15.456 Financial Engineering

MIT Sloan School of Management
Paul F. Mende

Course overview and introduction to asset pricing
September 4, 2020

Agenda

- Welcome to 15.456 – Financial Engineering
- Announcements
  - Pre-requisites
  - Problem Set 1
- Organization and logistics
- Course overview
- Payoffs, probabilities, and prices
- Appendix: Facts from linear algebra

Course introduction and overview

What Is Financial Engineering?

- Financial challenges of firms, households and governments:
  - What investments to make/how to value assets?
  - How to finance investments?
  - How to manage risk?
  - How to innovate to improve the efficiency for above activities?

- Financial engineering is the application of finance principles and relevant tools to arrive at concrete solutions to the specific problems in meeting these challenges.

This Course

- To cover both well-settled and frontier areas of FE.
- The focus is on applications, methods and solutions.
- These methods, tools and solutions can be useful in expanding the broader and richer applications in other classes:
  - Options and futures (15.437)
  - Fixed income securities and derivatives (15.438)
  - International capital markets (15.447)
  - Functional and strategic finance (15.466)
  - Asset management, life-cycle investing and retirement finance (15.467)
  - Financial Data Science and Computing (15.458)

Course Overview

- H1 (Mende):
  - Arbitrage pricing and applications
  - Continuous-time finance: stochastic calculus and financial modeling

- H2 (Kogan):
  - Financial optimization
  - Market equilibrium in frictionless markets
  - Equilibrium models with frictions

Course Requirements

- Grading
  - Class participation (~10%)
  - Assignments + Quizzes (~25%)
  - Midterm exam (20%) – Friday, October 16, in-class
  - Final exam (45%) – MIT schedule

- Assignments:
  - Problem sets
    - **Individually written** with code attached
    - May consult with current 15.456 classmates and textbooks.
    - **May not use solutions or materials from other or prior classes**
    - Presentation should be clear and professional.
    - Upload high-quality scans (if handwritten) -- no phone photos. Upload PDF if prepared in TeX/LaTeX
  - Quizzes
    - Offline, self-paced
    - In-class, unannounced.

Course Expectations

- Value@Sloan
- Academic Integrity
- Classroom Behavior:
  - Attend all classes and arrive/leave on time. (Inform faculty and TA in advance for special circumstances.)
  - No non-class activities during class meeting.
  - Cameras on. Smile!
  - Microphones on mute.
  - Unmute yourself and speak up with questions and comments
- Some helpful suggestions:
  - Take copious notes during lectures (lecture notes are not complete).
  - Review the lectures afterwards (with your study group or alone).
  - Work on assignments – "*Finance and engineering are not spectator sports.*"
  - Ask questions!

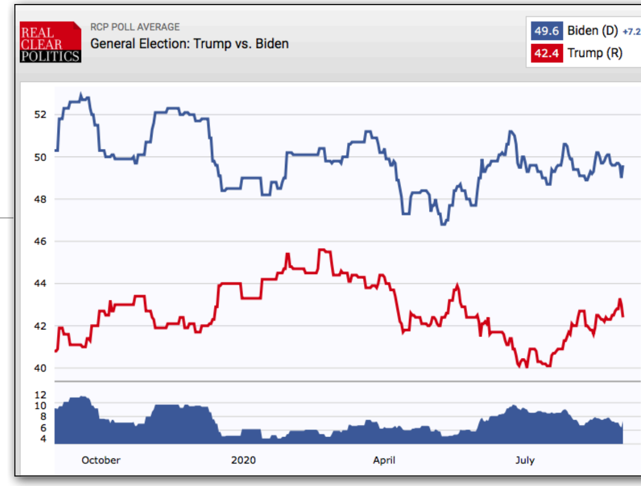
Course Resources

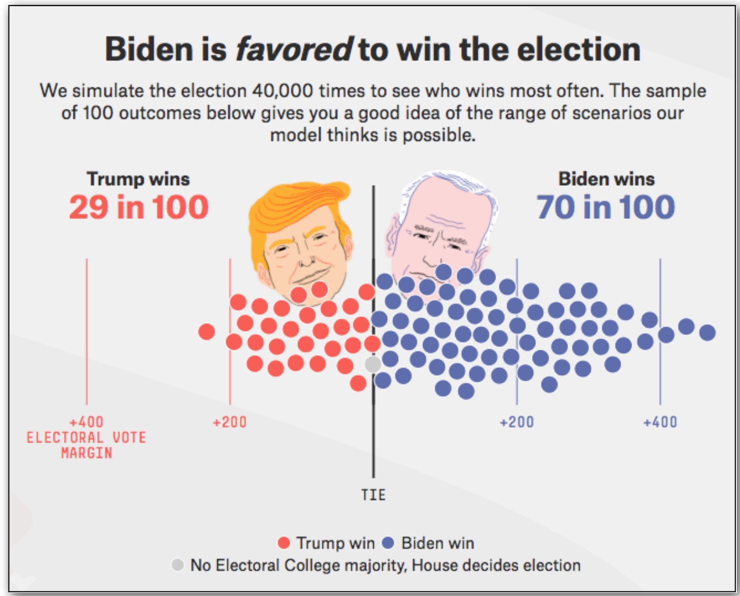
- Instructors
  - Paul Mende (H1) – Office hours Wednesday 11am-12noon ET and by appointment
  - Leonid Kogan (H2)
- TA
  - Ali Kakhbod (H1+H2) – Office hours and recitation TBD
- Canvas
  - All course information will be posted there
  - References: required and optional materials
  - Problem set download and submission
  - No late assignments

