calculus was a goldmine for someone working on finance theory in continuous time. In the preface to his book, “Continuous-Time Finance,” Merton has written

The mathematics of the continuous-time model contains some of the most beautiful applications of probability and optimization theory. But, of course, not all that is beautiful in science need also be practical. And surely, not all that is practical in science is beautiful. Here we have both.

Merton developed a much more elegant derivation of the Black-Scholes formula, a derivation based on an arbitrage argument. Black has said “A key part of the options paper I wrote with Myron Scholes was the arbitrage argument for deriving the formula. Bob gave us that argument. It should probably be called the Black-Merton-Scholes paper.”

In 1997, Merton shared the Nobel Prize in Economics with Scholes. Sadly, Black had died at a young age and could not share the prize, since the Nobel Prize cannot be awarded posthumously. Merton has been called “the Isaac Newton of modern finance.”

## 7.17 Summary

- An option gives the holder the right but not the obligation to do something, for example, to purchase a certain amount of a certain stock at a fixed price within a certain time frame.
- A call option gives one the right to purchase (call in) a stock. A put gives one the right to sell (put away) a stock.
- European options can be exercised only at their expiration date. American options can be exercised on or before their expiration date.
- Arbitrage is making a guaranteed profit without investing capital.
- Arbitrage pricing means determining the unique price of a financial instrument that guarantees that the market is free of arbitrage opportunities.
- Options can be priced by arbitrage using binomial trees.
- The “measure” of a binomial tree model or other stochastic process model gives the set of path probabilities of that model.
- There exists a risk-neutral measure such that expected prices calculating with respect to this measure are equal to arbitrage determined prices.
- In a binomial tree model with price changes proportional to the current price, as the number of steps increases the limit process is a geometric Brownian motion and the price of the option in the limit is given by the Black-Scholes formula.
- To price an option by the Black-Scholes formula, one needs an estimate of the stock price’s volatility. This can be obtained from historical data. Conversely, the implied volatility of a stock is the volatility which makes the actual market price equal to the price given by the Black-Scholes formula.

- Within the Black-Scholes model, the early exercise of calls is never optimal but the early exercise of puts is sometimes optimal. Therefore, European and American calls have equal prices, but American puts are generally worth more than European puts.
- Put-call parity is the relationship

$$P = C + e^{-rT}E - S_0$$

between  $P$ , the price of a European put, and  $C$ , the price of a European call. It is assumed that both have exercise price  $E$  and expiration date  $T$ .  $S_0$  is the price of the stock.

## 7.18 References

- Baxter, M., and Rennie, A. (1998), *Financial Calculus: An Introduction to Derivative Pricing*, Cambridge University Press.
- French, K. R., and Roll, R., (1986), "Stock return variances; the arrival of information and the reaction of traders," *Journal of Financial Economics*, 17, 5–26.
- Hull, John C. (1995), *Introduction to Futures and Options Markets*, Prentice Hall, Englewood Cliffs, NJ.
- Merton, R.C. (1992), *Continuous-Time Finance, revised ed.*, Blackwell, Cambridge, Ma. and Oxford, UK.



# Chapter 8

## GARCH models: 4/24/01

### 8.1 Introduction

Despite the popularity of ARMA models, they have a significant limitation, namely, that they assume a constant volatility. In finance, where correct specification of volatility is of the utmost importance, this can be a severe limitation. In this chapter we look at time series models that have randomly varying volatility.

ARMA models are used to model the conditional expectation of the current observation,  $Y_t$ , of a process given the past observations. ARMA models do this by writing  $Y_t$  as a linear function of the past plus a white noise term. ARMA models also allow us to predict future observations given the past and present. The prediction of  $Y_{t+1}$  given  $Y_t, Y_{t-1} \dots$  is simply the conditional expectation of  $Y_{t+1}$  given  $Y_t, Y_{t-1} \dots$ .

However, ARMA models have rather boring conditional variances—the conditional variance of  $Y_t$  given the past is always a constant. What does this mean for, say, modeling stock returns? Suppose we have noticed that recent daily returns have been unusually volatile. We might suppose that tomorrow's return will also be more variable than usual. However, if we are modeling returns as an ARMA process, we cannot capture this type of behavior because the conditional variance is constant. So we need better

time series models if we want to model the nonconstant volatility often seen in financial time series.

In this chapter we will study models of nonconstant volatility. ARCH is an acronym meaning AutoRegressive Conditional Heteroscedasticity.<sup>1</sup> In ARCH models the conditional variance has a structure very similar to the structure of the conditional expectation in an AR model. We will first study the ARCH(1) model, which is similar to an AR(1) model. Then we will look at ARCH( $p$ ) models which are analogous to AR( $p$ ) models. Finally, we will look at GARCH (Generalized ARCH) models which model conditional variances much like the conditional expectation of an ARMA model.

## 8.2 Modeling conditional means and variances

Before looking at GARCH models, we will study some general principles on how one models non-constant variance.

The general form for the regression of  $Y_t$  on  $X_{1,t}, \dots, X_{p,t}$  is

$$Y_t = f(X_{1,t}, \dots, X_{p,t}) + \epsilon_t \quad (8.1)$$

where  $\epsilon_t$  has expectation equal to 0 and a constant variance  $\sigma^2$ . The function  $f$  is the conditional expectation of  $Y_t$  given  $X_{1,t}, \dots, X_{p,t}$ . To appreciate this fact, notice that if we take the conditional (given the  $X_{i,t}$  values) expectation of (8.1),  $f(X_{1,t}, \dots, X_{p,t})$  is treated as a constant and the conditional expectation of  $\epsilon_t$  is 0. Moreover, the conditional variance is simply the variance of  $\epsilon_t$ , that is,  $\sigma^2$ . Frequently,  $f$  is linear so that

$$f(X_{1,t}, \dots, X_{p,t}) = \beta_0 + \beta_1 X_{1,t} + \dots + \beta_p X_{p,t}.$$

**Principle:** To model the conditional mean of  $Y_t$  given  $X_{1,t}, \dots, X_{p,t}$ , write  $Y_t$  as the conditional mean *plus* white noise.

---

<sup>1</sup>Heteroscedasticity is a fancy way of saying non-constant variance. Homoscedasticity means constant variance. Alternate spellings are heteroskedasticity and homoskedasticity.



Equation (8.1) can be modified to allow a nonconstant conditional variance. Let  $\sigma^2(X_{1,t}, \dots, X_{p,t})$  be the conditional variance of  $Y_t$  given  $X_{1,t}, \dots, X_{p,t}$ . Then the model

$$Y_t = f(X_{1,t}, \dots, X_{p,t}) + \sigma(X_{1,t}, \dots, X_{p,t})\epsilon_t \quad (8.2)$$

gives the correct conditional mean and variance.

**Principle:** To allow a nonconstant conditional variance in the model, *multiply* the white noise term by the conditional standard deviation. This product is added to the conditional mean as in the previous principle.

The function  $\sigma(X_{1,t}, \dots, X_{p,t})$  must be non-negative since it is a standard deviation. If the function  $\sigma(\cdot)$  is linear, then its coefficients must be constrained to ensure non-negativity. Modeling non-constant conditional variances in regression is treated in depth in the book by Carroll and Ruppert (1988). Models for conditional variances are often called “variance function models.” The GARCH models of this chapter are a special class of variance function models.

### 8.3 ARCH(1) processes

Let  $\epsilon_1, \epsilon_2, \dots$  be Gaussian white noise with unit variance, that is, let this process be independent  $N(0,1)$ . Then

$$E(\epsilon_t | \epsilon_{t-1}, \dots) = 0,$$

and

$$\text{Var}(\epsilon_t | \epsilon_{t-1}, \dots) = 1. \quad (8.3)$$

Property (8.3) is called *conditional homoscedasticity*.

The process  $a_t$  is an ARCH(1) process if

$$a_t = \epsilon_t \sqrt{\alpha_0 + \alpha_1 a_{t-1}^2}. \quad (8.4)$$

We require that  $\alpha_0 \geq 0$  and  $\alpha_1 \geq 0$  because a standard deviation cannot be negative. It is also required that  $\alpha_1 < 1$  in order for  $a_t$  to be stationary with a finite variance. If  $\alpha_1 = 1$  then  $a_t$  is stationary but its variance is  $\infty$ ; see below. Equation (8.4) is somewhat like an AR(1) but in  $a_t^2$ , not  $a_t$ , and the ARCH(1) model induces an ACF in  $a_t^2$  that is like an AR(1)'s ACF.

Define

$$\sigma_t^2 = \text{Var}(a_t | a_{t-1}, \dots)$$

to be the conditional variance of  $a_t$  given past values. Since  $\epsilon_t$  is independent of  $a_{t-1}$  and  $\text{Var}(\epsilon_t) = 1$

$$E(a_t | a_{t-1}, \dots) = 0, \quad (8.5)$$

and

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2. \quad (8.6)$$

Understanding equation (8.6) is crucial to understanding how GARCH processes work. This equation shows that if  $a_{t-1}$  has an unusually large deviation from its expectation of 0, so that  $a_{t-1}^2$  is large, then the conditional variance of  $a_t$  is larger than usual. Therefore,  $a_t$  is also expected to have an unusually large deviation from its mean. This volatility will propagate since  $a_t$  having a large deviation makes  $\sigma_{t+1}^2$  large so that  $a_{t+1}$  will tend to be large. Similarly, if  $a_{t-1}$  is unusually small, then  $\sigma_t^2$  will be small, and  $a_t$  is expected to also be small, etc. Because of this behavior, unusual volatility in  $a_t$  tends to persist, though not forever. The conditional variance tends to revert to the unconditional variance provided that  $\alpha_1 < 1$  so that the process is stationary with a finite variance.

The unconditional, i.e., marginal, variance of  $a_t$  denoted by  $\gamma_a(0)$  is gotten by taking expectations in (8.5) which give us

$$\gamma_a(0) = \alpha_0 + \alpha_1 \gamma_a(0).$$

This equation has a positive solution if  $\alpha_1 < 1$ :

$$\gamma_a(0) = \alpha_0 / (1 - \alpha_1).$$

If  $\alpha_1 \geq 1$  then  $\gamma_a(0)$  is infinite. It turns out that  $a_t$  is stationary nonetheless. The integrated GARCH model (I-GARCH) has  $\alpha_1 = 1$  and is discussed in Section 8.10.

Straightforward calculations using (8.6) show that the ACF of  $a_t$  is

$$\rho_a(h) = 0 \quad \text{if } h \neq 0.$$

In fact, any process such that the conditional expectation of the present observation given the past is constant is an uncorrelated process. In introductory statistics courses, it is often mentioned that independence implies zero correlation but not vice versa. A process, such as the GARCH processes, where the conditional mean is constant but the conditional variance is non-constant is a good example of a process that is uncorrelated but not independent. The dependence of the conditional variance on the past is the reason the process is not independent. The independence of the conditional mean on the past is the reason that the process is uncorrelated.

Although  $a_t$  is uncorrelated just like the white noise process  $\epsilon_t$ , the process  $a_t^2$  has a more interesting ACF: if  $\alpha_1 < 1$  then

$$\rho_{a^2}(h) = \alpha_1^{|h|}, \quad \forall h.$$

If  $\alpha_1 \geq 1$ , then  $a_t^2$  is nonstationary, so of course it does not have an ACF.

### 8.3.1 Example

A simulated ARCH(1) process is shown in Figure 8.1. The top-left panel shows the independent white noise process,  $\epsilon_t$ . The top right panel shows  $\sigma_t = \sqrt{1 + .95a_{t-1}^2}$ , the conditional standard deviation process. The bottom left panel shows  $a_t = \sigma_t\epsilon_t$ , the ARCH(1) process. As discussed in the next section, an ARCH(1) process can be used as the noise term of an AR(1) process. This is shown in the bottom right panel. The AR(1) parameters are  $\mu = .1$  and  $\phi = .8$ .

The variance of  $a_t$  is  $\gamma_a(0) = 1/(1 - .95) = 20$  so the standard deviation is  $\sqrt{20} = 4.47$ .

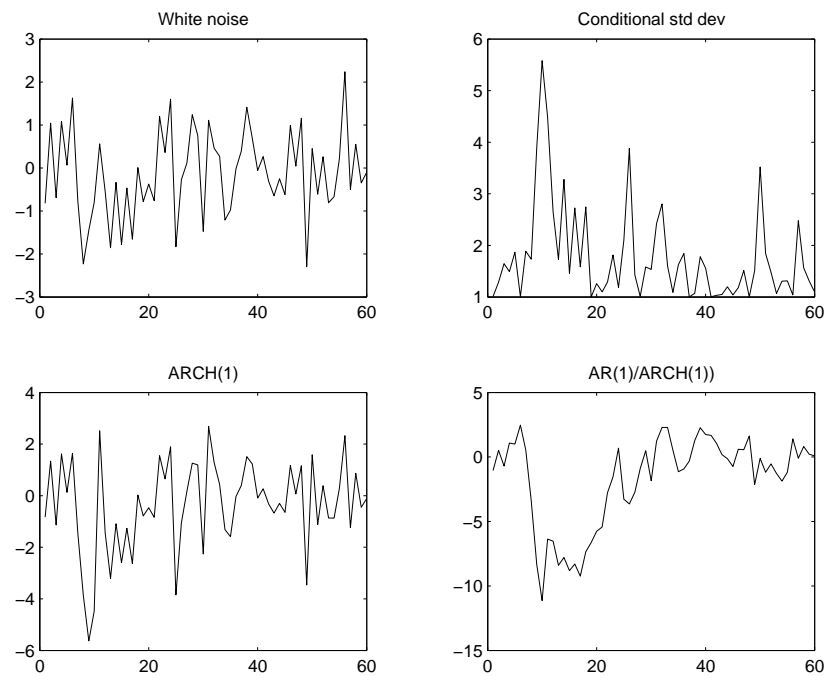


Figure 8.1: Simulation of 60 observations from an ARCH(1) process and an AR(1)/ARCH(1) process. The parameters are  $\alpha_0 = 1$ ,  $\alpha_1 = .95$ ,  $\mu = .1$ , and  $\phi = .8$ .

The processes were started out all equal to 0 and simulated for 70 observations. The first 10 observations were treated as a burn-in period where the process was converging to its stationary distribution. In the figure, only the last 60 observations are plotted.

The white noise process in the top left panel is normally distributed and has a standard deviation of 1, so it will be less than 2 in absolute value about 95% of the time. Notice that just before  $t = 10$ , the process is a little less than  $-2$  which is a somewhat large deviation from the mean of 0. This deviation causes the conditional standard deviation ( $\sigma_t$ ) shown in the top right panel to increase and this increase persists for about 10 observations though it slowly decays. The result is that the ARCH(1) process exhibits more volatility than usual when  $t$  is between 10 and 15.

Figure 8.2 shows a simulation of 600 observations from the same processes as in Figure 8.1. A normal probability plot of  $a_t$  is also included. Notice that this ARCH(1) exhibits extreme non-normality. This is typical of ARCH processes. Conditionally they are normal with a nonconstant variance, but their marginal distribution is non-normal with a constant variance.

## 8.4 The AR(1)/ARCH(1) model

As we have seen, an AR(1) has a nonconstant conditional mean but a constant conditional variance, while an ARCH(1) process is just the opposite. If we think that both the conditional mean and variance of a process will depend on the past then we need the features of both the AR and ARCH models. Thus, we will combine the two models. In this section we start simple and combine an AR(1) model with an ARCH(1) model.

