# No-arbitrage Asset Pricing

Alexis Akira Toda*

November 15, 2017

## 1 No-arbitrage asset pricing

Consider an economy with two periods, denoted by $t = 0, 1$. Suppose that at $t = 1$ the state of the economy can be one of $s = 1, \dots, S$. There are $J$ assets in the economy, indexed by $j = 1, \dots, J$. One share of asset $j$ trades for price $q_j$ at time 0 and pays $A_{sj}$ in state $s$. (It can be $A_{sj} < 0$, in which case the holder of one share of asset $j$ must deliver $-A_{sj} > 0$ in state $s$.) Let $q = (q_1, \dots, q_J)$ the vector of asset prices and $A = (A_{sj})$ be the matrix of asset payoffs. Define

$$W = W(q, A) = \begin{bmatrix} -q' \\ A \end{bmatrix}$$

be the $(1 + S) \times J$ matrix of net payments of one share of each asset in each state. Here, state 0 is defined by time 0 and the presence of $-q = (-q_1, \dots, -q_J)$ means that in order to receive $A_{sj}$ in state $s$ one must purchase one share of asset $j$ at time 0, thus paying $q_j$ (receiving $-q_j$).

Let $\theta \in \mathbb{R}^J$ be a *portfolio*. ($\theta_j$ is the number of shares of asset $j$ an investor buys. $\theta_j < 0$ corresponds to shortselling.) The net payments of the portfolio $\theta$ is the vector

$$W\theta = \begin{bmatrix} -q'\theta \\ A\theta \end{bmatrix} \in \mathbb{R}^{1+S}.$$

Here the investor pays $q'\theta$ at $t = 0$ for buying the portfolio $\theta$, and receives $(A\theta)_s$ in state $s$ at $t = 1$.

Let $\langle W \rangle = \left\{ W\theta \,\middle|\, \theta \in \mathbb{R}^J \right\} \subset \mathbb{R}^{1+S}$ be the set of payoffs generated by all portfolios, called the *asset span*. We say that the asset span $\langle W \rangle$ exhibits *no-arbitrage* if

$$\langle W \rangle \cap \mathbb{R}^{1+S}_+ = \{0\}.$$

That is, it is impossible to find a portfolio that pays a non-negative amount in every state and a positive amount in at least one state. Then we can show the following theorem, due to Harrison and Kreps (1979).

**Theorem 1** (Fundamental Theorem of Asset Pricing)**.** *The asset span $\langle W \rangle$ exhibits no-arbitrage if and only if there exists $p \in \mathbb{R}^S_{++}$ such that $[1, p']W = 0$.*

*In this case, the asset prices are given by*

$$q_j = \sum_{s=1}^S p_s A_{sj}.$$

*atoda@ucsd.edu

$p_s > 0$ *is called the* state price *in state* $s$.

*Proof.* Suppose that such a $p$ exists. If $0 \neq w = (w_0, \dots, w_S) \in \mathbb{R}^{1+S}_+$, then

$$[1, p']w = w_0 + \sum_{s=1}^S p_s w_s > 0,$$

so $w \notin \langle W \rangle$. This shows $\langle W \rangle \cap \mathbb{R}^{1+S}_+ = \{0\}$.

Conversely, suppose that there is no arbitrage. Then $\langle W \rangle \cap \Delta = \emptyset$, where $\Delta = \left\{ w \in \mathbb{R}^{1+S}_+ \,\middle|\, \sum_{s=0}^S w_s = 1 \right\}$ is the unit simplex. Clearly $\langle W \rangle, \Delta$ are convex and nonempty, and $\Delta$ is compact. By the (strong version of) separating hyperplane theorem, we can find $0 \neq \lambda \in \mathbb{R}^{1+S}$ such that

$$\langle \lambda, w \rangle < \langle \lambda, d \rangle \tag{1}$$

for any $w \in \langle W \rangle$ and $d \in \Delta$. If there is $\theta$ such that $\langle \lambda, W\theta \rangle \neq 0$, letting $w = W\alpha\theta$ and $\alpha \to \pm\infty$, we will violate (1). Hence $\lambda' W \theta = \langle \lambda, W\theta \rangle = 0$ for any $\theta$, so $\lambda' W = 0$. Then (1) becomes

$$0 < \langle \lambda, d \rangle$$

for all $d \in \Delta$. Letting $d = e_s$ (unit vector) for $s = 0, 1, \dots, S$, we get $\lambda_s > 0$. Letting $p_s = \lambda_s / \lambda_0$ for $s = 1, \dots, S$, the vector $p = (p_1, \dots, p_S)$ satisfies $p \gg 0$ and $[1, p']W = 0$. Writing down this equation component-wise, we get $q_j = \sum_{s=1}^S p_s A_{sj}$. □

Since $p_s > 0$ for all $s$, we have $\sum_{s=1}^S p_s > 0$. Since the risk-free asset pays 1 in every state, its price is

$$\frac{1}{1+r} = \sum_{s=1}^S p_s > 0.$$

Letting $\nu_s = p_s / \sum_s p_s > 0$, we have $\sum_s \nu_s = 1$ and

$$q_j = \frac{1}{1+r} \sum_{s=1}^S \nu_s A_{sj} = \frac{1}{1+r} \tilde{\mathrm{E}}[A_{sj}].$$

Therefore the asset price is the discounted expected payoff of the asset using the *risk-neutral* probability measure $\{\nu_s\}$. This formula is useful for computing option prices in continuous time. For more details, see Duffie (2001).