Probabilities and payoffs: warmup examples

Probabilities and payoffs

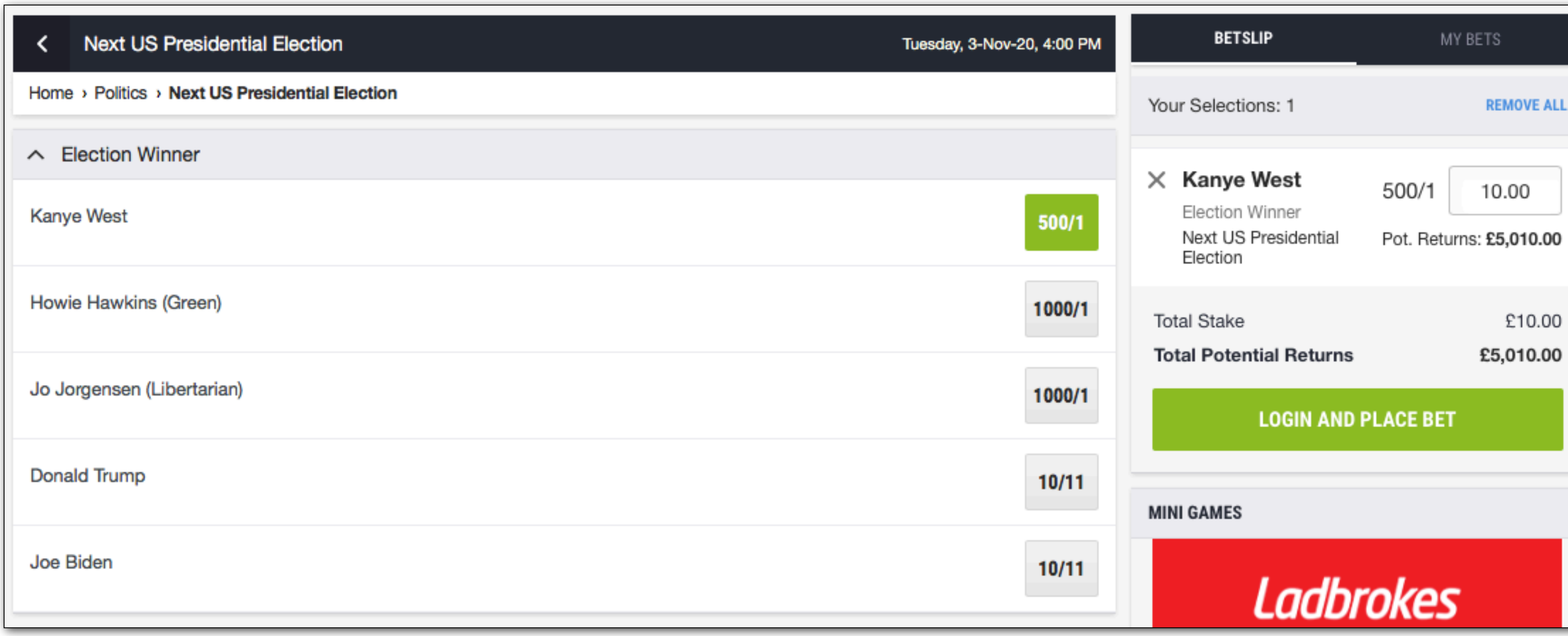
- Modeling from data: polls and surveys
- Probability forecasts
- Do you **contribute** to your favorite candidate?
- Do you **bet** on your favorite candidate?





Source: *fivethirtyeight.com*

Betting on the 2020 U.S. Presidential Election (from the U.K.)



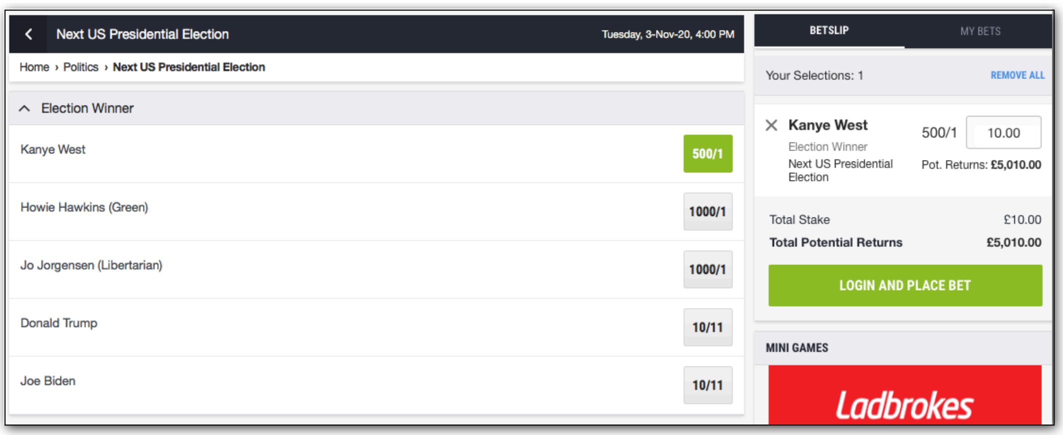
Bookmaking 101

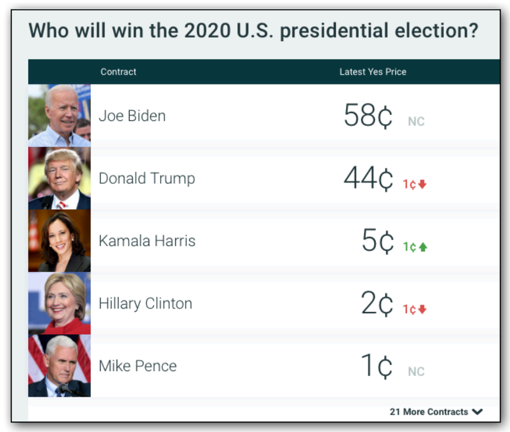
- Setting the odds based on probabilities, the "fair" payout *x* per dollar in each state is

$$x_i = \frac{1}{p_i} - 1$$

- Writing *x* = *n*/*d* in lowest terms, we say that the bet is "*n*-to-*d*."

- So if Kayne West has an 0.2% chance, he's a 500-1 longshot.





Source: *predictit.com*

Bookmaking 101

- Whatever the outcome, the bookmaker pays the winners in proportion to their bets and keeps the money from all the losers. His P/L in each state is

$$\pi_i = -x_i B_i + \sum_{j \neq i} B_j$$

- For example, suppose there are two outcomes
  - Probabilities 90% vs. 10%
  - \$10,000 is bet on the favorite, and \$2,000 is riding on the longshot:

$$\begin{aligned} \pi_1 &= \ +\$889 \\ \pi_2 &= -\$8{,}000 \end{aligned}$$

- Expected value is zero and bets are "fair," but 10% of the time a big loss occurs.

What are the odds?

- Betting was active in the summer of 2013 as Prince William and Catherine, Duchess of Cambridge, waited for the birth of a new heir to the British throne.

- How should bookmakers set their odds and payoffs?

- An unusual real-world case in which the exact odds (boy vs. girl) are known. So why aren't payoffs equal?

Bookmaking 101 (cont.)

- Suppose one **ignored** the "true" probabilities and set payoffs based on sentiment. If demand is 2 to 1, set

$$\hat{x}_1 = B_2/B_1 = 1/2,$$
$$\hat{x}_2 = B_1/B_2 = 2$$

- Now the P/L is zero **in each state**.