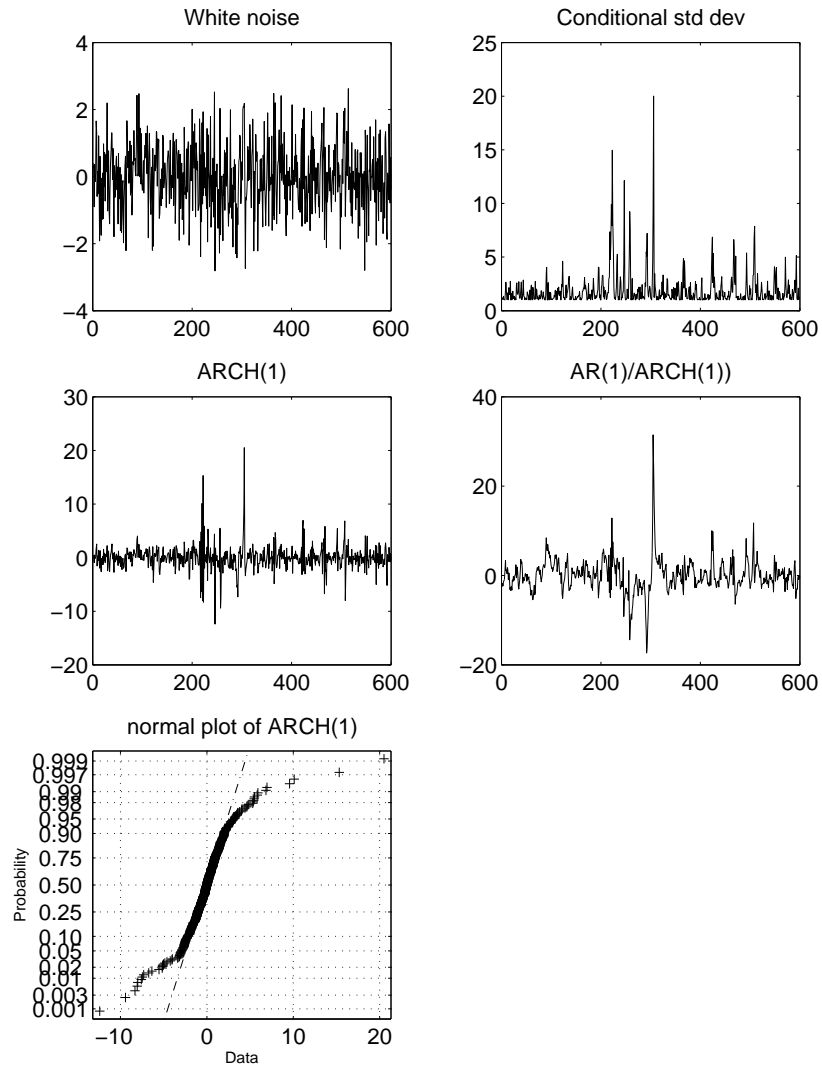


Figure 8.2: Simulation of 600 observations from an ARCH(1) process and an AR(1)/ARCH(1) process. The parameters are  $\alpha_0 = 1$ ,  $\alpha_1 = .95$ ,  $\mu = .1$ , and  $\phi = .8$ .

Let  $a_t$  be an ARCH(1) process and suppose that

$$u_t - \mu = \phi(u_{t-1} - \mu) + a_t.$$

$u_t$  looks like an AR(1) process, except that the noise term is not independent white noise but rather an ARCH(1) process.

Although  $a_t$  is not independent white noise, we saw in the last section that it is an uncorrelated process;  $a_t$  has the same ACF as independent white noise. Therefore,  $u_t$  has the same ACF as an AR(1) process:

$$\rho_u(h) = \phi^{|h|} \quad \forall \quad h.$$

Moreover,  $a_t^2$  has the ARCH(1) ACF:

$$\rho_{a^2}(h) = \alpha_1^{|h|} \quad \forall \quad h.$$

We need to assume that both  $|\phi| < 1$  and  $\alpha_1 < 1$  in order for  $u$  to be stationary with a finite variance. Of course,  $\alpha_0 \geq 0$  and  $\alpha_1 \geq 0$  and also assumed.

The process  $u_t$  is such that its conditional mean and variance, given the past, are both nonconstant so a wide variety of real time series can be modeled.

### Example

A simulation of an AR(1)/ARCH(1) process is shown in the bottom right panel of Figure 8.1. Notice that when the ARCH(1) noise term in the bottom left panel is more volatile, then the AR(1)/ARCH(1) process moves more rapidly.

## 8.5 ARCH( $q$ ) models

As before, let  $\epsilon_t$  be Gaussian white noise with unit variance. Then  $a_t$  is an ARCH( $q$ ) process if

$$a_t = \sigma_t \epsilon_t$$

where

$$\sigma_t = \sqrt{\alpha_0 + \sum_{i=1}^q \alpha_i a_{t-i}^2}$$

is the conditional standard deviation of  $a_t$  given the past values of this process. Like an ARCH(1) process, an ARCH( $q$ ) process is uncorrelated and has a constant mean (both conditional and unconditional) and a constant unconditional variance, but its conditional variance is nonconstant. In fact, the ACF of  $a_t^2$  is the same as the ACF on an AR( $p$ ) process.

## 8.6 GARCH( $p, q$ ) models

The GARCH( $p, q$ ) model is

$$a_t = \epsilon_t \sigma_t$$

where

$$\sigma_t = \sqrt{\alpha_0 + \sum_{i=1}^q \alpha_i a_{t-i}^2 + \sum_{i=1}^p \beta_i \sigma_{t-i}^2}$$

The process  $a_t$  is uncorrelated with a stationary mean and variance and  $a_t^2$  has an ACF like an ARMA process.

A very general time series model lets  $a_t$  be GARCH( $p_G, q_G$ ) and uses  $a_t$  as the noise term in an ARIMA( $p_A, d, q_A$ ) model.<sup>2</sup>

GARCH models include ARCH models as a special case, and we will use the term “GARCH” to refer to both ARCH and GARCH models.

Figure 8.3 is a simulation from a GARCH(1,1) process and from a AR(1)/GARCH(1,1) process. The GARCH parameters are  $\alpha_0 = 1$ ,  $\alpha_1 = .08$ ,  $\beta_1 = .9$ . The large value of  $\beta_1$  give the conditional standard deviation process a long-term memory. Notice that the conditional standard deviation is less “bursty” than for an ARCH(1) process.

---

<sup>2</sup>We use subscripts on  $p$  and  $q$  to distinguish between the GARCH (G) and ARIMA (A) parameters.



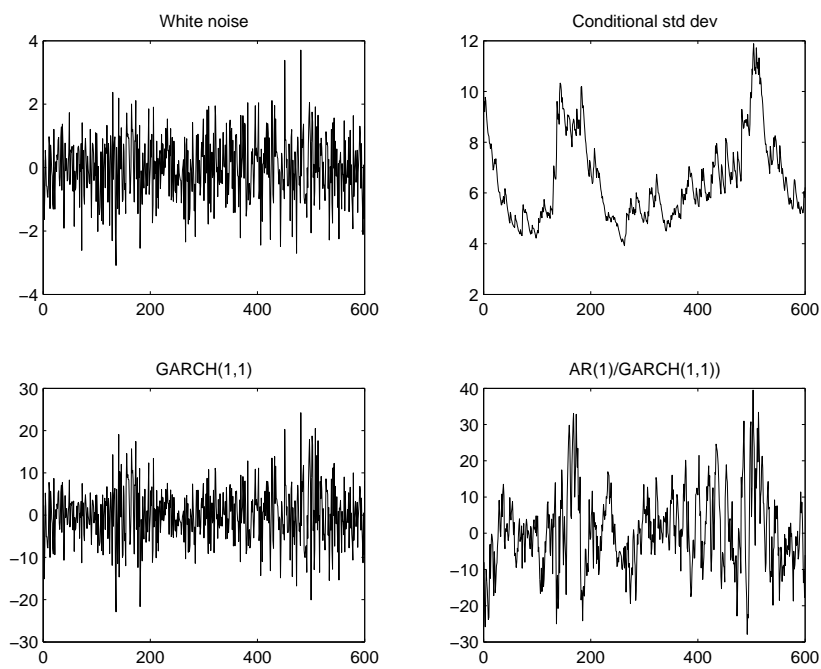


Figure 8.3: Simulation GARCH(1,1) and AR(1)/GARCH(1,1) processes. The parameters are  $\alpha_0 = 1$ ,  $\alpha_1 = .08$ ,  $\beta_1 = .9$ , and  $\phi = .8$ .

## 8.7 Heavy-tailed distributions

Researchers have long noticed that stock returns have “heavy-tailed” or “outlier-prone” probability distributions. This means that they have more extreme outliers than expected from a normal distribution. The reason for the outliers may be that the conditional variance is not constant. In fact, GARCH processes exhibit heavy-tails. Therefore, when we use GARCH models in finance we can model both the conditional heteroscedasticity and the heavy-tailed distributions of financial market data.

To understand how a non-constant variance induces outliers, we look at a simple case. Consider a distribution which is 90%  $N(0, 1)$  and 10%  $N(0, 25)$ . This is an example of a “normal mixture” distribution. The variance of this distribution is  $(.9)(1) + (.1)(25) = 3.4$  so its standard deviation is 1.844. This distribution is MUCH different than a  $N(0, 3.4)$  distribution, even though both distributions have the same mean (0) and variance (3.4). To appreciate this, look at Figure 8.4.

You can see in the top left panel that the two densities look quite different. The normal density looks much more dispersed than the normal mixture, but we know that they actually have the same variances. What’s happening? Look at the detail of the right tails in the top right panel. The normal mixture density is much higher than the normal density when  $x$  (the variable on the horizontal axis) is greater than 6. This is the “outlier” region (along with  $x < -6$ ). The normal mixture has more outliers and they come from the 10% of the population with a variance of 25. Outliers have a powerful effect on the variance and this small fraction of outliers inflates the variance from 1.0 (the variance of 90% of the population) to 3.4.

Let’s see how much more probability the normal mixture distribution has in the outlier range  $|x| > 6$  compared to the normal distribution.<sup>3</sup> For a

---

<sup>3</sup>There is nothing special about “6” to define the boundary of the outlier range. I just needed a specific number to make numerical comparisons. Clearly,  $|x| > 7$  or  $|x| > 8$ , say, would have been just as appropriate as outlier ranges.

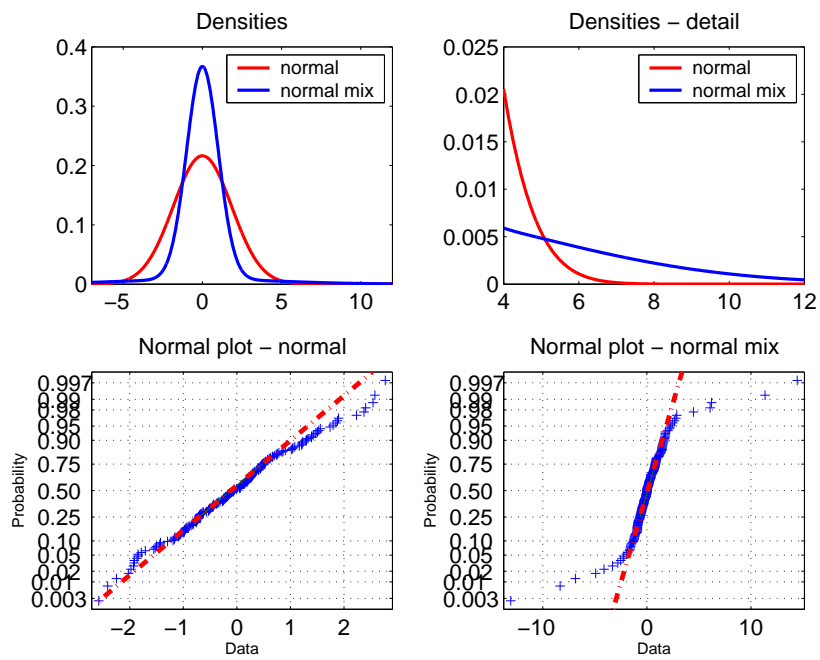


Figure 8.4: Comparison on normal and heavy-tailed distributions.

$N(0, \sigma^2)$  random variable  $X$ ,

$$P\{|X| > x\} = 2(1 - \Phi(x/\sigma)).$$

Therefore, for the normal distribution with variance 3.4,

$$P\{|X| > 6\} = 2(1 - \Phi(6/\sqrt{3.4})) = .0011.$$

For the normal mixture population which has variance 1 with probability .9 and variance 25 with probability .1 we have that

$$P\{|X| > 6\} = 2\{.9(1 - \Phi(6)) + .1(1 - \Phi(6/5))\} = (.9)(0) + (.1)(.23) = .023.$$

Since  $.023/.001 \approx 21$ , the normal mixture distribution is 21 times more likely to be in this outlier range than the normal distribution.

Normal probability plots of samples of size 200 from the normal and the normal mixture distributions are shown in the bottom panels. Notice how the outliers in the normal mixture sample give the data a nonlinear, almost S-shaped, pattern. The deviation of the normal sample from linearity is small and is due entirely to randomness.

In this example, the variance is conditional upon which component of the mixture an observation comes from. The conditional variance is 1 with probability .9 and 25 with probability .1. Because the conditional variance is discrete, in fact, with only two possible values, the example was easy to analyze. The marginal distribution of a GARCH process is also a normal mixture, but with a continuous distribution of components corresponding to the continuous distribution of the conditional variance. Although GARCH processes are more complex than the simple model in this section, the same theme applies — conditional heteroscedasticity induces heavy-tailed marginal distributions even though the conditional distributions are light-tailed normal distributions.

## 8.8 Comparison of ARMA and GARCH processes

Table 8.8 compares Gaussian white noise, ARMA, GARCH, and ARMA/GARCH processes according to various properties: conditional means,

conditional variances, conditional distributions, marginal means, marginal variances, and marginal distributions.

Property	Gaussian WN	ARMA	GARCH	ARMA/GARCH
Cond. mean	constant	non-const	0	non-const
Cond. var	constant	constant	non-const	non-const
Cond. dist'n	normal	normal	normal	normal
Marg. mean & var.	constant	constant	constant	constant
Marg. dist'n	normal	normal	heavy-tailed	heavy-tailed

All of the processes are stationary so that their marginal means and variances are constant. Gaussian white noise is the “baseline” process. Because it is an independent process the conditional distributions are the same as the marginal distribution. Thus, its conditional means and variances are constant and both its conditional and marginal distributions are normal. Gaussian white noise is the “driver” or “source of randomness” behind all the other processes. Therefore, they all have normal conditional distributions just like Gaussian white noise.

## 8.9 Fitting GARCH models

A time series was simulated using the same program that generated the data in Figure 8.1, the only difference being that 300 observations were generated rather than only 60 as in the figure. The data were saved as “garch02.dat” and analyzed with SAS using the following program.