Letting $\pi_s$ be the objective probability of state $s$ and $m_s = \frac{p_s}{\pi_s}$, we have

$$q_j = \sum_{s=1}^S p_s A_{sj} = \sum_{s=1}^S \pi_s m_s A_{sj} = \mathrm{E}[m A_j].$$

The random variable $m$ is called the *stochastic discount factor*, or SDF for short. Letting $R_j = A_j / q_j$ be the gross return of the asset, we have

$$\mathrm{E}[m R_j] = 1$$

for any asset. The risk-free rate $R_f$ satisfies $\mathrm{E}[mR_f] = 1 \iff R_f = 1/\mathrm{E}[m]$. Using the definition of covariances,

$$\mathrm{Cov}[X, Y] = \mathrm{E}[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])] = \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y],$$

we obtain

$$\begin{aligned} 0 = \mathrm{E}[m(R_j - R_f)] &= \mathrm{E}[m](\mathrm{E}[R_j] - R_f) - \mathrm{Cov}[m, R_j - R_f] \\ \iff \mathrm{E}[R_j] - R_f &= -\frac{1}{\mathrm{E}[m]} \mathrm{Cov}[m, R_j - R_f] \\ &= -R_f \,\mathrm{Cov}[m, R_j], \end{aligned}$$

which is known as the covariance pricing formula.

# 2 Linear factor model

A big issue in empirical asset pricing is what is the stochastic discount factor $m$. Practitioners typically do not care about theory and are more interested in linear factor pricing models. A linear factor model assumes that the SDF takes the form

$$m_t = a - b' f_t,$$

where $f_t$ is a vector of factors and $a, b$ are constants. The most famous example is the classic capital asset pricing model (CAPM), where the single factor is $f_t = R_{mt}$, the market return.

If there are $K$ factors and $b = (b_1, \dots, b_K)'$, then the covariance pricing formula becomes

$$\mathrm{E}[R_j] - R_f = R_f \sum_{k=1}^{K} b_k \,\mathrm{Cov}[f_k, R_j].$$

It is common to define the *beta* of the asset $j$ with respect to factor $k$ by

$$\beta_{k,j} = \frac{\mathrm{Cov}[f_k, R_j]}{\mathrm{Var}[f_k]}.$$

Letting $R_f \,\mathrm{Var}[f_k] = \gamma_k$ be the risk premium of factor $k$ and $\gamma_0 = R_f$ be the zero-beta rate, it follows that

$$\mu_j = \mathrm{E}[R_j] = \gamma_0 + \sum_{k=1}^{K} \gamma_k \beta_{k,j}.$$

A classic methodology to estimate and test a linear factor model is the Fama and MacBeth (1973) two-pass regression. In the first pass, one regress the asset returns $R_{j,t}$ on a constant and the factors to estimate the alpha and beta in

$$R_{j,t} = \alpha_j + \sum_{k=1}^{K} \beta_{k,j} f_{k,t} + \epsilon_{j,t}.$$

If the model is true, then $\alpha_j$ must be common across assets. In the second pass, one regress $\hat{\mu}_j$ (the time series sample mean of $R_{j,t}$) on a constant and the estimated betas to estimate the zero beta rate $\gamma_0$ and the risk premia $\gamma_k$. The $R^2$ from the second pass regression is viewed as a goodness-of-fit of the asset pricing model.

Of course, nowadays it would be more sophisticated to estimate the parameters by GMM exploiting the moment condition

$$\mathrm{E}_t[m_{t+1}R_{j,t+1}] = 1$$

or

$$\mathrm{E}_t[m_{t+1}(R_{j,t+1} - R_{f,t})] = 0.$$

One can test the model by the $J$ test if there are more moment conditions than the number of factors. The empirical literature on linear factor pricing is huge and I am not knowledgeable enough to review here.

The problem with GMM is that the asymptotic distribution of the $J$ statistic is calculated under the null. If the model is incorrect, then the asymptotic distribution would be something different, so the $J$ test tends to have low power (*i.e.*, tends to underreject the null when the null is untrue). Since all economic models are merely an approximation of reality (and hence necessarily wrong), this is a big problem.

In fact, there is a recent literature that shows that applying standard GMM may lead to spurious results. For example, suppose that the model $m_t = a - b'f_t$ is approximately true, but one includes a spurious factor $g_t$, assumed to be independent of all asset returns. Then the moment condition is

$$\mathrm{E}_t[(a - b'f_t - cg_t)(R_{j,t} - R_{f,t})] = 0.$$

This condition holds trivially by setting $b = 0$, $c = 1$, and $a = \mathrm{E}[g_t]$, since by assumption $g_t$ is independent of returns. Therefore if one estimates the model parameters by GMM but the estimating equations contains a spurious factor, then the coefficient on the spurious factor will be significant, the coefficients on the useful factors will be insignificant, and the model will fit perfectly. Such cases are discussed in Kan and Zhang (1999a,b) and Burnside (2016). Since most macroeconomic factors such as consumption growth are weakly correlated with asset returns, the GMM estimation of such models may lead to spurious results.

Kan et al. (2013) and Gospodinov et al. (2014) develop the asymptotic theory of the Fama-MacBeth two-pass regression and the GMM of linear factor models under possible model misspecification. (I am sure there is new development in this field—check the website of Gospodinov.)

# 3 Binomial option pricing

As an application of the no-arbitrage asset pricing, in this section I explain the binomial option pricing model of Cox et al. (1979).