- Think of the "implied probabilities" as

$$\hat{p}_1 = 1/(1+\hat{x}_1) = 67\%$$
$$\hat{p}_2 = 1/(1+\hat{x}_2) = 33\%$$

A Tale of Two Bookies

- Bookie A uses "true" probabilities. **Ratio** of payouts is equal.
- Bookie B uses **implied** probabilities. She computes payouts to equalize P/L in both states of the world. The "true" probabilities are disregarded entirely.
- What's missing from this picture?

	Girl	Boy
True odds	50%	50%
Total Wager	**£40,000**	**£20,000**
Bookie A pays	£1.00	£1.00
Bookie B pays	£0.50	£2.00
Bookie A implied	50%	50%
Bookie B implied	67%	33%
Bookie A P/L	-£20,000	£20,000
Bookie B P/L	£0	£0

Bookmaking 101 (cont.)

- Suppose one **ignored** the "true" probabilities and set payoffs based on sentiment. If demand is 2 to 1, set

$$\hat{x}_1 = B_2/B_1 = 1/2,$$
$$\hat{x}_2 = B_1/B_2 = 2$$

- Now the P/L is zero **in each state**.

- Think of the "implied probabilities" as

$$\hat{p}_1 = 1/(1+\hat{x}_1) = 67\%$$
$$\hat{p}_2 = 1/(1+\hat{x}_2) = 33\%$$

- However, we're not running a charity. Modify payouts to allow for an extra profit π in each state:

$$x_i \to x_i - \pi/B_i,$$
$$\hat{x}_i \to \hat{x}_i - \pi/B_i$$

- For the latter case, this is the **minimal variance** solution. P/L is bounded above.

The truth is ... not always relevant to the best policy

- What is the best strategy for the bookmaker?
- If demand is static, Bookie B will outlive Bookie A. Both have same expected return.
- Can divergence between sentiment and true probabilities persist?
- What is the best opportunity for the bettor?

	Girl	Boy
True odds	50%	50%
Total Wager	**£40,000**	**£20,000**
Bookie A pays	£0.90	£0.90
Bookie B pays	£0.425	£1.85
Bookie A implied	52.6%	52.6%
Bookie B implied	70.2%	35.1%
Bookie A P/L	-£16,000	£22,000
Bookie B P/L	£3,000	£3,000

Arbitrage bounds

- The payouts must not permit **arbitrage**, which would occur if a bettor (not the bookmaker) can make bets that never lose and can earn a riskless profit.

- Bookie B fixed odds based on equilibrium sentiment plus a fixed profit.

$$\hat{x}_1 = B_2/B_1 - \pi/B_1,$$
$$\hat{x}_2 = B_1/B_2 - \pi/B_2$$

- These payoffs obey

$$\hat{x}_1 \hat{x}_2 \leq 1 \quad \implies \quad \hat{p}_1 + \hat{p}_2 \geq 1$$

- A gambler offered these odds might wager amounts *w* in a **different ratio**. We cannot have both

$$w_1 \hat{x}_1 - w_2 > 0,$$
$$w_2 \hat{x}_2 - w_1 > 0$$

Modeling unique events

- Since this is a one-time event, how do we know which forecasts were best?
- Given the outcome, how would you change the model next time?

Throw Out the Probability Models

Nassim Nicholas Taleb, a former derivatives trader, is a distinguished professor of risk engineering at Polytechnic Institute of New York University. He is the author of "The Black Swan: The Impact of the Highly Improbable."

UPDATED APRIL 2, 2012, 4:44 PM

After the events that started in 2007 and the subsequent reactions by economists, anyone who takes the current economics establishment seriously needs to spend time in a sanatorium.

This does not mean we should write off the entire body of knowledge. By now, we can see what works and what does not work. Simply, a certain class of consequential rare events, what I've called "black swans," are not predictable and their probabilities unmeasurable, so anything that relies on a computation of the probability of these events should go out of the window. Now. Such models induce fragilities and bring harm. We're better off with no model than with a defective model, something people understand intuitively, but they tend to forget when they don't have "skin in the game." If you are a passenger on a plane and the pilot tells you he has a faulty map, you get off the plane; you don't stay and say "well, there is nothing better." But in economics, particularly finance, they keep teaching these models on grounds that "there is nothing better," causing harmful risk-taking. Why? Because the professors don't bear the harm of the models.

We would have great jumps in knowledge if we avoided teaching these models, and replaced them with anything, even gardening classes.

Models in finance

- Modeling financial markets
  - Remove extraneous features
  - Focus on core ideas, drivers, causes
  - Choose mathematical forms that are tractable analytically or computationally
  - Choose estimation procedures consistent with observable data
- Strictly speaking, the mathematical hypotheses required for analytical results are almost always violated.
- When does it matter?

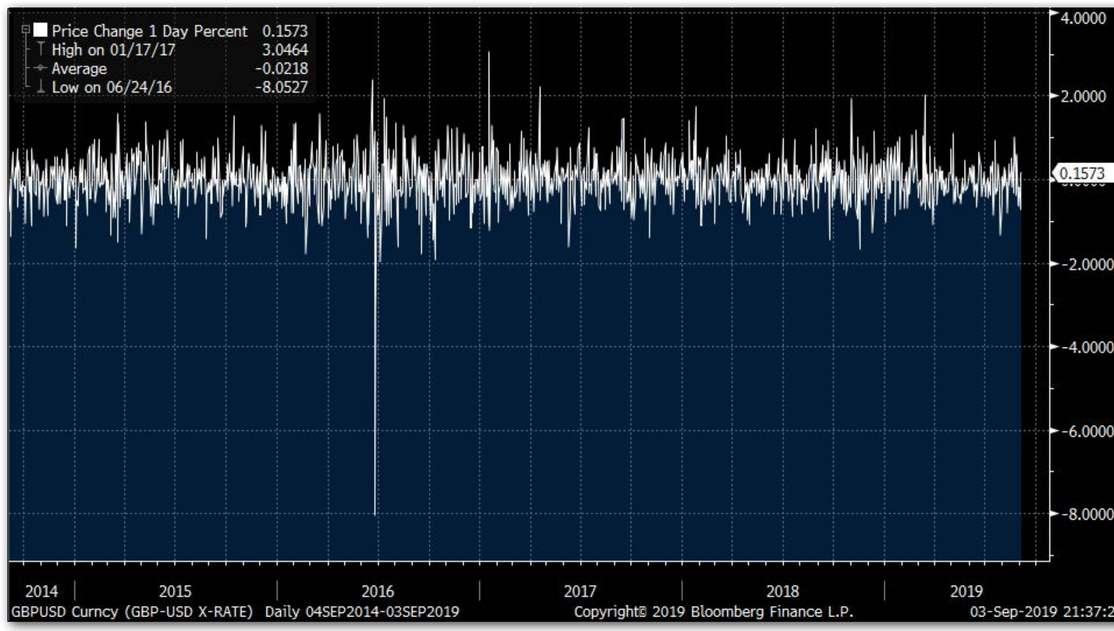
Source: Wikimedia Commons

Models in finance

- Good models
  - Reproduce observed behaviors qualitatively and quantitatively
  - Correspond to an idealized limit
  - Have well-defined range of validity
  - Allow for corrections, extensions, and approximations