### Listing of the SAS program for the simulated data

```
options linesize = 65 ;
data arch ;
infile 'c:\courses\or473\sas\garch02.dat' ;
input y ;
```

```

run ;
title 'Simulated ARCH(1)/AR(1) data' ;
proc autoreg ;
model y =/nlag = 1  archtest garch=(q=1);
run ;

```

This program uses the “autoreg” command that fits AR models. Since  $nlag = 1$ , an AR(1) model is being fit. However, the noise is not modeled as independent white noise. Rather an ARCH(1) model is used because of the specification “garch=(q=1)” in the “model” statement below the “autoreg” command. More complex GARCH models can be fit using, for example, “garch=(p=2,q=1).” The specification “archtest” requests tests of ARCH effects, that is, tests the null hypothesis of conditional homoscedasticity versus the alternative of conditional heteroscedasticity.

The output from this SAS program are listed below. The tests of conditional homoscedasticity all reject with p-values of .0001 or smaller. The estimates are  $\hat{\phi} = -.8226$ , which is  $+.8226$  in our notation. This is close to the true value of 0.8.

The estimates of the ARCH parameters are  $\hat{\alpha}_0 = 1.12$  and  $\hat{\alpha}_1 = .70$ . The true values are  $\alpha_0 = 1$  and  $\alpha_1 = .95$ . The standard errors of the ARCH parameters are rather large. This is a general phenomenon; time series usually have less information about variance parameters than about the parameters specifying the conditional expectation. An approximate 95% confidence interval for  $\alpha_1$  is

$$.70 \pm (2)(0.117) = (.446, .934),$$

which does not quite include the true parameter, 0.95. This could have just been bad luck, though it may indicate that  $\hat{\alpha}_1$  is downward biased. The confidence interval is based on the assumption of unbiasedness and is not valid if there is a sizeable bias.



Simulated ARCH(1)/AR(1) data 2  
 13:01 Wednesday, April 4, 2001

The AUTOREG Procedure

Preliminary MSE 3.5076

Estimates of Autoregressive Parameters

Lag	Standard Coefficient	Error	t Value
1	-0.780567	0.036209	-21.56

Algorithm converged.

GARCH Estimates

SSE	1056.42037	Observations	300
MSE	3.52140	Uncond Var	3.72785257
Log Likelihood	-549.43844	Total R-Square	0.6077
SBC	1121.69201	AIC	1106.87688
Normality Test	1.5134	Pr > ChiSq	0.4692

Variable	Standard DF	Approx Estimate	Error	t Value	Pr >  t
Intercept	1	0.4810	0.3910	1.23	0.2187
AR1	1	-0.8226	0.0266	-30.92	<.0001
ARCH0	1	1.1241	0.1729	6.50	<.0001
ARCH1	1	0.6985	0.1167	5.98	<.0001

### 8.9.1 Example: S&P 500 returns

This example is Example 10.5 in Pindyck and Rubinfeld (1998). The data are monthly from 1960 to 1996. The variables are the S&P 500 index (FSP-COM), the return on the S&P 500 (RETURNSP), the dividend yield on the S&P 500 index (FSDXP), the 3-month T-bill rate (R3), the change in the 3-month T-bill rate (DR3), the wholesale price index (PW), and the rate of wholesale price inflation (GPW). In this analysis, only RETURNSP, DR3, and GPW are used.



It is expected that variation in stock returns are in part caused by changes in interest rates and changes in the rate of inflation. Therefore, a regression model where RETURN<sub>SP</sub> is regressed on DR3 and GPW is used. Regression models that regress returns on macroeconomic variables in this way are sometimes called “factor models” — see Bodie, Kane, and Marcus (1999). Figure 8.5 shows the residuals from this regression. The residuals represent the part of the S&P 500 returns that cannot be explained by changes in interest rates and the inflation rate. In the figure, there is some sign of nonconstant volatility. Also, there is no reason to assume that the residuals are uncorrelated as is assumed in a standard regression model. At the very least, this assumption should be checked. If the data contradict the assumption, then a model with correlated errors should be used.

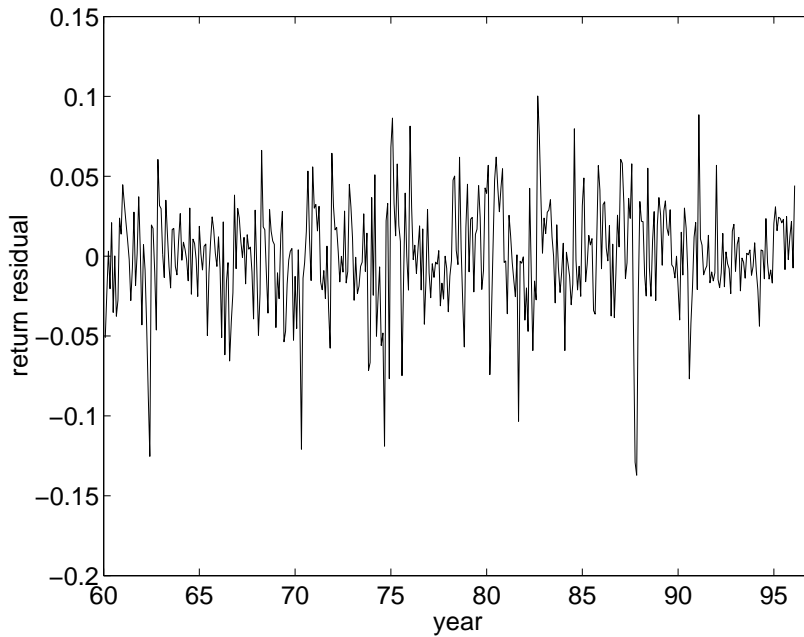


Figure 8.5: *Residuals when the S&P 500 returns are regressed against the change in the 3-month T-bill rates and the rate of inflation.*

An analysis more appropriate for this data set is to use a regression model to specify the conditional expectation of RETURNSP given DR3 and GPW, but not to assume a “standard” regression model with errors (which the residuals estimate) that are independent white noise. Rather we will assume the model

$$\text{RETURNSP} = \gamma_0 + \gamma_1 \text{DR3} + \gamma_2 \text{GPW} + u_t \quad (8.7)$$

where  $u_t$  is an AR(1)/GARCH(1,1) process.<sup>4</sup> Therefore,

$$u_t = \phi_1 u_{t-1} + a_t,$$

where  $a_t$  is a GARCH(1,1) process:

$$a_t = \epsilon_t \sigma_t$$

where

$$\sigma_t = \sqrt{\alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2}.$$

Below is a listing of the SAS program used to fit this model. The regression model with AR(1)/GARCH(1,1) errors is specified by the command:

```
proc autoreg ;
model returnsp = DR3 gpw/nlag = 1 archtest garch=(p=1,q=1);
```

In this command,

- the statement “returnsp = DR3 gpw ” specifies the regression model, that is, that “returnsp” is the dependent variable and “DR3” and “gpw” are the independent variables.
- “nlag = 1” specifies the AR(1) structure.
- “garch=(p=1,q=1)” specifies the GARCH(1,1) structure.
- “archtest” specifies that tests of conditional heteroscedasticity be performed

---

<sup>4</sup>We denote the regression coefficients by gamma rather than beta, as is standard, because beta is used for parameters in the GARCH model for  $a_t$ .

**Listing of the SAS program**

```

options linesize = 65 ;
data arch ;
infile 'c:\courses\or473\data\pindyck105.dat' ;
input month year RETURNSP FSPCOM FSDXP R3 PW GPW;
DR3 = dif(R3) ;
run ;
title 'S&P 500 monthly data from Pindyck & Rubinfeld, Ex 10.5' ;
title2 'AR(1)/GARCH(1,1) model' ;
proc autoreg ;
model returnsp =/nlag = 1  archtest garch=(p=1,q=1);
run ;
title2 'Regression model with AR(1)/GARCH(1,1)' ;
proc autoreg ;
model returnsp = DR3 gpw/nlag = 1  archtest garch=(p=1,q=1);
run ;

```

The SAS output is listed below.

From examination of the output, the following conclusions can be reached:

- The p-values of the Q and LM tests are all very small, less than .0001. Therefore, the errors in the regression model exhibit conditional heteroscedasticity.
- Ordinary least squares estimates of the regression parameters are:

Variable	DF	Estimate	Standard Error	t Value	Approx Pr >  t
Intercept	1	0.0120	0.001755	6.86	<.0001
DR3	1	-0.8293	0.3061	-2.71	0.0070
GPW	1	-0.8550	0.2349	-3.64	0.0003

- Using residuals from the OLS estimates, the estimated residual autocorrelations are:

Estimates of Autocorrelations		
Lag	Covariance	Correlation
0	0.00108	1.000000
1	0.000253	0.234934

- Also, using OLS residuals, the estimate AR parameter is:

Estimates of Autoregressive Parameters

Lag	Coefficient	Standard Error	t Value
1	-0.234934	0.046929	-5.01

- Assuming AR(1)/GARCH(1,1) errors, the estimated parameters of the regression are:

Variable	DF	Estimate	Standard Error	t Value	Approx Pr >  t
Intercept	1	0.0125	0.001875	6.66	<.0001
DR3	1	-1.0665	0.3282	-3.25	0.0012
GPW	1	-0.7239	0.1992	-3.63	0.0003

- Notice that these differ slightly from OLS estimates.
- Since all p-values are small, both independent variables are significant.
- However, the Total R-square value is only 0.0551, so the regression has little predictive value.
- The estimated GARCH parameters are:

AR1	1	-0.2016	0.0603	-3.34	0.0008
ARCH0	1	0.000147	0.0000688	2.14	0.0320
ARCH1	1	0.1337	0.0404	3.31	0.0009
GARCH1	1	0.7254	0.0918	7.91	<.0001

- the estimate of  $\phi$  is  $-.2016$  in SAS's notation but  $+.2016$  in our notation. Thus, there is a **positive** association between returns and lagged returns
- Since all p-values are small, all GARCH parameters are significant.
- The GARCH1 estimate (0.7254) is larger than the ARCH1 (0.1337) estimate; this implies that the conditional variance will exhibit reasonably long persistence of volatility.

## Listing of SAS output

```

S&P 500 monthly data from Pindyck & Rubinfeld, Ex 10.5      1
  Regression model with AR(1)/GARCH(1,1)
                        17:04 Tuesday, April 10, 2001

      The AUTOREG Procedure

      Dependent Variable      RETURNSP

      Ordinary Least Squares Estimates

SSE           0.46677572      DFE           430
MSE           0.00109        Root MSE      0.03295
SBC           -1711.5219      AIC           -1723.7341
Regress R-Square      0.0551      Total R-Square      0.0551
Durbin-Watson      1.5203

      Q and LM Tests for ARCH Disturbances

      Order          Q      Pr > Q          LM      Pr > LM

      1             26.8804  <.0001          26.5159  <.0001
      2             27.1508  <.0001          27.1519  <.0001
      3             28.2188  <.0001          28.4391  <.0001
      4             28.6957  <.0001          28.4660  <.0001
      5             33.4112  <.0001          32.6168  <.0001
      6             34.0892  <.0001          32.6962  <.0001
      7             34.4187  <.0001          32.9617  <.0001
      8             34.6542  <.0001          32.9636  <.0001
      9             35.2228  <.0001          33.3330  0.0001
     10             35.3047  0.0001          33.4174  0.0002
     11             35.8274  0.0002          33.9440  0.0004
     12             36.0142  0.0003          33.9507  0.0007

      Variable      DF      Estimate      Standard      Approx
                   DF      Estimate      Error      t Value      Pr > |t|

Intercept          1         0.0120      0.001755         6.86         <.0001
DR3                1        -0.8293      0.3061         -2.71         0.0070
GPW                1        -0.8550      0.2349         -3.64         0.0003

      Estimates of Autocorrelations

      Lag      Covariance      Correlation

      0         0.00108         1.000000
      1         0.000253         0.234934

```

S&P 500 monthly data from Pindyck & Rubinfeld, Ex 10.5 2  
 Regression model with AR(1)/GARCH(1,1)  
 17:04 Tuesday, April 10, 2001

The AUTOREG Procedure

Estimates of Autocorrelations

Lag	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1
0											*****										
1											*****										

Preliminary MSE 0.00102

Estimates of Autoregressive Parameters

Lag	Coefficient	Standard Error	t Value
1	-0.234934	0.046929	-5.01

Algorithm converged.

GARCH Estimates

SSE	0.44176656	Observations	433
MSE	0.00102	Uncond Var	0.00104656
Log Likelihood	889.071523	Total R-Square	0.1058
SBC	-1735.6479	AIC	-1764.143
Normality Test	43.0751	Pr > ChiSq	<.0001

Variable	DF	Estimate	Standard Error	t Value	Approx Pr >  t
Intercept	1	0.0125	0.001875	6.66	<.0001
DR3	1	-1.0665	0.3282	-3.25	0.0012
GPW	1	-0.7239	0.1992	-3.63	0.0003
AR1	1	-0.2016	0.0603	-3.34	0.0008
ARCH0	1	0.000147	0.0000688	2.14	0.0320
ARCH1	1	0.1337	0.0404	3.31	0.0009
GARCH1	1	0.7254	0.0918	7.91	<.0001

## 8.10 I-GARCH models

I-GARCH or integrated GARCH processes were designed to model data that has persistent changes in volatility. A GARCH( $p, q$ ) process is station-

ary with a finite variance if

$$\sum_{i=1}^q \alpha_i + \sum_{i=1}^p \beta_i < 1.$$

A GARCH( $p, q$ ) process is called an I-GARCH process if

$$\sum_{i=1}^q \alpha_i + \sum_{i=1}^p \beta_i = 1.$$

I-GARCH processes are either non-stationary or have an infinite variance.

Infinite variance implies heavy-tailed, though a distribution can be heavy-tailed with a finite variance. To appreciate what an infinite variance processes can look like, we will do some simulation. Figure 8.6 shows 40,000 observations of ARCH(1) processes with  $\alpha_1 = .9, 1,$  and  $1.8$ . The same white noise process is used in each of the ARCH(1) processes. All three ARCH(1) processes are stationary but only the one with  $\alpha_1 = .9$  has a finite variance. The second process is an I-GARCH process (actually, I-ARCH since  $q = 0$ ). The third process has  $\alpha_1 > 1$  and so is more extreme than an I-GARCH process. Notice how all three processes do revert to their conditional mean of 0. The larger the value of  $\alpha_1$  the more the volatility comes in sharp bursts. The processes with  $\alpha_1 = .9$  and  $\alpha_1 = 1$  looks similar; there is no sudden change in behavior when the variance becomes infinite. The process with  $\alpha_1 = .9$  already has a heavy-tail despite having a finite variance. Increasing  $\alpha_1$  from 0.9 to 1 does not increase the tail weight dramatically.

Normal plots of the simulated data in Figure 8.6 are shown in Figure 8.7. Clearly, the larger the value of  $\alpha_1$ , the heavier the tails of the marginal distribution.