Consider a $T$ period economy, and time is indexed by $t = 0, 1, \dots, T$. Suppose that there are two assets, a stock and a bond. The gross risk-free rate is constant at $R$, and the stock price at time $t$ is denoted by $S_t$, which is a random variable. Assume that the stock can go up or down, so

$$S_{t+1} = \begin{cases} US_t, & \text{(if stock goes up)} \\ DS_t, & \text{(if stock goes down)} \end{cases}$$

where $U > R > D$. Question: what is the price of a call option with strike $K$?

This question seems hopeless to answer since we have not even specified the probabilities of up and down. It turns out that the answer does not depend on the probability, so an optimist and a pessimist will still agree on the price of the option.

Recall that a call (put) option with strike price $K$ and maturity $T$ is a contract such that the holder has the right (but not the obligation) to buy (sell) the stock at price $K$ until the maturity. The act of buying/selling the stock at the specified price is called *exercising*. If the investor can exercise the option at any time on or before maturity, it is called *American*. If the option can be exercised only at maturity, it is called *European*. For more details see textbooks such as Shreve (2004).

## 3.1 European options

Let us compute the price of a European call option. First, consider the simplest case where there is no time, so $T = 0$. Let $C$ be the call price. If the investor exercises the option, he gets $S_0 - K$ by buying the stock at strike price $K$ and selling at the market value $S_0$. If the investor does not exercise the option, it expires, and he gets 0. A rational investor will choose the better alternative, so

$$C = \max\{S_0 - K, 0\}.$$

Next, consider the case with one period to go. If the stock price goes up at $t = 1$, by the above argument the option price becomes $C_u = \max\{US_0 - K, 0\}$. Similarly, in the down state at $t = 1$, the option price is $C_d = \max\{DS_0 - K, 0\}$. Letting $p_s$ be the state price of state $s = u, d$, by no-arbitrage we have

$$C = p_u C_u + p_d C_d.$$

Therefore it remains to compute $p_u, p_d$. To this end we use the no-arbitrage condition for the stock and bond. Since the stock price is $S_0$ at $t = 0$, and it is $US_0$ in the up state and $DS_0$ in the down state, we have

$$S_0 = p_u US_0 + p_d DS_0 \iff 1 = p_u U + p_d D.$$

Since the risk-free asset pays $R$ in all states for one unit of money invested, we have

$$1 = p_u R + p_d R.$$

Solving the system of two linear equations in two unknowns, we get

$$\begin{bmatrix} p_u \\ p_d \end{bmatrix} = \frac{1}{R} \begin{bmatrix} p \\ 1-p \end{bmatrix},$$

where $p = \frac{R-D}{U-D}$. Therefore the call price is

$$C = \frac{1}{R}(pC_u + (1-p)C_d) = \frac{1}{1+r}\tilde{\mathrm{E}}[C_s],$$

where $r$ is the net risk-free rate and $\tilde{\mathrm{E}}$ denotes the expectation under the risk-neutral probability $p$.

The general case is completely analogous. If there are $T$ periods to go, payoffs must be discounted by $R^T$. Since there are two states (up or down) following any state, the the risk-neutral probability is $(p, 1-p)$ each, the risk-neutral probability at $T$ is a binomial distribution with probability $p$. The probability that there are $n$ up states is $\binom{T}{n} p^n (1-p)^{T-n}$, and in this case the final stock price is $S_T = U^n D^{T-n}$. Thus the price of a European call option must be

$$C = \frac{1}{R^T} \sum_{n=0}^{T} \binom{T}{n} p^n (1-p)^{T-n} \max \left\{ U^n D^{T-n} S_0 - K, 0 \right\}.$$

The pricing of European put option is also analogous. Recalling that the payoff of a put when the stock price is $S$ and the strike is $K$ is $P = \max\{K - S, 0\}$, by the same argument the price of a European put is

$$P = \frac{1}{R^T} \sum_{n=0}^{T} \binom{T}{n} p^n (1-p)^{T-n} \max \left\{ K - U^n D^{T-n} S_0, 0 \right\}.$$

An important property of the European options is the put-call parity:

$$C - P = S_0 - KR^{-T}$$

always. Therefore if we know the call price, we can compute the put price by $P = C - S_0 + KR^{-T}$, so we do not need to repeat the calculation. To prove the put-call parity, note that the payoff of a call is $\max\{S_T - K, 0\}$, and that of the put is $\max\{K - S_T, 0\}$. But since

$$\begin{aligned} \max\{S_T - K, 0\} - \max\{K - S_T, 0\} &= \max\{S_T - K, 0\} + \min\{S - K_T, 0\} \\ &= S_T - K, \end{aligned}$$

if someone buys one call and short one put, the terminal payoff is equal to that of holding the stock and paying $K$ at the terminal date. The present value of this portfolio is exactly $S_0 - KR^{-T}$, so the put-call parity holds.

## 3.2 American options

Next, consider the pricing of American options. Since American options can be exercised early, the price of an American option must be at least that of a European option, which cannot be exercised early. We can still use the same idea to price American options.

If $T = 0$, American and European options are identical because the current date is the maturity date.

If $T = 1$, the investor must choose whether to exercise the option at $t = 0$. If he exercises, he gets $S_0 - K$. If he does not exercise, it is the same as holding a European option. Therefore the price of an American option must be

$$C = \max \left\{ S_0 - K, \frac{1}{R}(pC_u + (1-p)C_d) \right\}.$$

In general, letting $C(S, T)$ be the price of an American option when the stock price is $S$ and maturity is $T$ (I suppress the dependency on $K, R$), then we have

$$C(S, T) = \max \left\{ S - K, \frac{1}{R}(pC(US, T-1) + (1-p)C(DS, T-1)) \right\}. \tag{2}$$

The following proposition, due to Merton (1973), says that the price of American and European options coincide if $R \geq 1$.