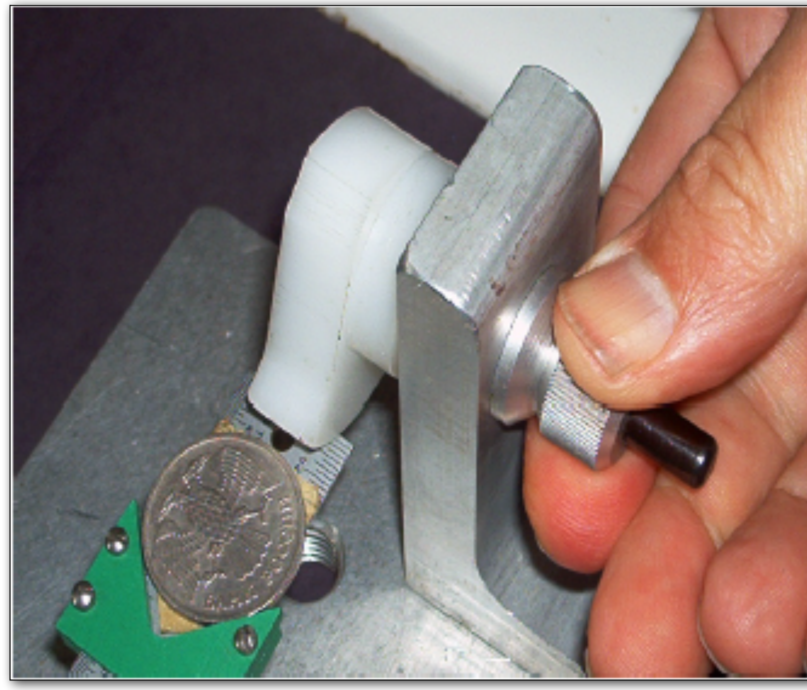
- *"Toy models"*
  - Only capture qualitative features and perhaps semi-quantitative behavior
  - Cannot be extended to meaningfully match real-world behaviors

$$\boxed{\sigma_t^2 = \omega + \alpha\xi_{t-1}^2 + \beta\sigma_{t-1}^2}$$



Models in finance

- Model breakdown
  - Knowing how models fail is just as important as knowing where and when they are no longer valid
  - Required model assumptions may stop being applicable or reasonable
  - Where one model fails, a new one can take over



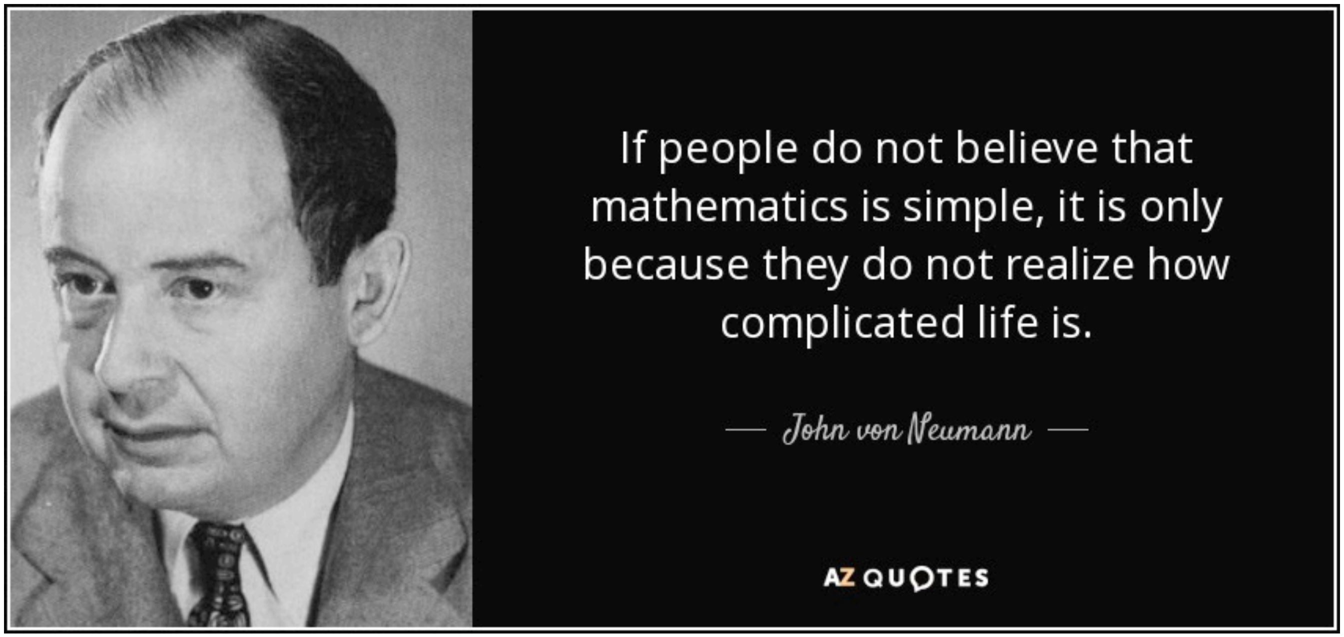
Source: Diaconis, Holmes, and Montgomery

Models in finance

- Model breakdown
  - Knowing how models fail is just as important as knowing where and when they are no longer valid
  - Required model assumptions may stop being applicable or reasonable
  - Where one model fails, a new one can take over

- Example:
  - Standard log-normal model of asset prices **require positive prices.**
  - What should be done with models if prices go **negative**?