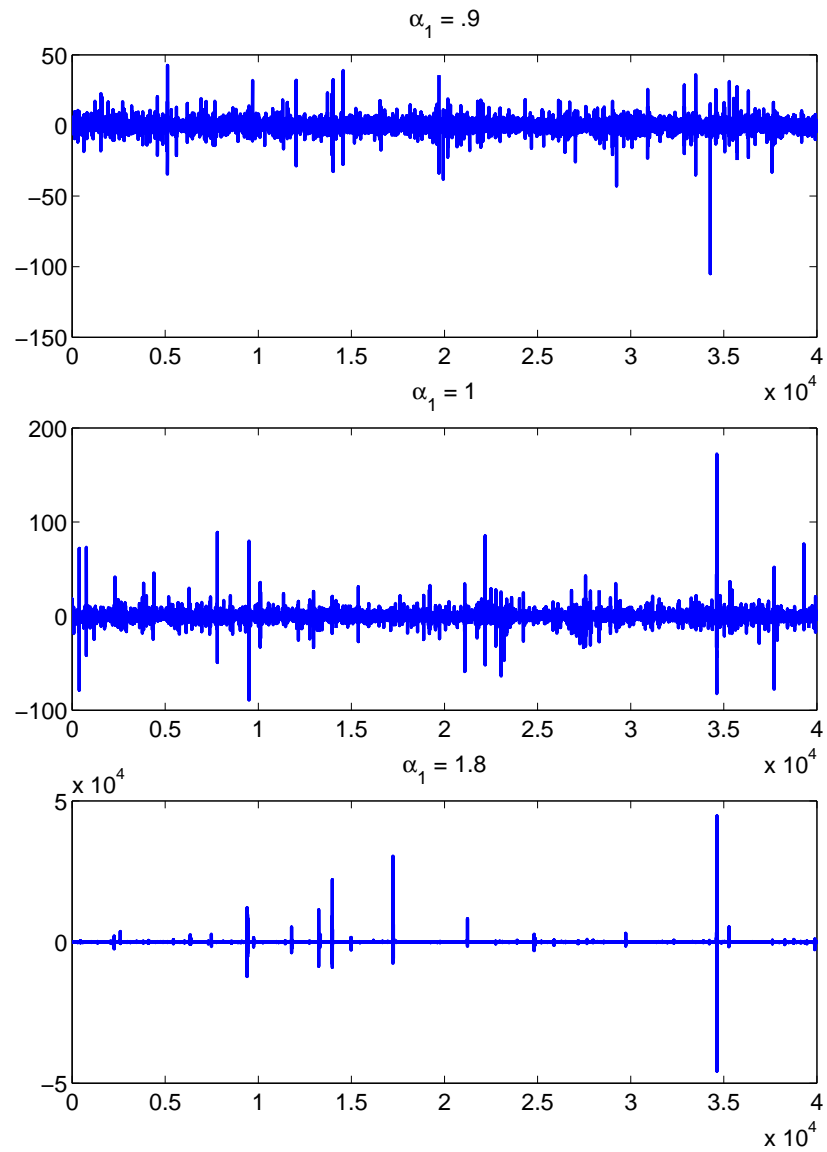


Figure 8.6: Simulated ARCH(1) processes with  $\alpha_1 = .9, 1, \text{ and } 1.8$ .



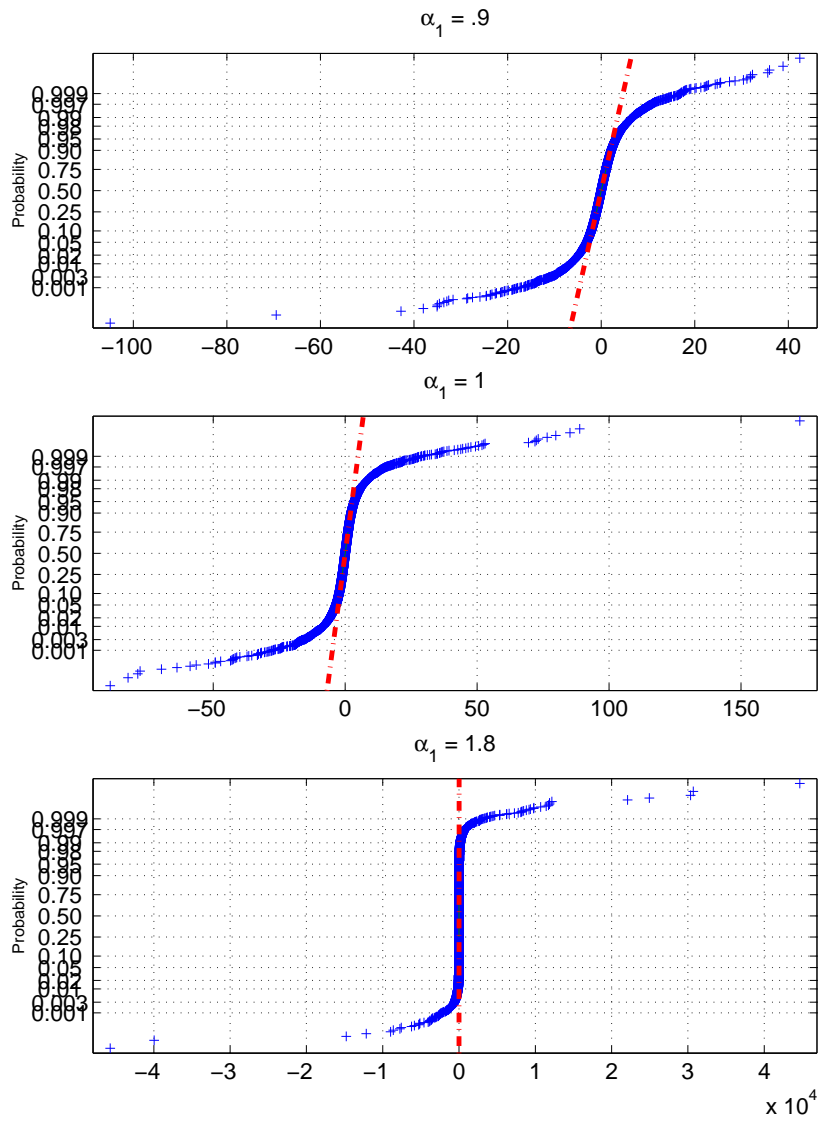


Figure 8.7: Normal plots of ARCH(1) processes in Figure 8.6.

None of the processes in Figure 8.6 show much persistence of higher volatility. To model persistence of higher volatility, one needs an I-GARCH( $p, q$ ) process with  $q \geq 1$ . Figure 8.8 shows simulations from I-GARCH(1,1) processes. Since  $\alpha_1 + \beta_1 = 1$  for these processes,  $\beta_1 = 1 - \alpha_1$ , and the process is completely specified by  $\alpha_0$  and  $\alpha_1$ . In this figure,  $\alpha_0$  is fixed at 1 and  $\alpha_1$  is varied. Notice that the conditional variance is very bursty when  $\alpha_1 = .95$ . When  $\alpha_1 = .05$ , the conditional standard deviation looks somewhat like a random walk.

I-GARCH processes can be fit by SAS by adding the specification “type = integrated” into the program, e.g., for the previous example with S&P 500 returns:

```
proc autoreg ;
model returnsp =/nlag = 1 garch=(p=1,q=1,type=integrated);
run ;
```

For this example, the I-GARCH(1,1) model seems to fit worse than a GARCH(1,1) model according to AIC; see Section 8.13.

### 8.10.1 What does it mean to have an infinite variance?

A random variable need not have a finite variance. Also, its expectation need not exist at all. To appreciate these facts, let  $X$  be a random variable with density  $f_X$ . The expectation of  $X$  is

$$\int_{-\infty}^{\infty} x f_X(x) dx$$

provided that this integral is defined. If

$$\int_{-\infty}^0 x f_X(x) dx = -\infty \tag{8.8}$$

and

$$\int_0^{\infty} x f_X(x) dx = \infty \tag{8.9}$$

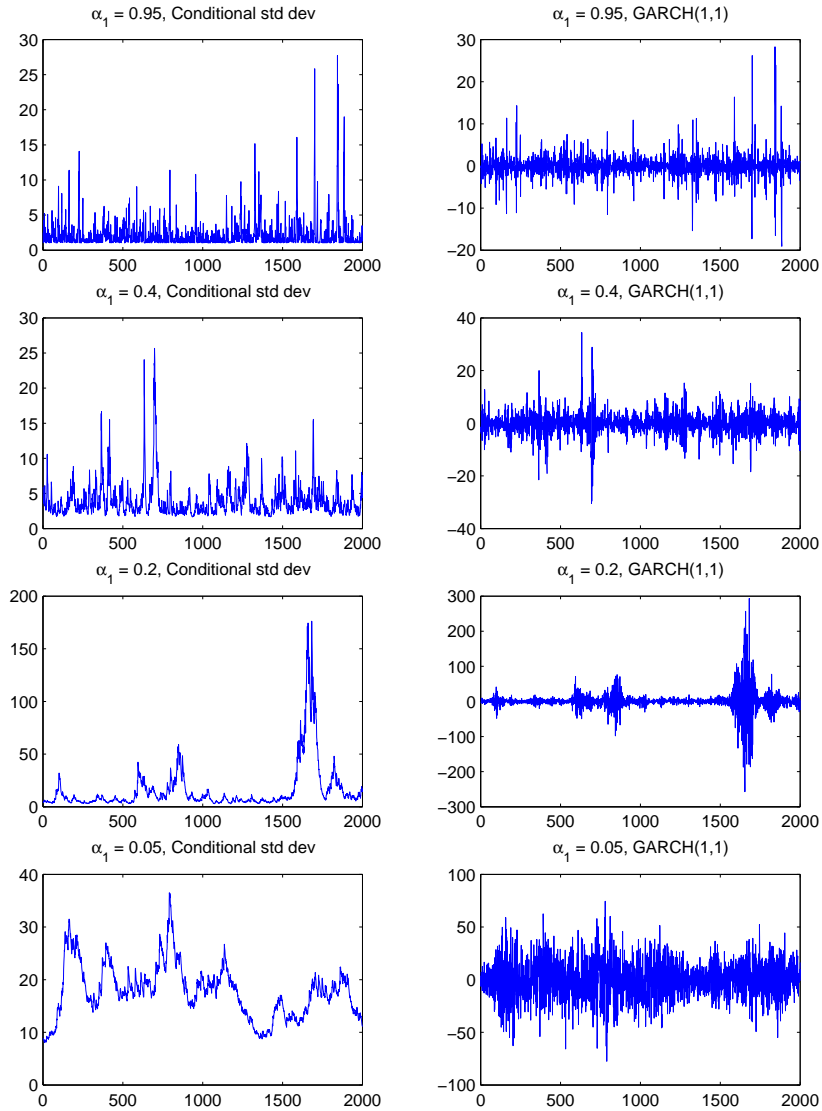


Figure 8.8: Simulations of I-GARCH(1,1) processes.

then the expectation is, formally,  $-\infty + \infty$  which is not defined. If integrals on the left hand sides of (8.8) and (8.8) are both finite, then  $E(X)$  exists and equals the sum of these two integrals.

### Exercise

Suppose that  $f_X(x) = 1/6$  if  $|x| < 1$  and  $f_X(x) = 1/(6x^2)$  if  $|x| \geq 1$ . Show that

$$\int_{-\infty}^{\infty} f_X(x) dx = 1$$

so that  $f_X$  really is a density, but that

$$\int_{-\infty}^0 x f_X(x) dx = -\infty$$

and

$$\int_0^{\infty} x f_X(x) dx = \infty$$

One consequence of the expectation not existing is this. Suppose we have a sample of iid random variables with density  $f_X$ . The law of large numbers says that the sample mean will converge to  $E(X)$  as the sample size goes to infinity. However, the law of large numbers holds only if  $E(X)$  is defined. Otherwise, there is no point to which the sample mean can converge and it will just wander without converging.

Figure 8.9 shows the sample mean of the first  $t$  observations plotted against  $t$  for the data in Figures 8.6 and 8.10. The sample mean appears to converge to 0 when  $\alpha_1 = .9$  or 1, but when  $\alpha_1 = 1.8$  it is unclear what the sample mean is doing. The sample mean decays towards 0 when the process is not in a high volatility period, but can shoot up or down during a burst of volatility.

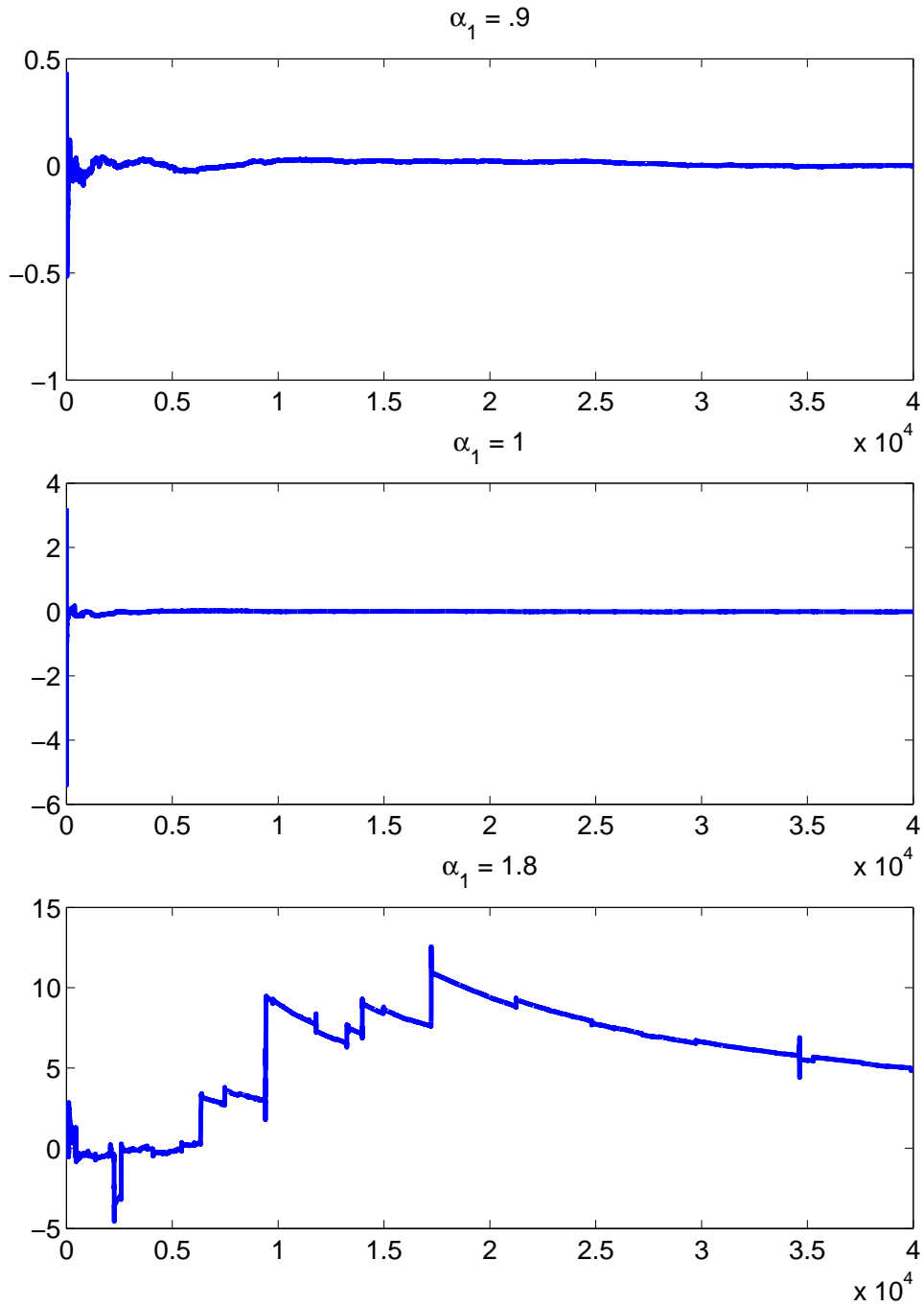


Figure 8.9: Sample means of simulated ARCH(1) processes with  $\alpha_1 = .9, 1,$  and  $1.8$ .