**Proposition 2.** *Suppose that $R \geq 1$. Then it is never optimal to exercise the call option prematurely. Consequently, the American and European calls have identical prices.*

*Proof.* Let us show that $C(S,T)$ is convex in $S$. We show by induction. If $T=0$, then $C(S,0)=\max\{S-K,0\}$ is clearly convex in $S$. Suppose that $C(S,T-1)$ is convex. Since $S \mapsto US$ is linear, $C(US,T-1)$ is convex. Similarly, $C(DS,T-1)$ is convex. Since the (positively) weighted sum of convex functions is convex,

$$\frac{1}{R}(pC(US,T-1)+(1-p)C(DS,T-1))$$

is convex. Since the maximum of two convex functions is convex, $C(S,T)$ in (2) is convex.

Using the convexity of $C$, if $R \geq 1$, it follows that

$$\begin{aligned}
&\frac{1}{R}(pC(US,T-1)+(1-p)C(DS,T-1)) \\
&\geq \frac{1}{R}C((pU+(1-p)D)S,T-1) \\
&= \frac{1}{R}C(RS,T-1) \\
&\geq \frac{1}{R}\max\{RS-K,0\} \\
&= \max\{S-K/R,0\} \\
&\geq S-K/R \geq S-K,
\end{aligned}$$

where the last inequality is due to $R \geq 1$. Therefore the first term in the right-hand side of (2) is always less than the second term, so rational agents will not exercise the option prematurely. $\square$

Since the payoff of a put option is also convex, you might guess that the above proof also works for put options by changing $S-K$ to $K-S$. However, if $R>1$, the very last inequality does not hold. (It will hold if $R \leq 1$, so in this case an early exercise of put options is not optimal.) In general, American puts have higher value than European puts, and premature exercise may be optimal. Numerically computing the American put price is straightforward by using the binomial tree and iterating (2) with $C(S,T)$ and $S-K$ replaced by $P(S,T)$ and $K-S$, respectively.

## 3.3 Continuous-time limit and Black-Scholes formula

By taking a clever continuous-time limit, Cox et al. (1979) derives the celebrated Black and Scholes (1973) option pricing formula. I omit the details, but let us derive the formula directly using no-arbitrage pricing.

Using the risk-neutral pricing formula, the call option price must be

$$C = \mathrm{e}^{-rT}\tilde{\mathrm{E}}[\max\{S_T-K,0\}],$$

where $r$ is the risk-free rate, $T$ is maturity, $S_T$ is the stock price at maturity, $K$ is the strike price, and $\tilde{\mathrm{E}}$ is the expectation under the risk-neutral probability.

Suppose that the stock price $S_t$ obeys a geometric Brownian motion, so

$$\mathrm{d}S_t = \mu S_t \,\mathrm{d}t + \sigma S_t \,\mathrm{d}B_t,$$

where $\mu$ is the expected return and $\sigma > 0$ is volatility. Think of the risk-neutral probability as the subjective probability of a risk-neutral agent. From this agent's perspective, the stock price must follow the geometric Brownian motion

$$\mathrm{d}S_t = r S_t \,\mathrm{d}t + \sigma S_t \,\mathrm{d}B_t,$$

for otherwise there will be an arbitrage opportunity.

Using Itô's formula (I will come back to that later), we obtain

$$\mathrm{d}(\log S_t) = \left(r - \frac{\sigma^2}{2}\right)\mathrm{d}t + \sigma \,\mathrm{d}B_t,$$

so

$$\log S_t \sim N((r - \sigma^2/2)t, \sigma^2 t).$$

Therefore we can write the logarithm of the terminal stock price as

$$\log S_T = (r - \sigma^2/2)T + \sigma\sqrt{T}z,$$

where $z \sim N(0,1)$. The call price is then

$$\begin{aligned}
C &= \mathrm{e}^{-rT}\int_{-\infty}^{\infty} \max\left\{S_0 \mathrm{e}^{(r-\sigma^2/2)T+\sigma\sqrt{T}z} - K, 0\right\} \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-z^2/2}\,\mathrm{d}z \\
&= \int_{-\infty}^{\infty} \max\left\{S_0 \mathrm{e}^{-\sigma^2 T/2+\sigma\sqrt{T}z} - K\mathrm{e}^{-rT}, 0\right\} \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-z^2/2}\,\mathrm{d}z.
\end{aligned}$$

The expression inside the max is strictly increasing in $z$, and it is zero if and only if

$$S_0 \mathrm{e}^{-\sigma^2 T/2+\sigma\sqrt{T}z} - K\mathrm{e}^{-rT} = 0 \iff z = \frac{\sigma\sqrt{T}}{2} - \frac{1}{\sigma\sqrt{T}}\log\left(\frac{S_0}{K\mathrm{e}^{-rT}}\right).$$

Call this $z$ as $\bar{z}$. Letting $\Phi$ be the cumulative distribution function of the standard normal, it follows that

$$C = \int_{\bar{z}}^{\infty} \left(S_0 \mathrm{e}^{-\sigma^2 T/2+\sigma\sqrt{T}z} - K\mathrm{e}^{-rT}\right) \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-z^2/2}\,\mathrm{d}z.$$