Source: Bloomberg, 4/21/2020


If people do not believe that mathematics is simple, it is only because they do not realize how complicated life is.
— John von Neumann —
AZ QUOTES

Payoffs, probabilities, and prices

Simplest model of financial markets

- State space: discrete and finite
- Time: discrete and two-period
- Securities
  - Contracts to receive a future payoff
  - Price now is known
  - Future payoff is uncertain
- Example:
  - Probabilities and payoffs for 4 securities
  - Stock, Bond, Calls (*K=1.5, K=1)*
  - What is left out?

State	Probability	Bond	Stock	Call #1	Call #2
I	1/2	$1	$3	$1.5	$2
II	1/6	$1	$2	$0.5	$1
III	1/3	$1	$1	$0	$0

The payoff matrix

- Let there be *n* different securities and *s* states of the world. Then represent the payoffs for each security as a column vector of length *s*, and collect them to form the **payoff matrix**

$$A : \mathbb{R}^n \to \mathbb{R}^s$$

- In our example,

$$A = \begin{pmatrix} 1 & 3 & 1.5 & 2 \\ 1 & 2 & 0.5 & 1 \\ 1 & 1 & 0 & 0 \end{pmatrix}$$

- A **portfolio** is represented as a vector of quantities held for each security

The payoff matrix

- A **portfolio** is represented as a vector of quantities held for each security.

- The **portfolio payoff** is determined by action on the portfolio vector with the payoff matrix. For example if

$$A = \begin{pmatrix} 1 & 3 & 1.5 & 2 \\ 1 & 2 & 0.5 & 1 \\ 1 & 1 & 0 & 0 \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} -1 \\ 2 \\ -2 \\ -1 \end{pmatrix}, \quad A\mathbf{x} = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$$

- So it's simple to find the payoff of any portfolio. What about the reverse?

- What if we had asked: what portfolio ***x*** gives the payoff (0 1 1)'?

The payoff matrix

- In other words, we need to know how to solve the general equation $A\mathbf{x} = \mathbf{b}$

- If *A* is **non-singular**, there is a unique solution, $\mathbf{x} = A^{-1}\mathbf{b}$
  Requires:
  - *A* invertible
  - Columns of *A* are independent. (Rows, too)
  - Securities not redundant
  - #securities = #states, *n* = *s*

- If ***n* < *s*,** then in general there is no solution, except for those $\mathbf{b} \in \mathrm{Image}(A)$

- If ***n* > *s***, then there might or might not be a solution. Solutions are not unique: one can add any element of the kernel, or null space, of *A.*

Market completeness

- A **complete market** is one in which **every** payoff can be generated by **some** portfolio
  - Equivalently: Image(A) = the full space of payoffs
  - Equivalently: rank(A) = *s* = # of independent future states of the world
  - Equivalently: the linear transformation A is "onto," or "surjective."

- An **incomplete market** is one in which **some** payoffs cannot be generated by **any** portfolio.
  - Equivalently: Image(A) = a proper subspace within the space of payoffs
  - Equivalently: rank(A) < s = # independent future states of the world

The payoff matrix

- In still other words, we want to know when it is possible to create a portfolio that matches a given or desired payoff.

- Examples
  - OTC derivatives
  - Business risk management

- One way to solve this business problem
  - Compute the **perfect hedge** (...if exists...and if so, is it unique?)
  - Add a markup (...if customers find this service worth paying for)
  - Correct for deviations from idealized solution

Redundant securities

- In both complete and incomplete markets, there could be **redundant securities**. This occurs when one or more sets of securities have payoffs that are linearly dependent.
  - Equivalently: one or more securities have payoffs that can be **replicated** by a portfolio of other securities.
  - Equivalently: the kernel (or null space) of A is non-empty; dim(ker(A)) > 0
  - Equivalently: there exist "arbitrage portfolios" which are non-trivial portfolios that have **zero** payoff.

- If redundant securities exist, the association of portfolios with payoffs is **not unique**.
  - For a given portfolio, the payoff is unique, $A\mathbf{x} = \mathbf{b}$
  - But for a given payoff, there can be more than one portfolio ***x*** that solves the equation. Take a solution, add any multiple of any "arbitrage portfolio" (i.e., an element of ker(A)), and it will have the **same payoff**. (Of course, in an incomplete market the existence of solution depends on **b**. So there will either be an infinite number of solutions or none at all.)

Replicating payoffs

- Example: can the *K=1* call be replicated from the other 3 securities?

$$A = \begin{pmatrix} 1 & 3 & 1.5 \\ 1 & 2 & 0.5 \\ 1 & 1 & 0 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}$$

$$A^{-1} = \begin{pmatrix} 1 & -3 & 3 \\ -1 & 3 & -2 \\ 2 & -4 & 2 \end{pmatrix}, \quad \text{so } \mathbf{x} = A^{-1}\mathbf{b} = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}$$

- Therefore the *K=1* call is **redundant**: its payoff is identical to that of a portfolio of **basis assets**.

Replicating payoffs

- Example: can the *K=1* call be replicated from **just two** other securities?

$$A = \begin{pmatrix} 1 & 3 \\ 1 & 2 \\ 1 & 1 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}$$

- *A* cannot be invertible since it isn't even square. But there is still a solution,

$$A\mathbf{x} = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} \implies \mathbf{x} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

Replicating payoffs

- Example: can the *K=1.5* call be replicated from the other 3 securities?

$$A = \begin{pmatrix} 1 & 3 & 2 \\ 1 & 2 & 1 \\ 1 & 1 & 0 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 1.5 \\ 0.5 \\ 0 \end{pmatrix}$$

- The **rank** of *A* = 2, so the payoff matrix is singular. An inverse does not exist, and the option cannot be replicated. The market is **incomplete**.

Complete markets

- A **complete market** is one in which any payout can be generated by a portfolio of basis assets.

- In a complete market, the payoff matrix has **rank *s*.**
  - If, in addition, the number of securities $n=s$, then the existence of an inverse means there is a unique solution for any **b**.

- If $n > s$, select a complete basis and drop n-s redundant securities

Note:

$$r(A) = r(A^\mathsf{T}) \leq \min(n, s),$$
$$r(AB) \leq \min(r(A), r(B))$$
$$r(AA^\mathsf{T}) = r(A)$$

Prices

- So far we have talked only about future payoffs for securities.
  - What are their prices now?
  - What is the difference between payoffs and returns?