Now suppose that the expectation of  $X$  exists and equals  $\mu_X$ . Then the variance of  $X$  equals

$$\int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) dx.$$

If this integral is  $+\infty$ , then the variance is infinite.

The law of large numbers also implies that the sample variance will converge to the variance of  $X$  as the sample size increases. If the variance of  $X$  is infinity, then the sample variance will converge to infinity.

Figure 8.10 shows the sample variance of the first  $t$  observations plotted against  $t$  for the data in Figure 8.6.

In the top panel, the sample variance should be converging to  $10 = (1 - \alpha_1)^{-1}$ . Maybe it is converging to 10, but it is hard to tell even with 40,000 observations. In the middle and bottom panels the variance is infinity so the sample variance will converge to infinity. This convergence does appear to be happening in the bottom panel, but it is hard to see in the middle panel. Of course, in the middle panel the value of  $\alpha_1$  is on the borderline between finite and infinite variance, and the infinite variance may take a very long time to have its effect.

## 8.11 GARCH-M processes

We have seen that one can fit regression models with AR/GARCH errors. In fact, we have done that with the S&P 500 data. In some examples, it makes sense to use the conditional standard deviation as one of the regression variables. For example, when the dependent variable is a return we might expect that higher conditional variability will cause higher returns. This is because the market demands a higher risk premium for higher risk.

Models where the conditional standard deviation is a regression variable

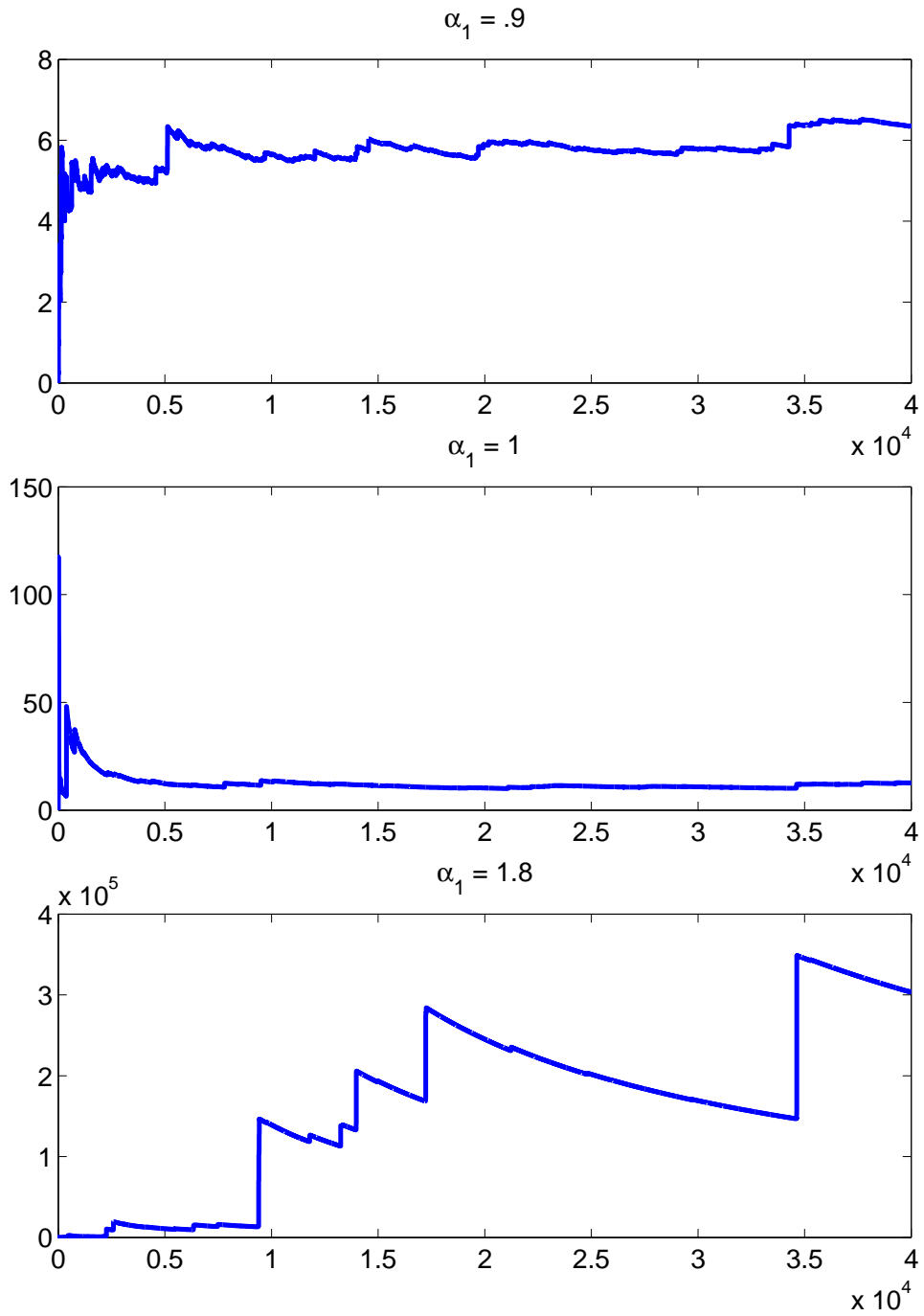


Figure 8.10: Sample variances of simulated ARCH(1) processes with  $\alpha_1 = .9, 1,$  and  $1.8$ .

are called GARCH-in-mean, or GARCH-M, models. They have the form

$$Y_t = \mathbf{X}_t^T \boldsymbol{\gamma} + \delta \sigma_t + a_t,$$

where  $a_t$  is a GARCH process with conditional standard deviation  $\sigma_t$ .

GARCH-M models can be fit in SAS by adding the keyword “mean” to the GARCH specification, e.g.,

```
proc autoreg ;
model returnsp =/nlag = 1 garch=(p=1,q=1,mean);
run ;
```

or for I-GARCH-M

```
proc autoreg ;
model returnsp =/nlag = 1 garch=(p=1,q=1,mean,type=integrated);
run ;
```

For the S&P 500 returns data, a GARCH(1,1)-M was fit in SAS. The estimate of  $\hat{\delta}$  was .5150 with a standard error of .3695. This gives a t-value of 1.39 and a p-value of .1633. Since the p-value is reasonably large we could accept the null hypothesis that  $\delta = 0$ . Therefore, we see no strong evidence that there are higher returns during times of higher volatility. The volatility of the S&P 500 is *market risk* so this finding is a bit surprising. It may be that the effect is small ( $\hat{\delta}$  is positive, after all) and cannot be detected with certainty. The AIC criterion *does* select the GARCH-M model; see Section 8.13.

## 8.12 E-GARCH

The exponential GARCH, or E-GARCH, model is

$$\log(\sigma_t) = \alpha_0 + \sum_{i=1}^q \alpha_i g(\epsilon_{t-i}) + \sum_{i=1}^p \beta_i \log(\sigma_{t-i}),$$

where

$$g(\epsilon_t) = \theta \epsilon_t + \gamma \{|\epsilon_t| - E(|\epsilon_t|)\}$$



and  $\epsilon_t = a_t/\sigma_t$ . Since  $\log(\sigma_t)$  can be negative, there are no constraints on the parameters.

Notice that

$$g(\epsilon_t) = -\gamma E(|\epsilon_t|) + (\gamma + \theta)|\epsilon_t| \quad \text{if } \epsilon_t > 0,$$

and

$$g(\epsilon_t) = -\gamma E(|\epsilon_t|) + (\gamma - \theta)|\epsilon_t| \quad \text{if } \epsilon_t < 0,$$

It is a good calculus exercise to show that  $E(|\epsilon_t|) = \sqrt{2/\pi} = .7979$ .

Typically,  $-1 < \hat{\theta} < 0$  so that  $0 < \gamma + \theta < \gamma - \theta$ . For example,  $\hat{\theta} = -.7$  in the S&P 500 example; see below. The function  $g$  with  $\theta = -.7$  is plotted in the top left panel of Figure 8.11. Notice that  $g(\epsilon_t)$  is negative if  $|\epsilon_t|$  is close to zero; small values of noise decrease  $\sigma_t$ . If  $|\epsilon_t|$  is large, then  $\sigma_t$  increases. With a negative value of  $\theta$ ,  $\sigma_t$  increases more rapidly as a function of  $|\epsilon_t|$  when  $\epsilon_t$  is negative than when  $\epsilon_t$  is positive,

In finance, the “leverage effect” predicts that a asset’s price will become more volatile when its price decreases. This is the type of behavior obtained when  $\theta < 0$ . The ability to accommodate leverage effects was the reason that the E-GARCH model was introduced by Daniel Nelson.

The function  $g$  for several other values of  $\theta$  are also shown in Figure 8.11. When  $\theta = 0$  (top right) the function is symmetric about 0. The bottom right panel where  $\theta = -1$  shows an extreme case where  $g(\epsilon_t)$  is negative for all positive  $\epsilon_t$

SAS fits the E-GARCH model with  $\gamma$  fixed as 1 and  $\theta$  estimated. The E-GARCH model is specified by using “type=exp” as in

```
proc autoreg ;
model returnsp =/nlag = 1 garch=(p=1,q=1,mean,type=exp) ;
run ;
```

This command specifies both a GARCH-in-mean effect and the E-GARCH model. Omitting “mean” removes the GARCH-in-mean effect.

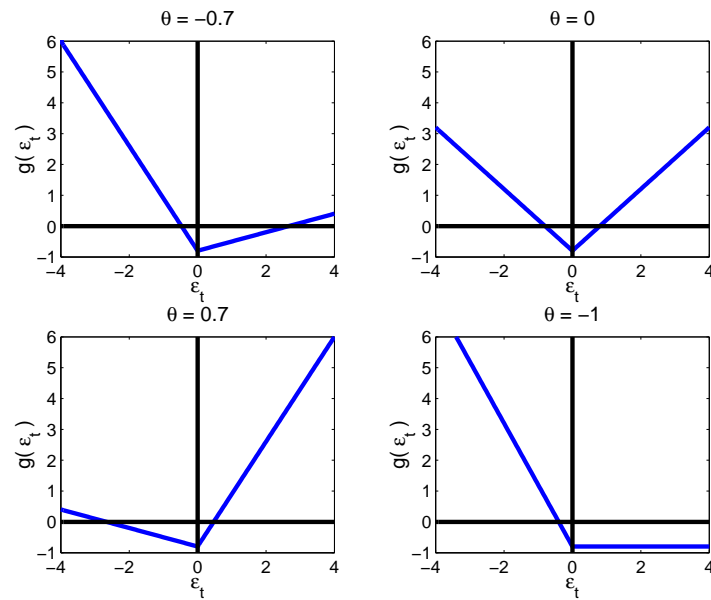


Figure 8.11: The  $g$  function for the S&P 500 data (top left panel) and several other values of  $\theta$ .

### 8.13 Back to the S&P 500 example

SAS can fit six different AR(1)/GARCH(1,1) models since SAS allows “type” to be “integrated,” “exp,” or “nonneg.” The last is the default and specifies a GARCH model with non-negativity constraints. Moreover, for each of these three types we can specify that a GARCH-in-mean effect be included or not. Table 8.1 contains the AIC statistics for the six models. The models are ordered from best fitting to worse fitting according to AIC—remember that a smaller AIC is better.

It seems that the E-GARCH-M model is best, though the E-GARCH model fits nearly as well. The E-GARCH-M model will be used in the remaining discussion. To see if more AR or GARCH parameters would improve the fit, AR(2) and E-GARCH(1,2)-M, E-GARCH(2,1)-M, and E-GARCH(2,2)-M models were tried, but none of these lowered AIC or had all parameters significant at  $p = .1$ . Thus, AR(1)/E-GARCH(1,1) appears to be a good fit to the noise and adding a GARCH-in-mean term to the regression model

<b>Model</b>	<b>AIC</b>	<b><math>\Delta</math> AIC</b>
E-GARCH-M	-1783.9	0
E-GARCH	-1783.1	0.8
GARCH-M	-1764.6	19.3
GARCH	-1764.1	19.8
I-GARCH-M	-1758.0	25.9
I-GARCH	-1756.4	27.5

Table 8.1: AIC statistics for six AR(1)/GARCH(1,1) models fit to the S&P 500 returns data.  $\Delta$  AIC is the change in AIC between a given model and E-GARCH-M.

seems reasonable although it does not improve the fit very much.

The fit to this model is in the SAS output listed below.

**Listing of SAS output for the E-GARCH-M model:**

S&P 500 monthly data from Pindyck & Rubinfeld, Ex 10.5      2  
 Regression model with AR(1)/E-GARCH(1,1)-M  
 11:52 Sunday, April 15, 2001

## The AUTOREG Procedure

## Estimates of Autoregressive Parameters

Lag	Coefficient	Standard Error	t Value
1	-0.234934	0.046929	-5.01

Algorithm converged.

## Exponential GARCH Estimates

SSE	0.44211939	Observations	433
MSE	0.00102	Uncond Var	.
Log Likelihood	900.962569	Total R-Square	0.1050
SBC	-1747.2885	AIC	-1783.9251
Normality Test	24.9607	Pr > ChiSq	<.0001

Variable	DF	Estimate	Standard Error	t Value	Approx Pr >  t
Intercept	1	-0.003791	0.0102	-0.37	0.7095
DR3	1	-1.2062	0.3044	-3.96	<.0001
GPW	1	-0.6456	0.2153	-3.00	0.0027
AR1	1	-0.2376	0.0592	-4.01	<.0001
EARCH0	1	-1.2400	0.4251	-2.92	0.0035
EARCH1	1	0.2520	0.0691	3.65	0.0003
EGARCH1	1	0.8220	0.0606	13.55	<.0001
THETA	1	-0.6940	0.2646	-2.62	0.0087
DELTA	1	0.5067	0.3511	1.44	0.1490

## 8.14 The GARCH zoo

There are many more types of GARCH models than the few mentioned so far. I've discussed only the most widely used models that can be fit in SAS. The number of models seems limited only by the number of letters in the alphabet, not the imagination of econometricians! Here's a sample of other GARCH models mentioned in Bollerslev, Engle, and Nelson (1994):

- QARCH = quadratic ARCH
- TARARCH = threshold ARCH
- STARARCH = structural ARCH
- SWARCH = switching ARCH
- QTARCH = quantitative threshold ARCH
- vector ARCH
- diagonal ARCH
- factor ARCH

## 8.15 Applications of GARCH in finance

GARCH models were developed by econometricians working with business and finance data, and their applications to finance have been extensive. The review paper by Bollerslev, Engle, and Nelson lists hundreds of references.

Finance models such as the CAPM and the Black-Scholes model for option pricing assume a constant conditional variance. When this assumption is false, use of these models can lead to serious errors. Therefore, generalization of finance models to include GARCH errors has been a hot topic. See Bollerslev, Engle, and Woolridge (1988) and Duan (1996a, 1996b) for some examples of finance models with conditional heteroscedasticity.

Rossi (1996) is a collection of papers, many reprinted from finance journals, on modeling stock market volatility with GARCH models.