The second term (omitting the minus sign) is

$$K\mathrm{e}^{-rT}(1 - \Phi(\bar{z})) = K\mathrm{e}^{-rT}\Phi(-\bar{z}).$$

The first term is

$$\begin{aligned}
&\int_{\bar{z}}^{\infty} S_0 \mathrm{e}^{-\sigma^2 T/2+\sigma\sqrt{T}z} \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-z^2/2}\,\mathrm{d}z \\
&= S_0 \int_{\bar{z}}^{\infty} \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-(z-\sigma\sqrt{T})^2/2}\,\mathrm{d}z \\
&= S_0 \int_{\bar{z}-\sigma\sqrt{T}}^{\infty} \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-y^2/2}\,\mathrm{d}y \\
&= S_0(1 - \Phi(\bar{z} - \sigma\sqrt{T})) = S_0\Phi(-\bar{z} + \sigma\sqrt{T}).
\end{aligned}$$

Putting all the pieces together, we obtain

$$\begin{aligned} C &= S_0\Phi(-\bar{z}+\sigma\sqrt{T}) - K\mathrm{e}^{-rT}\Phi(-\bar{z}) \\ &= S_0\Phi(x) - K\mathrm{e}^{-rT}\Phi(x-\sigma\sqrt{T}), \end{aligned}$$

where

$$x = -\bar{z}+\sigma\sqrt{T} = \frac{1}{\sigma\sqrt{T}}\log\left(\frac{S_0}{K\mathrm{e}^{-rT}}\right) + \frac{\sigma\sqrt{T}}{2}.$$

This is the Black-Scholes formula.

## Exercises

**1.** 1. Let $\{C_i\}_{i\in I} \subset \mathbb{R}^N$ be a collection of convex sets. Prove that $\bigcap_{i\in I} C_i$ is convex.

2. Let $A$ be any set. Prove that there exists a smallest convex set that includes $A$ (convex hull of $A$).

**2.** Prove Lemma 3.

**3.** 1. Let $0 \neq a \in \mathbb{R}^N$ and $c \in \mathbb{R}$. Show that the hyperplane

$$H = \left\{x \in \mathbb{R}^N \,\middle|\, \langle a, x\rangle = c\right\}$$

and the half space

$$H^+ = \left\{x \in \mathbb{R}^N \,\middle|\, \langle a, x\rangle \geq c\right\}$$

are convex sets.

2. Let $A$ be an $M \times N$ matrix and $b \in \mathbb{R}^M$. The set of the form

$$P = \left\{x \in \mathbb{R}^N \,\middle|\, Ax \leq b\right\}$$

is called a *polytope*. Show that a polytope is convex.

**4.** 1. Let $a, b \in \mathbb{R}^N$. Prove the following *parallelogram law*:

$$\|a+b\|^2 + \|a-b\|^2 = 2\|a\|^2 + 2\|b\|^2.$$

2. Using the parallelogram law, prove (3).

**5.** Let $A = \left\{(x,y) \in \mathbb{R}^2 \,\middle|\, y > x^3\right\}$ and $B = \left\{(x,y) \in \mathbb{R}^2 \,\middle|\, x \geq 1, y \leq 1\right\}$.

1. Draw a picture of the sets $A, B$ on the $xy$ plane.

2. Can $A, B$ be separated? If so, provide an equation of a straight line that separates them. If not, explain why.

**6.** Let $C = \left\{(x,y) \in \mathbb{R}^2 \,\middle|\, y > \mathrm{e}^x\right\}$ and $D = \left\{(x,y) \in \mathbb{R}^2 \,\middle|\, y \leq 0\right\}$.

1. Draw a picture of the sets $C, D$ on the $xy$ plane.

2. Provide an equation of a straight line that separates $C, D$.

3. Can $C, D$ be strictly separated? Answer yes or no, then explain why.

**7.** Prove the following Stiemke's Lemma. Let $A$ be an $M \times N$ matrix. Then one and only one of the following statements is true:

1. There exists $x \in \mathbb{R}^N_{++}$ such that $Ax = 0$.
2. There exists $y \in \mathbb{R}^M$ such that $A'y > 0$.

**8.** A typical linear programming problem is

$$\begin{array}{ll} \text{minimize} & \langle c, x \rangle \\ \text{subject to} & Ax \geq b, \end{array}$$

where $x \in \mathbb{R}^N$, $0 \neq c \in \mathbb{R}^N$, $b \in \mathbb{R}^M$, and $A$ is an $M \times N$ matrix with $M \geq N$. A standard algorithm for solving a linear programming problem is the *simplex method*, which you should have already learned. The idea is that you keep moving from one vertex of the polytope

$$P = \left\{ x \in \mathbb{R}^N \,\middle|\, Ax \geq b \right\}$$

to a neighboring vertex as long as the function value decreases, and if there are no neighboring vertex with smaller function value, you stop.

1. Prove that the simplex method terminates in finite steps.
2. Prove that when the algorithm stops, you are at a solution of the original problem.

# A Convex sets

A set $C \subset \mathbb{R}^N$ is said to be *convex* if the line segment generated by any two points in $C$ is entirely contained in $C$. Formally, $C$ is convex if $x, y \in C$ implies $(1-\alpha)x + \alpha y \in C$ for all $\alpha \in [0, 1]$ (Figure 1). So a circle, triangle, and square are convex but a star-shape is not (Figure 2). One of my favorite mathematical jokes is that the Chinese character for "convex" is not convex (Figure 3).