- Present prices can be represented by a vector ***S*** (or just as often, its transpose)

- Value of a portfolio

$$MV = \sum_{i=1}^{n} S_i x_i = \begin{pmatrix} S_1 & S_2 & \cdots \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \end{pmatrix} = \mathbf{S}^* \mathbf{x} = S[\mathbf{x}]$$

- Note that ***S*** acts **linearly** on portfolio vectors, and that its dimension is also *n.*

Appendix: Facts from Linear Algebra

Linear dependence, basis, and dimension

Linear dependence

- A **linear combination** is a sum of vectors multiplied by arbitrary scalars

$$\mathbf{w} = \sum_i a_i \mathbf{v_i}$$

- A set of vectors is **linearly dependent** if there are constants, not all zero, such that

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n = 0$$

- If no such set of constants exist, the set of vectors is **linearly independent.**

Linear dependence

- Linear dependence means we can write at least one of the vectors in terms of the others.

$$\mathbf{v}_1 = -\frac{1}{a_1}(a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n)$$

(Here is where we use that the scalars must be a ***field*** *so that there is an inverse for non-zero scalars.)*

- So is there a **finite set** of vectors that be used to express all the others as linear combinations?

Spanning sets

- One way is to **define** a vector space so that it's true. We call the set of all linear combinations of a given set of vectors the **span** of that set.

$$\text{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\} = \left\{ \sum_{i=1}^{k} a_i \mathbf{v_i}, \quad \forall a_i \in \mathbb{R} \right\}$$

- Because it is to designed to be closed under addition and scalar multiplication, it forms a vector space.

- If every element of a vector space *V* can be expressed as a linear combination of a given set, then that set is said to **span the vector space**.

Basis

- If, in addition to spanning *V*, the vectors in the spanning set are linearly independent, then they form a **basis** for *V.*
    - A basis is a minimal, independent set of vectors that spans the space.
    - The number of vectors in the basis set is called the **dimension** of the vector space.
    - The choice of basis vectors is **not unique**.
    - Changing the basis, however, does not change the dimension.

Coordinates and notation

- Given a basis, the expression of any vector as a linear combination in terms of the basis vectors is **unique**.

$$\mathbf{v} = c_1\mathbf{u_1} + c_2\mathbf{u_2} + \cdots + c_n\mathbf{u_n}$$

- The coefficients are called the **coordinates with respect to the basis**.

- We use vector notation to denote this linear combination compactly as

$$\mathbf{v} = c_1 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + c_2 \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} + \ldots c_n \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}$$

Subspaces

- A **subspace** S of a vector space V is a vector space which is a subset of V.
- The span of a **subset** of basis vectors of V defines a subspace of V.
- Any linearly independent set of $k < n$ vectors of V defines a subspace.
  - The set can constitute a basis of the subspace.
  - The dimension of the subspace is k.
  - Example: polynomials of degree 3 or less vanishing at $x=1$.

Linear transformations

- Functions and mappings
- Linear functions
- A **linear transformation** on a vector space $T : V \to W$ obeys

$$T(\mathbf{v}_1 + \mathbf{v}_2) = T\mathbf{v}_1 + T\mathbf{v}_2,$$
$$T(c\mathbf{v}) = cT\mathbf{v}.$$

- This simple property of **linearity** means that any linear transformation is completely described by its action on the basis vectors of a space.

Linear transformations

- Consider the transformation *T* acting on an arbitrary vector, which is expressed as a linear combination of basis vectors. Then by linearity,

$$\begin{aligned} T\mathbf{v} &= T\left(c_1\mathbf{u_1} + c_2\mathbf{u_2} + \cdots + c_n\mathbf{u_n}\right) \\ &= c_1(T\mathbf{u_1}) + c_2(T\mathbf{u_2}) + \cdots + c_n(T\mathbf{u_n}) \end{aligned}$$

- Therefore if we know how *T* acts on **each** basis vector in the vector space *V*, we can express the action of *T* on **any** vector by taking linear combinations of these *n* results.

Matrix of a linear transformation

- Let's use column notation for vectors in the target space *W*.

- *T*'s action on each basis vector of *V* gives some vector in *W*, so let's write them in general form as

$$T\mathbf{u_1} = \begin{pmatrix} m_{11} \\ m_{21} \\ \vdots \\ m_{s1} \end{pmatrix}, \quad T\mathbf{u_2} = \begin{pmatrix} m_{12} \\ m_{22} \\ \vdots \\ m_{s2} \end{pmatrix}, \cdots \quad T\mathbf{u_n} = \begin{pmatrix} m_{1n} \\ m_{2n} \\ \vdots \\ m_{sn} \end{pmatrix}$$

- *T* is then characterized by *n* column vectors (the dimension of *V*), each of length *s* (the dimension of *W*).

Matrix of a linear transformation

- Combine these n columns to form the matrix M corresponding to the linear transformation.

$$M = \begin{pmatrix} m_{11} & m_{12} & \cdots & m_{1n} \\ m_{21} & m_{22} & \cdots & m_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ m_{s1} & m_{s2} & \cdots & m_{sn} \end{pmatrix}$$

- The matrix M depends on the choice of bases in V and W.

- When M acts on a column vector of V, the result will be a linear combination of the columns of M.

Matrix of a linear transformation

- In column notation,

$$M\mathbf{x} = \begin{pmatrix} m_{11} & m_{12} & \cdots & m_{1n} \\ m_{21} & m_{22} & \cdots & m_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ m_{s1} & m_{s2} & \cdots & m_{sn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} m_{11}x_1 + m_{12}x_2 + \cdots + m_{1n}x_n \\ m_{21}x_1 + m_{22}x_2 + \cdots + m_{2n}x_n \\ \vdots \\ m_{s1}x_1 + m_{s2}x_2 + \cdots + m_{sn}x_n \end{pmatrix}$$

- In components, this transformation rule reads

$$(M\mathbf{x})_i = \sum_{j=1}^{n} m_{ij}x_j, \qquad i = 1, 2, \ldots, s$$

Linear transformations of the plane

- In two dimensions, let's write this as

$$M\mathbf{x} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} ax + by \\ cx + dy \end{pmatrix}$$

- Acting on this equation from the left with a new linear transformation gives a rule for multiplication of matrices

$$M'M = \begin{pmatrix} a'a + b'c & a'b + b'd \\ c'a + d'c & c'b + d'd \end{pmatrix}$$

- Each column of the product is the result of acting with *M'* on the corresponding column of *M*.