## 8.16 Summary

- The marginal, or unconditional, distribution of a stationary process is the distribution of an observation from the process given no information about the previous or future observations
  - by stationarity the marginal distribution must be constant
  - in particular, the marginal mean and variance are constant
- Besides the marginal distribution, we are interested in the conditional distribution of the next observation given the current information set of present and past values of the process, and perhaps of other processes
- For ARMA processes the conditional mean is non-constant but the conditional variance is constant
- The constant conditional variance of ARMA processes makes them unsuitable for modeling the volatility of financial markets
- GARCH processes have non-constant conditional variance and were developed to model volatility
- GARCH processes can be used as the “noise” term of an ARMA process
  - ARMA/GARCH processes have both non-constant conditional mean and a non-constant conditional variance
  - GARCH and ARMA/GARCH processes can be estimated by maximum likelihood.
  - Proc Autoreg in SAS fits AR/GARCH models

- The simple ARCH( $q$ ) models have burst of volatility but cannot model persistent volatility
- The generalized ARCH (GARCH) models can model persistent volatility
- The marginal distribution of a GARCH process has heavier tails than the normal distribution.
  - heavy tails = outlier prone
  - in fact, for certain parameter values a GARCH process will have an infinite variance, which is an extreme case of heavy tails
    - \* I-GARCH (integrated GARCH) models are examples of GARCH models with infinite variance
- If the marginal variance is infinite, then the sample variance will converge to infinity as the sample size increase
- For extremely heavy tails, the marginal expectation may not exist
  - then there exists no point to which the sample mean can converge
    - \* the sample mean will wander aimlessly
- ARMA/GARCH processes can be used as the noise term in regression models
  - SAS's Proc Autoreg can use an AR/GARCH noise term in a regression model
- The **GARCH-M** models use the conditional standard deviation as an independent variable in the regression
- The “leverage effect” occurs when a negative return (drop in price) increases the volatility of future returns because the denominator of those returns is smaller.
- E-GARCH models were designed to capture the leverage effect

- in an E-GARCH model, the log of the conditional standard deviation is modeled as an ARMA process but with the white noise process  $\epsilon_t$  replaced by another white noise process  $g(\epsilon_t)$
- there is no need for non-negativity constraints on the parameters, such as those in an ordinary GARCH model, since the log standard deviation can be negative
- the parameter  $\theta$  in an E-GARCH model determines the leverage effects
  - \*  $\theta < 0 \Rightarrow$  leverage effects
  - \*  $\theta = 0 \Rightarrow$  no leverage
  - \*  $\theta > 0 \Rightarrow$  positive returns increase volatility (this would be the opposite of the leverage effect and is not expected to happen in practice)
- In the S&P 500 example we found that
  - returns are negatively associated with changes in interest rates (an increase in interest rates decreases returns)
  - returns are negatively associated with changes in wholesale prices
  - returns are positively associated with returns lagged one month ( $\hat{\phi} = -.2376$  is negative in the SAS output, but SAS's definition of  $\phi$  is the negative of ours—our  $\hat{\phi}$  is  $+.2376$ )
  - there are leverage effects since a E-GARCH model fits better than a GARCH model and  $\hat{\delta} = -.7$
  - there is slight evidence of a GARCH-in-mean effect, that is, there is some reason to believe that there is a risk premium
- There is a wide variety of other GARCH models in the literature, but the ones discussed here, ARCH( $q$ ), GARCH( $p, q$ ), E-GARCH, GARCH-M, and I-GARCH, are probably enough to know about since they can model a wide variety of data types
  - the models discussed in these notes are the ones that can be fit by SAS



- There is a large and growing literature on financial models with returns following GARCH processes

## 8.17 References

- Bodie, Z., Kane, A., and Marcus, A. (1999), *Investments*, Irwin/McGraw-Hill, Boston.
- Bollerslev, T. (1986), Generalized autoregressive conditional heteroskedasticity, *J. of Econometrics*, 31, 307–327.
- Bollerslev, T., Engle, R.F., and Nelson, D.B. (1994), ARCH models, in *Handbook of Econometrics, Volume IV*, Engle, R.F., and McFadden, D.L., editors, Elsevier.
- Bollerslev, T., Chou, R.Y., and Kroner, K.F. (1992), ARCH modeling in finance, *J. of Econometrics*, 52, 5-59.
- Bollerslev, T., and Engle, R.F. (1993), Common persistence in conditional variances, *Econometrica*, 61, 167–186.
- Bollerslev, T., Engle, R.F., and Nelson, D.B. (1994), ARCH models, In *Handbook of Econometrics, Vol IV*, R.F. Engle and D.L. McFadden, ed., Elsevier.
- Bollerslev, T., Engle, R.F., and Wooldridge, J.M. (1988). A capital asset pricing model with time-varying covariances, *J. of Political Economy*, 96, 116–131.
- Carroll, R.J., and Ruppert, D. (1988), *Transformation and Weighting in Regression*, Chapman & Hall, New York.
- Duan, J-C. (1996a). A unified theory of option pricing under stochastic volatility — from GARCH to diffusion, manuscript. (available at <http://www.bm.ust.hk/fina/staff/jcduan.html>)

- Duan, J.-C. (1996b). Term structure and bond option pricing under GARCH, manuscript. (available at <http://www.bm.ust.hk/fina/staff/jcduan.html>)
- Enders, W. (1995), *Applied Econometric Time Series*, Wiley, New York.
- Engle, R.F. (1982), Autoregressive conditional heteroskedasticity with estimates of variance of U.K. inflation, *Econometrica*, 50, 987–1008.
- Nelson, D.B. (1989). Modelling stock market volatility changes, ASA 1989 Proceedings of the Business and Economics Statistics Section, pp. 93–98. [Reprinted in Rossi (1996)]
- Pindyck, R.S., and Rubinfeld, D.L., (1998), *Econometric Models and Economic Forecasts*, Irwin/McGraw Hill, Boston.
- Rossi, P.E. (1996). *Modelling Stock Market Volatility*, Academic Press, San Diego.
- SAS Institute (1993), *SAS/ETS User's Guide, Version 6, 2nd Edition*, SAS Institute, Cary, NC.

# Chapter 9

## Fixed Income Securities: 4/30/01

### 9.1 Introduction

Corporations finance their operations by selling stock and bonds. Owning a share of stock means partial ownership of the company. You share in both the profits and losses of the company, so nothing is guaranteed.

Owning a bond is different. When you buy a bond you are loaning money to the corporation. The corporation is obligated to pay back the principle and to pay interest as stipulated by the bond. You receive a fixed stream of income, unless the corporation defaults on the bond. For this reason, bonds are called “fixed-income” securities.

It might appear that bonds are risk-free, almost stodgy. This is not the case. Many bonds are long-term, e.g., 20 or 30 years. Even if the corporation stays solvent or if you buy a US Treasury bond where default is virtually impossible, your income from the bond is guaranteed only if you keep the bond to maturity. If you sell the bond before maturity, your return will depend on changes in the price of the bond due to changes in interest rates.

The interest rate of your bond is fixed, but in the market interest rates fluctuate. Therefore, the market value of your bond fluctuates too. For

example, if you buy a bond paying 5% and the rate of interest increases to 6% then your bond is inferior to new bonds offering 6%. Consequently, the price of your bond will decrease. If you sell the bond you would lose money. So much for a “fixed income” stream!

If you ever bought a CD, which really is a bond that you buy from a bank or credit union, you will have noticed that the interest rate depends on the maturity of the CD. This is a general phenomenon. For example, on March 28, 2001, the interest rate of Treasury bills<sup>1</sup> was 4.23% for 3-month bills. The yields on Treasuries were 4.41%, 5.01%, and 5.46% for 2, 10, and 30 year maturities, respectively. The **term structure** of interest rates describes how rates of interest change with the maturity of bonds.

In this chapter we will study how bond prices fluctuate due to interest rate changes. We will also study how the term structure of interest rates can be determined.

## 9.2 Zero coupon bonds

Zero-coupon bonds, also called pure discount bonds, pay no principle or interest until maturity. A “zero” has a par value which is the payment made to the bond holder at maturity. The zero sells for less than par, which is the reason it is a “discount bond.”

For example, consider a 20-year zero with a par value of \$1000 and 6% interest compounded annually. The price is the present-value of \$1000 with discounting annually at 6%. That is, the price is

$$\frac{\$1000}{(1.06)^{20}} = \$311.80.$$

---

<sup>1</sup>Treasury bills have maturities of one year or less, Treasury notes have maturities from one to ten years, and Treasury bonds have maturities from 10 to 30 years.

If the interest is 6% but compounded every six months, then the price is

$$\frac{\$1000}{(1.03)^{40}} = \$306.56,$$

and if the interest is 6% compounded continuously then the price is

$$\frac{\$1000}{\exp\{(.06)(20)\}} = \$301.19.$$

### 9.2.1 Price and returns fluctuate with the interest rate

For concreteness, assume semi-annual compounding. Suppose you just bought the zero for \$306.56 and then six months later the interest rate increased to 7%. The price would now be

$$\frac{\$1000}{(1.035)^{39}} = \$261.41$$

so your investment would drop by  $(\$306.56 - \$261.41) = \$45.15$ . You will still get your \$1000 if you keep the bond for 20 years, but if you sell it now you will lose \$45.15. This is a return of

$$\frac{-45.15}{306.56} = -14.73\%$$

for a half-year or  $-29.46\%$  per year. And the interest rate only changed from 6% to 7%!

If the interest rate dropped to 5% after six months, then your bond would be worth

$$\frac{\$1000}{(1.025)^{39}} = \$381.74.$$

This would be an annual rate of return of

$$2 \left( \frac{381.74 - 306.56}{306.56} \right) = 49.05\%.$$

If the interest rate remained unchanged at 6%, then the price of the bond would be

$$\frac{\$1000}{(1.03)^{39}} = \$315.75.$$

The annual rate of return would be

$$2 \left( \frac{315.75 - 306.56}{306.56} \right) = 6\%.$$

Thus, if the interest rate does not change, you can earn a 6% annual rate of return by selling the bond before maturity. If the interest rate does change, however, the 6% annual rate of return is guaranteed only if you keep the bond until maturity.

### 9.3 Coupon bonds

Coupon bonds make regular interest payments.<sup>2</sup> Coupon bonds generally sell at par when issued. At maturity, one receives the principle and the final interest payment.

As an example, consider a 20-year coupon bond with a par value of \$1000 and 6% annual interest with semi-annual coupon payments. Each coupon payment will be \$30. Thus, the bond holder receives 40 payments of \$30, one every six months plus a principle payment of \$1000 after 20 years. One can check that the present value of all payments, with discounting at the 6% annual rate (3% semi-annual), equals \$1000:

$$\sum_{t=1}^{40} \frac{30}{(1.03)^t} + \frac{1000}{(1.03)^{40}} = 1000.$$

After six months if the interest rate is unchanged, then the bond (including the first coupon payment which is now due) is worth

$$\sum_{t=0}^{39} \frac{30}{(1.03)^t} + \frac{1000}{(1.03)^{39}} = (1.03) \left( \sum_{t=1}^{40} \frac{30}{(1.03)^t} + \frac{1000}{(1.03)^{40}} \right) = 1030,$$

which is a 6% annual return as expected. If the interest rate increases to 7%, then after six months the bond (plus the interest due) is only worth

$$\sum_{t=0}^{39} \frac{30}{(1.035)^t} + \frac{1000}{(1.035)^{39}} = (1.035) \left( \sum_{t=1}^{40} \frac{30}{(1.035)^t} + \frac{1000}{(1.035)^{40}} \right) = 924.49.$$

---

<sup>2</sup>At one time actual coupons were attached to the bond, one coupon for each interest payment. When a payment was due, its coupon could be clipped off and sent to the issuing company for payment.

This is an annual return of

$$2 \left( \frac{924.49 - 1000}{1000} \right) = -15.1\%.$$

If the interest rate drops to 5% after six months then the investment is worth

$$\sum_{t=0}^{39} \frac{30}{(1.025)^t} + \frac{1000}{(1.025)^{39}} = (1.025) \left( \sum_{t=1}^{40} \frac{30}{(1.025)^t} + \frac{1000}{(1.025)^{40}} \right) = 1,153.70, \quad (9.1)$$

and the annual return is

$$2 \left( \frac{1153.6 - 1000}{1000} \right) = 30.72\%.$$

### Some general formulas

Let's derive some useful formulas. If a bond with a par value of PAR matures in  $T$  years and makes semi-annual payments of  $C$  and the discount rate (rate of interest) is  $r$  per half-year, then the value of the bond when it is issued is

$$\begin{aligned} \sum_{t=1}^{2T} \frac{C}{(1+r)^t} + \frac{\text{PAR}}{(1+r)^{2T}} &= \frac{C}{r} \{1 - (1+r)^{-2T}\} + \frac{\text{PAR}}{(1+r)^{2T}} \\ &= \frac{C}{r} + \left\{ \text{PAR} - \frac{C}{r} \right\} (1+r)^{-2T}. \end{aligned} \quad (9.2)$$

If  $C = \text{PAR} \times r$ , then the value of the bond when issued is PAR. The value six months later is  $(1+r)$  times the value in equation (9.2). The MATLAB function "bondvalue.m" computes (9.2). The call to this function is `bondvalue(c, T, r, par)`.

For example, if the coupon is  $C = 30$ , if  $T = 30$ , and if after six months  $r = 6.2\%$ /half-year (or 3.1%/year), then the bond is worth

$$(1.031) \left[ \frac{30}{.031} \{1 - (1.031)^{-60}\} + \frac{1000}{(1.031)^{60}} \right] = 1003.1.$$

This value was computed by MATLAB with the call

1.031\*bondvalue(30,30,.031,1000).

Similarly, (9.1) was computed with the MATLAB call

1.025\*bondvalue(30,20,.025,1000).

### Derivation of (9.2)

The summation formula for a finite geometric series is

$$\sum_{i=0}^T r^i = \frac{1 - r^{T+1}}{1 - r}, \quad (9.3)$$

provided that  $r \neq 1$ . Therefore,

$$\sum_{t=1}^{2T} \frac{C}{(1+r)^t} = \frac{C}{1+r} \sum_{t=0}^{2T-1} \frac{1}{(1+r)^t} = \frac{C\{1 - (1+r)^{-2T}\}}{(1+r)(1 - (1+r)^{-1})} \quad (9.4)$$

$$= \frac{C}{r} \{1 - (1+r)^{-2T}\}. \quad (9.5)$$

## 9.4 Yield to maturity

Suppose a bond with  $T = 30$  and  $C = 40$  is selling for \$1200, \$200 above par. If the bond were selling at par, then the interest rate would be .04/half-year (= .08/year). The 4%/half-year rate is called the **coupon rate**.

But the bond is *not* selling at par. If you purchase the bond at \$1200 you will make *less* than 8% per year interest. There are two reasons when the rate of interest is less than 8%. First, the coupon payments are \$40 or  $40/1200 = 3.333\%$ /half-year of the \$1200 investment; 3.333% is called the **current yield**. Second, at maturity you only get back \$1000 of the \$1200 investment. The current yield overestimates the return since it does not account for this loss of capital.