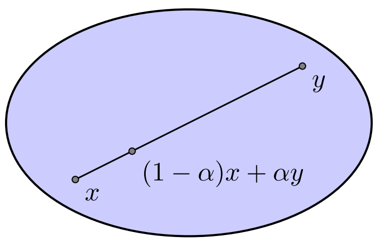


**Figure 1.** Definition of a convex set.

Let $A$ be any set. The smallest convex set that includes $A$ is called the *convex hull* of $A$ and is denoted by $\operatorname{co} A$. (Its existence is left as an exercise.) For example, in Figure 4, the convex hull of the set $A$ consisting of two circles is the entire region in between.

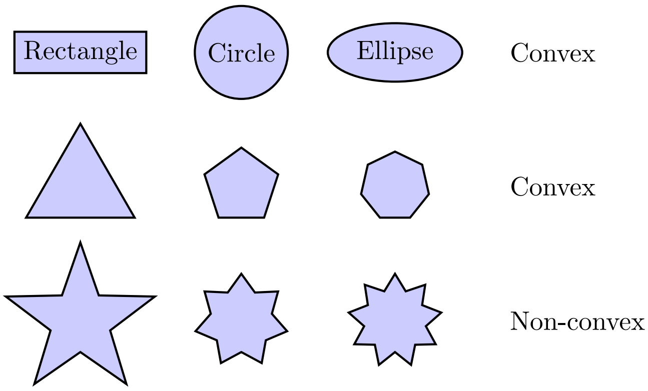


**Figure 2.** Examples of convex and non-convex sets.

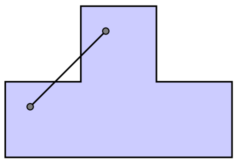

**Figure 3.** Chinese character for "convex" is not convex.

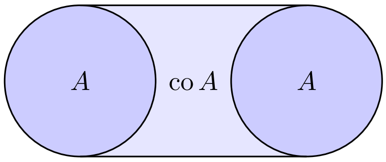


**Figure 4.** Convex hull.

Let $\{x_k\}_{k=1}^K \subset \mathbb{R}^N$ be any points. A point of the form

$$x = \sum_{k=1}^K \alpha_k x_k,$$

where $\alpha_k \geq 0$ and $\sum_{k=1}^K \alpha_k = 1$, is called a *convex combination* of the points $\{x_k\}_{k=1}^K$. The following lemma provides a constructive way to obtain the convex hull of a set.

**Lemma 3.** *Let $A \subset \mathbb{R}^N$ be any set. Then $\mathrm{co}\, A$ consists of all convex combinations of points of $A$.*

*Proof.* Exercise. $\square$

## B Hyperplanes and half spaces

You should know from high school that the equation of a line in $\mathbb{R}^2$ is

$$a_1 x_1 + a_2 x_2 = c$$

for some real numbers $a_1, a_2, c$, and that the equation of a plane in $\mathbb{R}^3$ is

$$a_1 x_1 + a_2 x_2 + a_3 x_3 = c.$$

Letting $a = (a_1, \dots, a_N)$ and $x = (x_1, \dots, x_N)$ be vectors in $\mathbb{R}^N$, the equation $\langle a, x\rangle = c$ is a line if $N = 2$ and a plane if $N = 3$, where

$$\langle a, x\rangle = a_1 x_1 + \cdots + a_N x_N$$

is the inner product of the vectors $a$ and $x$.[1] In general, we say that the set

$$\left\{x \in \mathbb{R}^N \,\middle|\, \langle a, x\rangle = c\right\}$$

is a *hyperplane* if $a \neq 0$. The vector $a$ is orthogonal to this hyperplane (is a *normal vector*). To see this, let $x_0$ be a point in the hyperplane. Since $\langle a, x_0\rangle = c$, by subtraction and linearity of inner product we get $\langle a, x - x_0\rangle = 0$. This means that the vector $a$ is orthogonal to the vector $x - x_0$, which can point to any direction in the plane by moving $x$. So it makes sense to say that $a$ is orthogonal to the hyperplane $\langle a, x\rangle = c$. The sets

$$\begin{aligned} H^+ &= \left\{x \in \mathbb{R}^N \,\middle|\, \langle a, x\rangle \geq c\right\}, \\ H^- &= \left\{x \in \mathbb{R}^N \,\middle|\, \langle a, x\rangle \leq c\right\} \end{aligned}$$

are called *half spaces*, since $H^+$ ($H^-$) is the portion of $\mathbb{R}^N$ separated by the hyperplane $\langle a, x\rangle = c$ towards the direction of $a$ ($-a$). Hyperplanes and half spaces are convex sets (exercise).

# C Separation of convex sets

Let $A, B$ be two sets. We say that the hyperplane $\langle a, x\rangle = c$ *separates* $A, B$ if $A \subset H^-$ and $B \subset H^+$ (Figure 5), that is,

$$\begin{aligned} x \in A &\implies \langle a, x\rangle \leq c, \\ x \in B &\implies \langle a, x\rangle \geq c. \end{aligned}$$

(The inequalities may be reversed.)

Clearly $A, B$ can be separated if and only if

$$\sup_{x \in A} \langle a, x\rangle \leq \inf_{x \in B} \langle a, x\rangle,$$

since we can take $c$ between these two numbers. We say that $A, B$ can be *strictly separated* if the inequality is strict, so

$$\sup_{x \in A} \langle a, x\rangle < \inf_{x \in B} \langle a, x\rangle.$$

The remarkable property of convex sets is the following separation property.