Matrix multiplication

- Matrix multiplication defined by **composition** of linear transformations

$$(M'M)_{ij} = \sum_{k=1}^{s} M'_{ik} M_{kj}$$

- Properties:
  - Associative $$M_1(M_2M_3) = (M_1M_2)M_3$$
  - Distributive $$M_1(M_2 + M_3) = M_1M_2 + M_1M_3$$
  - NOT commutative $$M_1M_2 \neq M_2M_1 \text{ (in general)}$$
  - Identity $$MI = IM = M$$

- (Bonus fact: Matrices also form a vector space of their own.)

Image and kernel

- Two important subspaces can be defined with respect to a linear operator

$$T : V \to W$$

- The **image** of *T* is the set of all vectors in *W* that can be reached from *V*

$$\operatorname{Im} T = \{\mathbf{w} | \exists \mathbf{v} \in V, \quad T\mathbf{v} = \mathbf{w}\} \subset W$$

- The **kernel** of *T* is the set of all vectors in *V* "annihilated" by *T*

$$\operatorname{Ker} T = \{\mathbf{v} \in V | T\mathbf{v} = 0\} \subset V$$

Image and kernel

- The **rank** of a linear transformation is the dimension of the image. It is the number of linearly independent columns of a matrix.

- Fundamental Theorem of Linear Transformations:

$$\dim V = \dim(\mathrm{Im}\, T) + \dim(\mathrm{Ker}\, T)$$

- If the kernel is empty, i.e., has dimension=0, then the rank of the transformation is the dimension of *V.*

- If in addition, *V* and *W* have the same dimension, then *T* is **invertible**.

Some properties of determinants

- A square matrix has an **inverse** if and only if Det *M* is non-zero
- For a 2x2 matrix, $$\mathrm{Det}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$$
- Product rule: $$\mathrm{Det}(MM') = (\mathrm{Det}\, M)(\mathrm{Det}\, M')$$
- Scalar multiplication: $$\mathrm{Det}(cM) = c^n\, \mathrm{Det}\, M, \quad M : \mathbb{R}^n \to \mathbb{R}^n$$
- A **singular** matrix, where Det *M=0,* has a non-trivial **kernel**

Some properties of determinants

- Det M is **linear** as function of its individual rows or columns
- Det M is **antisymmetric** under interchange of adjacent rows or columns
- If any rows or columns are linearly dependent, Det $M = 0$.
- The determinant is the (oriented) **volume** of the image of the unit (hyper)cube

Some properties of the trace

- The **trace** is defined as the sum of the **diagonal** elements of a square matrix.
- The trace is **invariant** under a change of basis.
- The trace of the **identity** matrix in an *n*-dimensional space is *n*.
- The trace of a product is invariant under **cyclic** changes in the order; e.g.,

$$\mathrm{Tr}\,AB = \mathrm{Tr}\,BA$$
$$\mathrm{Tr}\,ABC = \mathrm{Tr}\,BCA = \mathrm{Tr}\,CAB$$

Matrix inverse

- For 2x2 matrices, the inverse when Det *M* is non-zero is given by

$$M^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

- The inverse acts on the left or the right to give the identity matrix

$$M(M^{-1}) = (M^{-1})M = I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Matrices and inverses

- Example: Rotation matrices
  - Columns are **image** of basis vectors
  - Single parameter for angle
  - No fixed points
  - Determinant = 1
  - Inverse matrix = reverse rotation

$$M_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \quad M_{-\theta} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$

Linear functions, adjoints, and the dual space

Linear functionals

- Linear functions on a vector space can map *V* to the real numbers (a one-dimensional vector space)

- For example,

$$\begin{aligned} \alpha : \mathbb{R}^3 &\to \mathbb{R}, \\ \alpha[\mathbf{v}] &= 2x + 3y - z \end{aligned}$$

- This action can be represented by a **row vector** (i.e., a 1 x *n* matrix) acting on *V*.

$$\begin{aligned} \boldsymbol{\alpha} &= \begin{pmatrix} 2 & 3 & -1 \end{pmatrix}, \\ \alpha[\mathbf{v}] &= \begin{pmatrix} 2 & 3 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 2x + 3y - z \end{aligned}$$

Dual space

- The **dual space** *V** of a vector space *V* is the space of all **linear functions** from *V* to the reals.

- Because the functions are linear, the space is closed under addition and multiplication

$$(c_1\alpha + c_2\beta)[\mathbf{v}] = c_1(\alpha[\mathbf{v}]) + c_2(\beta[\mathbf{v}])$$

- Given a basis of *V*, it is natural to construct a **dual basis** of *V** by selecting those functions which give 1 on one of the basis vectors of *V* and zero otherwise.

$$\epsilon_i[\mathbf{u}_j] = \delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

Dual space basis

- To see that this set is **linearly independent**, let a general linear combination act on a basis vector of *V*; only one term survives:

$$\alpha = \sum_{i=1}^{n} c_i \epsilon_i, \quad \alpha[\mathbf{u}_j] = c_j$$

so the element cannot be zero unless all the coefficients are zero.

- To see that the proposed basis **spans** *V**, write out a general element as

$$\alpha = \alpha[\mathbf{u}_1]\epsilon_1 + \alpha[\mathbf{u}_2]\epsilon_2 + \cdots + \alpha[\mathbf{u}_n]\epsilon_n$$

and note that both sides gives the same result when acting on any basis element of *V*, so that the expansion on the right-hand side spans *V**.

Adjoint transformation

- The **adjoint** of a linear transformation is a mapping between the duals of the vector spaces, in the opposite direction.

$$T : V \to W$$

$$T^* : W^* \to V^*$$

- Its action on dual elements is defined so that their action gives the same result as if the original dual element acted on $T\mathbf{v}$:

$$(T^*\alpha)[\mathbf{v}] = \alpha[T\mathbf{v}]$$

- The matrix of the adjoint transformation is the **transpose** of the matrix

$$(M^*)_{ij} = (M^\mathsf{T})_{ij} = M_{ji}$$

Matrix component gymnastics

- Two multiply two matrices
  - The inner dimensions must agree
  - Hold outer indices fixed; sum over all inner ones

$$(AB)_{ij} = \sum_{k=1}^{K} A_{ik} B_{kj}$$

Matrix component gymnastics

- The **transpose** of a matrix is obtained by **interchanging** rows and columns
  - The transpose of an *s* x *n* matrix is *n* x *s.*
  - The transpose of a column vector is a row vector
  - The transpose of a transpose gives back the original form