The **yield to maturity** is a measure of the average rate of return, including the loss (or gain) of capital because the bond was purchased above (or



below) par. For this bond, the yield to maturity is the value of  $r$  that solves

$$1200 = \frac{40}{r} + \left\{ 1000 - \frac{40}{r} \right\} (1 + r)^{-60}. \quad (9.6)$$

The right hand side of (9.6) is (9.5) with  $C = 40$ ,  $T = 30$ , and  $\text{PAR} = 1000$ . It is easy to solve equation (9.6) numerically. The MATLAB program `yield.m` does the following:

- computes the bond price for each  $r$  value on a grid
- graphs bond price versus  $r$  (this is not necessary but it's fun to see the graph)
- interpolates to find the value of  $r$  when bond value equals 1200

One finds that the yield to maturity is 0.0324. Figure 9.1 shows the graph of bond price versus  $r$  and shows that  $r = .0324$  maps to a bond price of \$1200.

The yield to maturity of .0324 is less than the current yield of 0.0333 which is less than the coupon rate of  $40/1000 = .04$ . (All three rates are rates per half-year.) Thus, we see that

- coupon rate > current yield
  - since the bond sells above par
- current yield > yield to maturity
  - since yield to maturity accounts for the loss of capital when at the maturity date you only get back \$1000 of the \$1200 investment

Whenever, as in this example, the bond is selling above par, we have

$$\text{coupon rate} > \text{current yield} > \text{yield to maturity} \quad (9.7)$$

Everything is reversed if the bond is selling below par. For example, if the price of the bond were only \$900, then

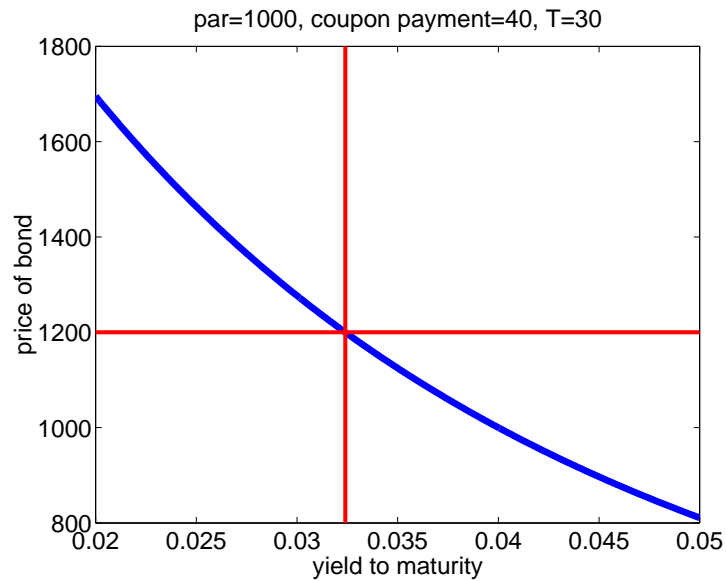


Figure 9.1: Bond price versus the interest rate  $r$  and determining by interpolation the yield to maturity when the price equals \$1200.

- the yield to maturity would be 0.0448 (as before, this value can be determined by “yield.m” using interpolation)
- the current yield would be  $40/900 = 0.0444$
- The coupon rate would still be  $40/1000 = .04$

Therefore we would have

$$\text{coupon rate} < \text{current yield} < \text{yield to maturity}, \quad (9.8)$$

which is just the opposite of (9.7).

### 9.4.1 Spot rates

The yield to maturity of a zero coupon bond of maturity  $n$  years is called the  $n$ -year **spot rate**.

A coupon bond is a bundle of zero coupon bonds, one for each coupon payment and a final one for the principle payment. The component zeros have different maturity dates and therefore different spot rates. The yield to maturity of the coupon bond is a complex “average” of the spot rates of the zeros in this bundle.

## 9.5 Term structure

On January 26, 2001, the *Ithaca Journal* stated that 1-year T-bill rate was 4.83% and the 30-year Treasury bond rate was 6.11%. This is typical—short and long term rates usually do differ. Such differences can be seen in Figure 10.5 of Campbell et al. or Figure 15.7 of Bodie, Kane, and Marcus (1999).

Often short term rates are lower than long-term rates. This makes sense since long term bonds are riskier. Long term bond prices fluctuate more with interest rate changes and these bonds are often sold before maturity. In contrast, a 90-day or even 1-year T-bill is often kept to maturity and so is really a risk-free “fixed income security.” However, during periods of very high short-term rates, the short-term rates may be higher than the long term rates. The reason is that the market believes that rates will return to historic levels and no one will commit to the high interest rate for, say, 20 or 30 years.

The **term structure** of interest rates is a description of how, *at a given time*, yield to maturity depends on maturity. Term structure for all maturities up to  $n$  years can be described by any one of the following:

- prices of zero coupon bonds of maturities 1-year, 2-years,  $\dots$ ,  $n$ -years denoted here by  $P(1), P(2), \dots, P(n)$
- spot rates (yields of maturity of zero coupon bonds) of maturities 1-year, 2-years,  $\dots$ ,  $n$ -years denoted by  $y_1, \dots, y_n$
- forwards rates  $r_1, \dots, r_n$

As will be seen below, each of the sets

- $\{P(1), \dots, P(n)\}$
- $\{y_1, \dots, y_n\}$ ,

and

- $\{r_1, \dots, r_n\}$

can be computed from either one of the other sets. For example, (9.10) gives  $\{P(1), \dots, P(n)\}$  in terms of  $\{r_1, \dots, r_n\}$ . Also, equations (9.11) and (9.12) give  $\{y_1, \dots, y_n\}$  in terms of  $\{P(1), \dots, P(n)\}$  or  $\{r_1, \dots, r_n\}$ , respectively.

Term structure can be described by breaking down the time interval between the present time and the maturity time of a bond into short time segments with a constant interest rate within each segment, but with interest rates varying between segments. For example, a 3-year loan can be considered as three consecutive 1-year loans.

**Example:**

As an illustration, suppose that the three 1-year loans have the forward interest rates listed in Table 9.1.

Year ( $i$ )	Interest rate ( $r_i$ )
1	6%
2	7%
3	8%

Table 9.1: *Forward interest rate example*

Using the forward rates in Table 9.1, we see that a par \$1000 1-year zero would sell for

$$\frac{1000}{1 + r_1} = \frac{1000}{1.06} = \$943.40 = P(1).$$

A par \$1000 2-year zero would sell for

$$\frac{1000}{(1+r_1)(1+r_2)} = \frac{1000}{(1.06)(1.07)} = \$881.68 = P(2).$$

A par \$1000 3-year zero would sell for

$$\frac{1000}{(1+r_1)(1+r_2)(1+r_3)} = \frac{1000}{(1.06)(1.07)(1.08)} = 816.37 = P(3).$$

The general formula for the present value of \$1 paid  $n$  periods from now is

$$\frac{1}{(1+r_1)(1+r_2)\cdots(1+r_n)}. \quad (9.9)$$

Here  $r_i$  is the **forward interest rate** during the  $i$ th period. By “forward rate” we mean the price for that period that is agreed upon *now*.

Letting  $P(n)$  be the price of an  $n$ -year zero par \$1000 coupon bond,  $P(n)$  is \$1000 times the discount factor in (9.9), that is,

$$P(n) = \frac{1000}{(1+r_1)\cdots(1+r_n)}. \quad (9.10)$$

### Back to the example

Let's look at the yields to maturity. For a 1-year zero, the yield to maturity  $y_1$  solves

$$\frac{1000}{(1+y_1)} = 993.40,$$

which implies that  $y_1 = .06$ . Nothing surprising here, since  $r_1 = .06$ ! For a 2-year zero, the yield to maturity is  $y_2$  that solves

$$\frac{1000}{(1+y_2)^2} = 881.68.$$

Thus,

$$y_2 = \sqrt{\frac{1000}{881.68}} - 1 = .0650.$$

It is easy to show that  $y_2$  is also given by

$$y_2 = \sqrt{(1 + r_1)(1 + r_2)} - 1 = \sqrt{(1.06)(1.07)} - 1 = .0650$$

For a 3-year zero, the yield to maturity  $y_3$  solves

$$\frac{1000}{(1 + y_3)^3} = \frac{1000}{881.68}.$$

Also,

$$y_3 = \{(1 + r_1)(1 + r_2)(1 + r_3)\}^{1/3} - 1 = \{(1.06)(1.07)(1.08)\}^{1/3} - 1 = .0700,$$

or, more precisely .069969. Thus,  $(1 + y_3)$  is the geometric average of 1.06, 1.07, and 1.08 and approximately equal to their arithmetic average.

Recall that  $P(n)$  is the price of a par \$1000  $n$ -year zero coupon bond. The general formulas for the yield to maturity  $y_n$  of an  $n$ -year zero are

$$y_n = \left\{ \frac{1000}{P(n)} \right\}^{1/n} - 1, \quad (9.11)$$

and

$$y_n = \{(1 + r_1) \cdots (1 + r_n)\}^{1/n} - 1. \quad (9.12)$$

Equations (9.11) and (9.12) give the yields to maturity in terms of the bond prices and forward rates, respectively. Also,

$$P(n) = \frac{1000}{(1 + y_n)^n}, \quad (9.13)$$

which give  $P(n)$  in terms of the yield to maturity.

As mentioned before, interest rates for future years are called **forward rates**. A forward contract is an agreement to buy or sell an asset at some fixed future date at a fixed price. Since  $r_2, r_3, \dots$  are rates at future dates that are fixed now when a long-term bond is purchased, they are forward rates.

maturity	price
1 year	\$920
2 year	\$830
3 year	\$760

Table 9.2: Bond price example

The general formula for determining forward rates from yields to maturity is

$$r_1 = y_1, \quad (9.14)$$

and

$$r_n = \frac{(1 + y_n)^n}{(1 + y_{n-1})^{n-1}}. \quad (9.15)$$

Now suppose that we only observed bond prices. Can we calculate yields to maturity and forward rates? The answer is “yes, using (9.11) and then (9.15).”

**Example:**

Suppose that 1, 2, and 3-year par 1000 zeros are priced as Table 9.2. Then using (9.11), the yields to maturity are:

$$y_1 = \frac{1000}{920} - 1 = .087,$$

$$y_2 = \left\{ \frac{1000}{830} \right\}^{1/2} - 1 = .0976,$$

$$y_3 = \left\{ \frac{1000}{760} \right\}^{1/3} - 1 = .096,$$

Then, using (9.14) and (9.15)

$$r_1 = y_1 = .087,$$

$$r_2 = \frac{(1 + y_2)^2}{(1 + y_1)} = \frac{(1.0976)^2}{1.0876} - 1 = .108,$$

and

$$r_3 = \frac{(1 + y_3)^3}{(1 + y_2)^2} - 1 = \frac{(1.096)^3}{(1.0976)^2} - 1 = .093.$$

The formula for finding  $r_n$  from the prices of zero coupon bonds is

$$r_n = \frac{P(n-1)}{P(n)} - 1. \quad (9.16)$$

which can be derived from

$$P(n) = \frac{1000}{(1 + r_1)(1 + r_2) \cdots (1 + r_n)},$$

and

$$P(n-1) = \frac{1000}{(1 + r_1)(1 + r_2) \cdots (1 + r_{n-1})}.$$

To calculate  $r_1$  using (9.16), we need  $P(0)$ , the price of a 0-year bond, but  $P(0)$  is simply the par value. (Trivially, a bond which must be paid back immediately is worth exactly its par value.)

Example

Thus, using (9.16)

$$r_1 = \frac{1000}{920} - 1 = .0870,$$

$$r_2 = \frac{920}{830} - 1 = .1084,$$

and

$$r_3 = \frac{830}{760} - 1 = .0921.$$

## 9.6 Continuous compounding

Now we will assume continuous compounding with forward rates  $r_1, \dots, r_n$ . We will see that the use of continuous compounding rates simplifies the



relationships between the forward rates, the yields to maturity, and the prices of zero coupon bonds.

If  $P(n)$  is the price of a \$1000 par  $n$ -year zero coupon bond, then

$$P(n) = \frac{1000}{\exp(r_1 + r_2 + \cdots + r_n)}. \quad (9.17)$$

Therefore,

$$\frac{P(n-1)}{P(n)} = \frac{\exp(r_1 + \cdots + r_n)}{\exp(r_1 + \cdots + r_{n-1})} = \exp(r_n), \quad (9.18)$$

and

$$\log \left\{ \frac{P(n-1)}{P(n)} \right\} = r_n.$$

The yield to maturity of an  $n$ -year zero coupon bond solves the equation

$$P(n) = \frac{1000}{\exp(ny_n)},$$

and is easily seen to be

$$y_n = (r_1 + \cdots + r_n)/n.$$

Therefore,  $\{r_1, \dots, r_n\}$  is easily found from  $\{y_1, \dots, y_n\}$  by the relationship

$$r_1 = y_1,$$

and

$$r_n = ny_n - (n-1)y_{n-1} \quad \text{for } n > 1.$$

### Example

Using the prices in Table 9.2 (converted from par 1000 to par 1) we have  $P(1) = .930$ ,  $P(2) = .850$ , and  $P(3) = .760$ . Therefore,

$$r_1 = \log \left\{ \frac{1}{.930} \right\} = .0725,$$

$$r_2 = \log \left\{ \frac{.930}{.850} \right\} = .0899,$$

and

$$r_3 = \log \left\{ \frac{.850}{.760} \right\} = .1119.$$

Also,

$$y_1 = r_1 = .0725,$$

$$y_2 = (r_1 + r_2)/2 = .0813,$$

and

$$y_3 = (r_1 + r_2 + r_3)/3 = .0915.$$

### 9.6.1 Continuous forward rates

So far, we have assumed that forward interest rates vary from year to year, but that these rates are constant within each year. This assumption is, of course, unrealistic. The forward rates should be modeled as a function varying continuously in time, rather than as functions that are constant for one year at a time. It is unrealistic to assume a starting time and that interest rates change each year after this starting time. In fact, bonds are issued repeatedly and bonds of many maturities are on the market.

To specify the term structure in a realistic way, we will assume that there is a function  $r(t)$  called the **forward rate function** such that the price of a zero coupon bond of maturity  $T$  and with par equal to 1 is given by

$$P(T) = \exp \left\{ - \int_0^T r(t) dt \right\}. \quad (9.19)$$

Formula (9.19) is a generalization of formula (9.17). To appreciate this, suppose that

$$r(t) = r_k \quad \text{for} \quad k - 1 < t \leq k.$$

With this piecewise constant  $r$ ,

$$\int_0^T r(t) dt = r_1 + r_2 + \dots + r_T,$$

so that

$$\exp \left\{ - \int_0^T r(t) dt \right\} = \exp \{ -(r_1 + \dots + r_T) \}$$

and therefore (9.17) agrees with (9.19).

The yield to maturity of a bond with maturity date  $T$  is

$$y_T = \frac{1}{T} \int_0^T r(t) dt. \tag{9.20}$$

Think of (9.20) as the average of  $r(t)$  over the interval  $0 \leq t \leq T$ .

Jarrow, Ruppert, and Yu (2001) estimate  $r(t)$  in (9.19) under the assumption that  $r(t)$  is a member of a flexible class of functions called splines. Figure 9.2 shows estimated forward rate curves for US Treasury bonds. Figure 9.3 shows estimated forward rate curves AT&T bonds. These two figures come from Jarrow, Ruppert, and Yu (2001).