**Theorem 4** (Separating Hyperplane Theorem). *Let $C, D \subset \mathbb{R}^N$ be nonempty and convex. If $C \cap D = \emptyset$, then there exists a hyperplane that separates $C, D$. If $C, D$ are closed and one of them is compact, then they can be strictly separated.*

[1] The inner product is sometimes called the *vector product* or the *dot product*. Common notations for the inner product are $\langle a, x\rangle$, $(a, x)$, $a \cdot x$, etc.

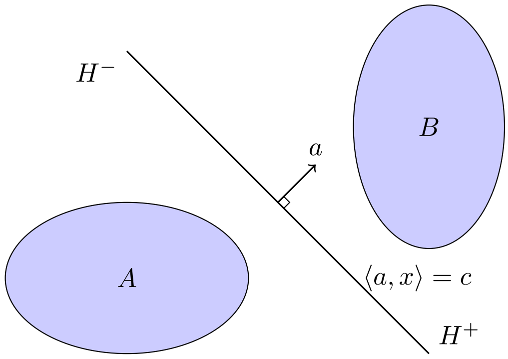


**Figure 5.** Separation of convex sets.

We need the following lemma to prove Theorem 4.

**Lemma 5.** *Let $C$ be nonempty and convex. Then any $x \in \mathbb{R}^N$ has a unique closest point $P_C(x) \in \operatorname{cl} C$, called the* projection of $x$ on $\operatorname{cl} C$. *Furthermore, for any $z \in C$ we have*

$$\langle x - P_C(x), z - P_C(x)\rangle \leq 0.$$

*Proof.* Let $\delta = \inf \{\|x - y\| \mid y \in C\} \geq 0$ be the distance from $x$ to $C$ (Figure 6).

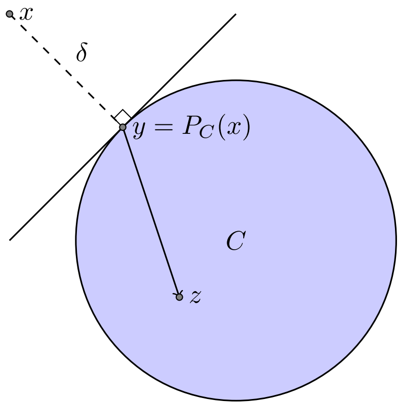


**Figure 6.** Projection on a convex set.

Take a sequence $\{y_k\} \subset C$ such that $\|x - y_k\| \to \delta$. Then by simple algebra we get

$$\|y_k - y_l\|^2 = 2\|x - y_k\|^2 + 2\|x - y_l\|^2 - 4\left\|x - \frac{1}{2}(y_k + y_l)\right\|^2. \tag{3}$$

Since $C$ is convex, we have $\frac{1}{2}(y_k + y_l) \in C$, so by the definition of $\delta$ we get

$$\|y_k - y_l\|^2 \leq 2\|x - y_k\|^2 + 2\|x - y_l\|^2 - 4\delta^2 \to 2\delta^2 + 2\delta^2 - 4\delta^2 = 0$$

as $k, l \to \infty$. Since $\{y_k\} \subset C$ is Cauchy, it converges to some point $y \in \operatorname{cl} C$. Then

$$\|x - y\| \leq \|x - y_k\| + \|y_k - y\| \to \delta + 0 = \delta,$$

so $y$ is the closest point to $x$ in $\operatorname{cl} C$. If $y_1, y_2$ are two closest points, then by the same argument we get

$$\|y_1 - y_2\|^2 \leq 2\|x - y_1\|^2 + 2\|x - y_2\|^2 - 4\delta^2 \leq 0,$$

so $y_1 = y_2$. Thus $y = P_C(x)$ is unique.

Finally, let $z \in C$ be any point. Take $\{y_k\} \subset C$ such that $y_k \to y = P_C(x)$. Since $C$ is convex, for any $0 < \alpha \leq 1$ we have $(1-\alpha)y_k + \alpha z \in C$. Therefore

$$\delta^2 = \|x - y\|^2 \leq \|x - (1-\alpha)y_k - \alpha z\|^2.$$

Letting $k \to \infty$ we get $\|x - y\|^2 \leq \|x - y - \alpha(z - y)\|^2$. Expanding both sides, dividing by $\alpha > 0$, and letting $\alpha \to 0$, we get $\langle x - y, z - y\rangle \leq 0$, which is the desired inequality. $\square$

The following proposition shows that a point that is not an interior point of a convex $C$ can be separated from $C$.

**Proposition 6.** *Let $C \subset \mathbb{R}^N$ be nonempty and convex and $\bar{x} \notin \operatorname{int} C$. Then there exists a hyperplane $\langle a, x\rangle = c$ that separates $\bar{x}$ and $C$,* i.e.*,*

$$\langle a, \bar{x}\rangle \geq c \geq \langle a, z\rangle$$

*for any $z \in C$. If $\bar{x} \notin \operatorname{cl} C$, then the above inequalities can be made strict.*

*Proof.* Suppose that $\bar{x} \notin \operatorname{cl} C$. Let $y = P_C(\bar{x})$ be the projection of $\bar{x}$ on $\operatorname{cl} C$. Then $\bar{x} \neq y$. Let $a = \bar{x} - y \neq 0$ and $c = \langle a, y\rangle + \frac{1}{2}\|a\|^2$. Then for any $z \in C$ we have

$$\begin{aligned}
&\langle \bar{x} - y, z - y\rangle \leq 0 \implies \langle a, z\rangle \leq \langle a, y\rangle < \langle a, y\rangle + \frac{1}{2}\|a\|^2 = c,\\
&\langle a, \bar{x}\rangle - c = \langle \bar{x} - y, \bar{x} - y\rangle - \frac{1}{2}\|a\|^2 = \frac{1}{2}\|a\|^2 > 0 \iff \langle a, \bar{x}\rangle > c.
\end{aligned}$$

Therefore the hyperplane $\langle a, x\rangle = c$ strictly separates $\bar{x}$ and $C$.