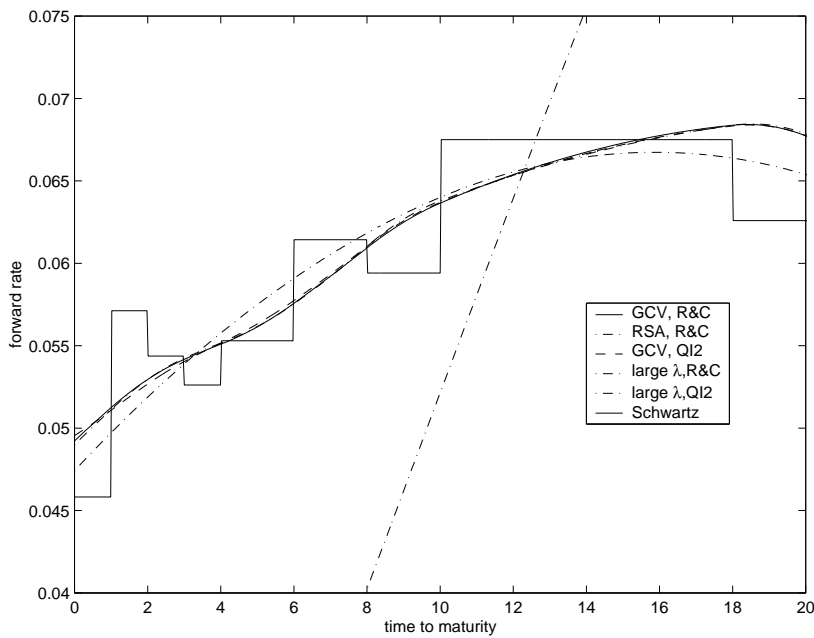


Figure 9.2: Spline estimates of forward rates of US Treasury bonds.

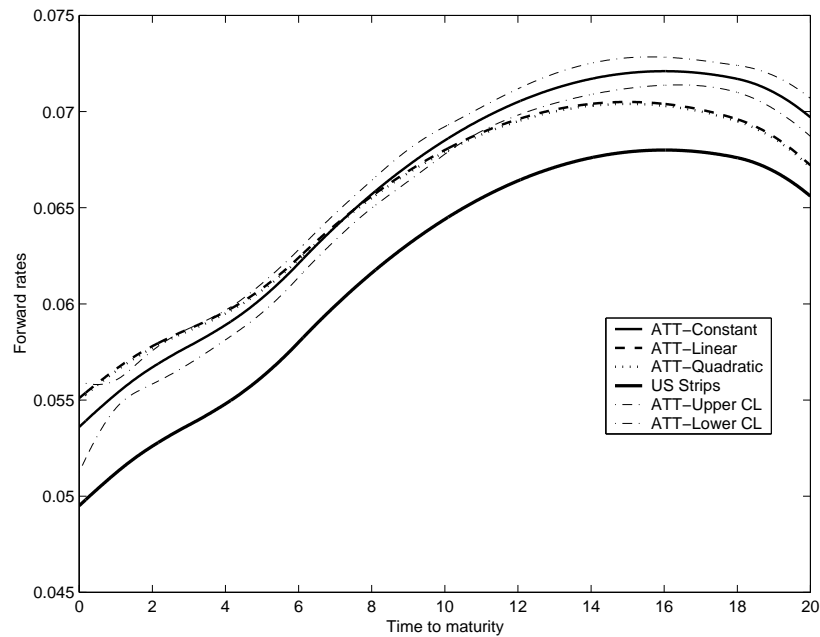


Figure 9.3: Spline estimates of forward rates of AT&T bonds

## 9.7 Summary

### 9.7.1 Introduction

- buy a bond = making a loan to the company
  - corporation is obligated to pay back the principle and interest (unless it defaults)
  - you receive a fixed stream of income
  - bonds are called “fixed-income” securities
  - for long term bond your income is guaranteed only if you keep the bond to maturity

### 9.7.2 Zero coupon bonds

- Zero-coupon bonds pay no principle or interest until maturity
- zero-coupon bond = pure discount bond
- par value is the payment made to the bond holder at maturity
- a zero sells for less than par
- **Example:** 20-year zero
  - par value of \$1000
  - interest 6% compounded every six months  $\Rightarrow$  price is

$$\frac{\$1000}{(1.03)^{40}} = \$306.56,$$

### 9.7.3 Risk due to interest rate changes

- bond prices fluctuate with the interest rate
  - **Example:** assume semi-annual compounding
  - you just bought the zero for \$306.56
    - \* six months later the interest rate increased to 7%
  - price would now be

$$\frac{\$1000}{(1.035)^{39}} = \$261.41$$

- investment would drop by  $(\$306.56 - \$261.41) = \$45.15$
- return of

$$\frac{-45.15}{306.56} = -14.73\%$$

for a half-year or  $-29.46\%$  per year

- however, if the interest rate remains unchanged then the bond is worth

$$\frac{\$1000}{(1.03)^{39}} = (1.03)(\$306.56)$$

- \* 3%/half-year return

### 9.7.4 Coupon bonds

- coupon bonds make regular interest payments
- consider a 20-year coupon bond with a par value of \$1000 and 6% annual interest with semi-annual coupon payments
  - coupon payment will be \$30
  - bond holder receives 40 payments of \$30
    - \* plus a principle payment of \$1000 after 20 years
  - present value of all payments, with discounting at the 6% annual rate (3% semi-annual), equals \$1000:

$$\sum_{t=1}^{40} \frac{30}{(1.03)^t} + \frac{1000}{(1.03)^{40}} = 1000.$$

- **General formula:**

- Notation
  - \* PAR = par value
  - \*  $C$  = coupon payment
  - \*  $T$  = maturity
  - \*  $r$  = interest rate per half-year
- bond price =

$$\begin{aligned} & \sum_{t=1}^{2T} \frac{C}{(1+r)^t} + \frac{\text{PAR}}{(1+r)^{2T}} \\ &= \frac{C}{r} + \left\{ \text{PAR} - \frac{C}{r} \right\} (1+r)^{-2T} \end{aligned}$$

- **Yield to maturity**
- **Example:** a bond with  $T = 30$  and  $C = 40$  is selling for \$1200,
- bond selling at par  $\Rightarrow$  interest rate = .04/half-year (= .08/year).
  - 4%/half-year rate = **coupon rate.**

- but *not* selling at par  $\Rightarrow$  if you purchase the bond at \$1200 you will make *less* than 8% per year
- two problems
  - coupon payments are \$40 or  $40/1200 = 3.333\%$ /half-year of the \$1200 investment
    - \* 3.333% is called the **current yield**
  - at maturity you only get back \$1000 of the \$1200 investment
  - **yield to maturity** = the average rate of return
- **Spot rates**
  - The yield to maturity of a zero coupon bond of maturity  $n$  years is called the  $n$  year **spot rate**.
  - A coupon bond is a bundle of zeros, each with a different maturity and therefore a different spot rate
    - \* the yield to maturity of a coupon bond is a complex “average” of these different spot rates

### 9.7.5 Term structure of interest rates

- term structure is description of how, *at a given time*, yield to maturity depends on maturity
- term structure for all maturities up to  $n$  years can be described by any one of the following sets:
  - prices of zero coupon bonds of maturities 1-year, 2-years, ...,  $n$ -years denoted here by  $P(1), P(2), \dots, P(n)$
  - spot rates (yields of maturity of zero coupon bonds) of maturities 1-year, 2-years, ...,  $n$ -years denoted by  $y_1, \dots, y_n$
  - forwards rates  $r_1, \dots, r_n$
- each of the above sets can be computed from either of the other sets.

### 9.7.6 Continuous compounding

- continuous compounding simplifies the relationships between
  - forward rates
  - yields to maturity of zeros (spot rates)
  - prices of zeros
- prices from forward rates:

$$P(1) = \frac{1000}{\exp(r_1)},$$

$$P(2) = \frac{1000}{\exp(r_1) \exp(r_2)},$$

etc., so that

$$P(n) = \frac{1000}{\exp(r_1 + r_2 + \cdots + r_n)}.$$

- forward rates from prices:

$$\frac{P(n-1)}{P(n)} = \frac{\exp(r_1 + \cdots + r_n)}{\exp(r_1 + \cdots + r_{n-1})} = \exp(r_n)$$

$\Rightarrow$

$$r_n = \log \left\{ \frac{P(n-1)}{P(n)} \right\}$$

- yield to maturity  $y_n$  solves

$$P(n) = \frac{1000}{\exp(ny_n)},$$

$\Rightarrow$

$$y_n = (r_1 + \cdots + r_n)/n.$$

- $\{r_1, \dots, r_n\}$  is easily found from  $\{y_1, \dots, y_n\}$  by:

$$r_1 = y_1,$$

and

$$r_n = ny_n - (n-1)y_{n-1} \quad \text{for } n > 1.$$



## 9.8 References

Jarrow, R., Ruppert, D., and Yu, Yan. (2001), Estimating the term structure of corporate debt with a semiparametric penalized spline model. manuscript.

Available at <http://www.orie.cornell.edu/~davidr/papers/>



# Chapter 10

## Behavioral finance: 5/1/01

### 10.1 Introduction

- behavioral finance is an alternative to the EMH
- this material taken mostly from the 2000 book by Andrei Shleifer of Harvard:
  - **Inefficient Markets: An Introduction To Behavioral Finance**
- EMH has been the central tenet of finance for almost 30 years
- power of the EMH assumption is remarkable
- EMH started in the 1960's
  - immediate success in theory and empirically
  - early empirical work gave overwhelming support to EMH
  - EMH invented at Chicago and Chicago became a world center of research in finance
  - Jensen (1978) “no other proposition in economics . . . has more solid empirical support”
- verdict is changing

- efficiency of arbitrage is much weaker than expected
- true arbitrage possibilities are rare
- near arbitrage is riskier than expected
- “Markets can remain irrational longer than you can remain solvent” — John Maynard Keynes

\* quoted by Roger Lowenstein in **When Genius Failed: The Rise and Fall of Long-Term Capital Management**

## 10.2 Defense of EMH

- three lines of defense of the EMH:
  - investors are rational
  - trading of irrational investors is random and their trades cancel each other
  - even if a “herd” of irrational investors trade in similar ways, rational arbitrageurs will eliminate their influence on market price
- each of these defenses is weaker than had been thought
- rational investing = “value a security by its fundamental value”
  - “fundamental value” = net present worth of all future cash flows
- rational investing  $\Rightarrow$  prices are (geometric) random walks
- but prices being random walks (or nearly so) does **not** imply rational investing
- there is good evidence that irrational trading is correlated
  - look at the internet stock bubble
- initial tests of the semi-strong form of efficiency supported that theory

- event studies showed that the market did react immediately to news and then stopping reacting
  - \* so reaction to news, as EMH predicts
  - \* also no reaction to stale news, again as EMH predicts
- Scholes (1972) found little reaction to “non news”
  - \* block sales had little effect on prices

### 10.3 Challenges to the EMH

- it is difficult to maintain that all investors are rational.
  - many investors react to irrelevant information
  - Black calls them noise traders
- investors act irrationally when they
  - fail to diversify
  - purchase actively and expensively managed mutual funds
  - churn their portfolios
- investors do not look at final levels of wealth when assessing risky situations (“prospect theory”)
- there is a serious “loss aversion”
- people do not follow Bayes rule for evaluating new information
  - too much attention is paid to recent history
- overreaction is commonplace
- these deviations from fully rational behavior are **not random**
- moreover, noise traders will follow each others mistakes
- thus, noise trading will be correlated across investors

- managers of funds are themselves human and will make these errors too
- managers also have their own types of errors
  - buying portfolios excessively close to a benchmark
  - buying the same stocks as other fund managers (so as not to look bad)
  - window dressing — adding stocks to the portfolio that have been performing well recently
  - on average, pension and mutual fund managers underperform passive investment strategies
    - \* these managers might be noise traders too

## 10.4 Can arbitrageurs save the day?

- the last defense of the EMH depends on arbitrage
- even if investor sentiment is correlated and noise traders create incorrectly priced assets
  - arbitrageurs are expected to take the other side of these traders and drive prices back to fundamental values
- a fundamental assumption of behavioral finance is that real-world arbitrage is risky and limited
- arbitrage depends on the existence of “close substitutes” for assets whose prices have been driven to incorrect levels by noise traders
- many securities do not have true substitutes
- often there are no risk-less hedges for arbitrageurs
- mispricing can get even worse, as the managers of LTCM learned
  - this is called noise trader risk

## 10.5 What do the data say?

- Schiller (1981), “Do stock prices move too much to be justified by subsequent changes in dividends”:
  - market prices are too volatile
  - more volatile than explained by a model where prices are expected net present values
  - this work has been criticized by Merton who said that Schiller did not correctly specify fundamental value
- De Bondt and Thaler (1985), “Does the stock market overreact?”:
  - frequently cited and reprinted paper
  - work done at Cornell
  - compare extreme winners and losers
  - find strong evidence of overreaction
  - for every year starting at 1933 they formed portfolios of the best performing stocks over the previous three years
    - \* “winner portfolios”
  - they also formed portfolios of the worse performing stocks
    - \* “loser portfolios”
  - then examined returns on these portfolios over the next five years
    - \* losers consistently outperformed winners
  - difference is difficult to explain as due to differences in risk, at least according to standard models such as CAPM
  - De Bondt and Thaler claim that investors overreact
    - \* extreme losers are too cheap
    - \* so they bounce back
  - the opposite is true of extreme winners

- historically, small stocks have earned higher returns than large stocks
  - no evidence that the difference is due to higher risk
  - superior returns of small stocks have been concentrated in January
  - small firm effect and January effect seem to have disappeared over the last 15 years
- market to book value is a measure of “cheapness”
  - high market to book value firms are “growth” stock
    - \* they tend to underperform
    - \* also they tend to be riskier, especially in severe down markets
- October 19, 1987 — Dow Jones index dropped 22.6%
  - there was no apparent news that day
- Cutler et al (1991): looked at 50 largest one-day market changes
  - many came on days with no major news announcements
- Roll (1988) tried to predict the share of return variation that could be explained by
  - economic influences
  - returns on other stocks in the same industry
  - public firm-specific news
- Roll’s findings:
  - $R^2 = .35$  for monthly data
  - $R^2 = .2$  for daily data
- Roll’s study also shows that there are no “close substitutes” for stocks



- this lack of close substitutes limits arbitrage
- stocks rise if the company is put on the S&P 500 index
  - this is reaction to “non news”
  - America Online rose 18% when included on the S&P
- In summary, there is now considerable evidence against the EMH
  - This evidence was not found during early testing of the EMH
  - Researchers needed to know what to look for

## 10.6 References

Cutler, D., Poterba, J., and Summers, L. (1991). Speculative dynamics, *Review of Economic Studies*, 53, 1839–1885.

De Bondt, W. and Thaler, R. (1985), ‘Does the stock market overreact?’, *J. of Finance*, 40, 793–805.

Jensen, M. (1978), Some anomalous evidence regarding market efficiency, *J. of Financial Economics*, 6, 95–101.

Roll, R. (1988).  $R^2$ , *J. of Finance*, 43, 541–566.

Shiller, R. (1981), Do stock prices move too much to be justified by subsequent changes in dividends, *American Economic Review*, 71, 421–436.

Shleifer, Andrei (2000), *Inefficient Markets: An Introduction to Behavioral Finance*, Oxford University Press, Oxford.