If $\bar{x} \in \operatorname{cl} C$, since $\bar{x} \notin \operatorname{int} C$ we can take a sequence $\{x_k\}$ such that $x_k \notin \operatorname{cl} C$ and $x_k \to \bar{x}$. Then we can find a vector $a_k \neq 0$ and a number $c_k \in \mathbb{R}$ such that

$$\langle a_k, x_k\rangle \geq c_k \geq \langle a_k, z\rangle$$

for all $z \in C$. By dividing both sides by $\|a_k\| \neq 0$, without loss of generality we may assume $\|a_k\| = 1$. Since $x_k \to \bar{x}$, the sequence $\{c_k\}$ is bounded. Therefore we can find a convergent subsequence $(a_{k_l}, c_{k_l}) \to (a, c)$. Letting $l \to \infty$, we get

$$\langle a, \bar{x}\rangle \geq c \geq \langle a, z\rangle$$

for any $z \in C$. $\square$

**Proof of Theorem 4.** Let $E = C - D = \{x - y \mid x \in C, y \in D\}$. Since $C, D$ are nonempty and convex, so is $E$. Since $C \cap D = \emptyset$, we have $0 \notin E$. In particular, $0 \notin \operatorname{int} E$. By Proposition 6, there exists $a \neq 0$ such that $\langle a, 0\rangle = 0 \geq \langle a, z\rangle$ for all $z \in E$. By the definition of $E$, we have

$$\langle a, x - y\rangle \leq 0 \iff \langle a, x\rangle \leq \langle a, y\rangle$$

for any $x \in C$ and $y \in D$. Letting $\sup_{x \in C} \langle a, x \rangle \leq c \leq \inf_{y \in D} \langle a, y \rangle$, it follows that the hyperplane $\langle a, x \rangle = c$ separates $C$ and $D$.

Suppose that $C$ is closed and $D$ is compact. Let us show that $E = C - D$ is closed. For this purpose, suppose that $\{z_k\} \subset E$ and $z_k \to z$. Then we can take $\{x_k\} \subset C$, $\{y_k\} \subset D$ such that $z_k = x_k - y_k$. Since $D$ is compact, there is a subsequence such that $y_{k_l} \to y \in D$. Then $x_{k_l} = y_{k_l} + z_{k_l} \to y + z$, but since $C$ is closed, $x = y + z \in C$. Therefore $z = x - y \in E$, so $E$ is closed.

Since $E = C - D$ is closed and $0 \notin E$, by Proposition 6 there exists $a \neq 0$ such that $\langle a, 0 \rangle = 0 > \langle a, z \rangle$ for all $z \in E$. The rest of the proof is similar. $\square$

# References


Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654, May-June 1973. doi:10.1086/260062.

Craig Burnside. Identification and inference in linear stochastic discount factor models with excess returns. *Journal of Financial Econometrics*, 14(2):295–330, 2016. doi:10.1093/jjfinec/nbv018.

John C. Cox, Stephen A. Ross, and Mark Rubinstein. Option pricing: A simplified approach. *Journal of Financial Economics*, 7(3):229–263, September 1979. doi:10.1016/0304-405X(79)90015-1.

Darrell Duffie. *Dynamic Asset Pricing Theory*. Princeton University Press, third edition, 2001.

Eugene F. Fama and James D. MacBeth. Risk, return, and equilibrium: Empirical test. *Journal of Political Economy*, 81(3):607–636, May-June 1973. doi:10.1086/260061.

Nikolay Gospodinov, Raymond Kan, and Cesare Robotti. Misspecification-robust inference in linear asset-pricing models with irrelevant risk factors. *Review of Financial Studies*, 27(7):2139–2170, 2014. doi:10.1093/rfs/hht135.

J. Michael Harrison and David M. Kreps. Martingales and arbitrage in multiperiod securities market. *Journal of Economic Theory*, 20(3):381–408, June 1979. doi:10.1016/0022-0531(79)90043-7.

Raymond Kan and Chu Zhang. Two-pass tests of asset pricing models with useless factors. *Journal of Finance*, 54(1):203–235, February 1999a. doi:10.1111/0022-1082.00102.

Raymond Kan and Chu Zhang. GMM tests of stochastic discount factor models with useless factors. *Journal of Financial Economics*, 54(1):103–127, October 1999b. doi:10.1016/S0304-405X(99)00033-1.

Raymond Kan, Cesare Robotti, and Jay Shanken. Pricing model performance and the two-pass cross-sectional regression methodology. *Journal of Finance*, 68(6):2617–2649, December 2013. doi:10.1111/jofi.12035.

Robert C. Merton. Theory of rational option pricing. *Bell Journal of Economics and Management Science*, 4(1):141–183, Spring 1973. doi:10.2307/3003143.

Steven E. Shreve. *Stochastic Calculus for Finance II—Continuous-Time Models*. Springer Finance. Springer, New York, 2004.