- Example:

$$M = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{pmatrix}, \quad \mathbf{v} = \begin{pmatrix} 9 \\ 10 \\ 11 \end{pmatrix}$$

$$M^{\mathsf{T}} = \begin{pmatrix} 1 & 5 \\ 2 & 6 \\ 3 & 7 \\ 4 & 8 \end{pmatrix}, \qquad \mathbf{v}^{\mathsf{T}} = \begin{pmatrix} 9 & 10 & 11 \end{pmatrix}$$

Matrix component gymnastics

- A matrix times a **vector** is a special case where we suppress one (uninteresting) index.
  - So in components, row and column vectors are handled **similarly**.
  - The column vector pairs up with the second index of the matrix; the row vector matches the first.
  - In examples, *K*,*L* may differ.

$$(B\mathbf{v})_i = \sum_{k=1}^{K} B_{ik} v_k$$

$$(\mathbf{w}^\mathsf{T} B)_j = \sum_{\ell=1}^{L} w_\ell B_{\ell j}$$

Symmetric matrices

- A matrix is **symmetric** if it is equal to its transpose

$$M = M^{\mathsf{T}}$$
$$M_{ij} = M_{ji}$$

- For 2x2 matrices, they take the form

$$M = \begin{pmatrix} a & b \\ b & d \end{pmatrix}$$

- Important examples include **correlation** and **covariance** matrices.

$$C_{ij} = \mathrm{E}\left[(R_i - \mu_i)(R_j - \mu_j)\right] = C_{ji} = C_{ij}^{\mathsf{T}}$$

Matrix component gymnastics

- Transpose of a product is the product of the transposes in **reverse order**,

$$(AB)^{\mathsf{T}} = B^{\mathsf{T}} A^{\mathsf{T}}$$

- In components,

$$\begin{aligned}
(AB)^{\mathsf{T}}_{ij} = (AB)_{ji} &= \sum_{k=1}^{K} A_{jk} B_{ki} \\
&= \sum_{k=1}^{K} (A^{\mathsf{T}})_{kj} (B^{\mathsf{T}})_{ik} = \sum_{k=1}^{K} (B^{\mathsf{T}})_{ik} (A^{\mathsf{T}})_{kj} \\
&= (B^{\mathsf{T}} A^{\mathsf{T}})_{ij}
\end{aligned}$$

Matrix component gymnastics

- For an element of the dual acting on a transformed vector,

$$\begin{aligned}
\beta[M\mathbf{v}] &= \sum_{i=1}^{m} \beta_i (M\mathbf{v})_i = \sum_{i=1}^{m}\sum_{j=1}^{n} \beta_i M_{ij} v_j \\
&= \sum_{i=1}^{m}\sum_{j=1}^{n} \beta_i (M^{\mathsf{T}})_{ji} v_j = \sum_{j=1}^{n} (M^{\mathsf{T}}\boldsymbol{\beta})_j v_j \\
&= (M^{\mathsf{T}}\beta)[v]
\end{aligned}$$

Systems of linear equations

Systems of linear equations

- Let's consider two kinds of systems of linear equations where in general we have ***s*** **equations** with ***n*** **unknowns**.

$$\begin{cases} M\mathbf{v} = \mathbf{b} & \text{inhomogeneous, or} \\ M\mathbf{v} = 0 & \text{homogeneous,} \end{cases}$$
$$\text{where } \mathbf{v} \in \mathbb{R}^n,\ \mathbf{b} \in \mathbb{R}^s,\ M : \mathbb{R}^n \to \mathbb{R}^s$$

- Expect to find, roughly, that the "inhomogeneous" equation has
  - **One** solution if $s = n$
  - **No** solutions if $s > n$
  - **Infinitely many** solutions if $s < n$

- The exact situation depends on the dimension of the image and kernel...
  - If $r = \text{rank}(M)$ is smaller than it could be, nature of solutions changes.

Case 1: *s* = *n*

- When $s{=}n$, there are **the same number** of equations as unknowns, and the matrix M is square. As we have seen, there are two sub-cases:

- If det $M \neq 0$, then M is invertible and there is a **unique solution**,

$$M\mathbf{v} = \mathbf{b} \implies \mathbf{v} = M^{-1}\mathbf{b}$$

- If det $M{=}0$, then the m equations are **not independent**. M has a **non-zero kernel**, and the rank is less than the dimension of the target space.

$$\dim(\mathrm{Im}\, M) = r = n - \dim(\ker M) < n = s$$

- So there will be some vectors **b** for which there is no solution.

Case 1: *s* = *n* and non-singular

- R example:

$$\boxed{\begin{aligned} x + 2y + 3z &= 1 \\ 4x + 5y + 6z &= 2 \\ 7x + 8y + 10z &= 3 \end{aligned}}$$

  - **solve(M)** inverts matrix
  - **solve(M,b)** gives **unique solution**

Solution:

$$\mathbf{v} = \begin{pmatrix} -1/3 \\ 2/3 \\ 0 \end{pmatrix}$$

```
> M <- matrix(c(1,2,3,4,5,6,7,8,10),byrow=T,nrow=3)
> b <- matrix(c(1,2,3),ncol=1)
> v <- solve(M,b)

> M
     [,1] [,2] [,3]
[1,]    1    2    3
[2,]    4    5    6
[3,]    7    8   10

> solve(M)
           [,1]      [,2] [,3]
[1,] -0.6666667 -1.333333    1
[2,] -0.6666667  3.666667   -2
[3,]  1.0000000 -2.000000    1

> v
           [,1]
[1,] -0.3333333
[2,]  0.6666667
[3,]  0.0000000


> M %*% v
     [,1]
[1,]    1
[2,]    2
[3,]    3
```

Case 1b: *s* = *n* and singular

- R example:

$$\begin{aligned} x + 2y + 3z &= 1 \\ 4x + 5y + 6z &= 2 \\ 7x + 8y + 9z &= 3 \end{aligned}$$

  - **solve** using **qr** gives a **particular solution** if it exists
  - Add any multiple of kernel for general solution

$$\text{Solution: } \mathbf{v} = \mathbf{v}_p + c\mathbf{z}$$

$$\mathbf{v}_p = \begin{pmatrix} 0 \\ 0 \\ 1/3 \end{pmatrix}, \ \mathbf{z} = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}$$

```
> M1 <- matrix(1:9,byrow=T,nrow=3)
> b1 <- matrix(c(1,2,3),ncol=1)
> v1 <- solve(M1,b1)
Error in solve.default(M1, b1) :
  system is computationally singular: reciprocal condition number =
1.54198e-18
> v1_p <- qr.solve(M1,b1)
Error in qr.solve(M1, b1) : singular matrix 'a' in solve

> M1
     [,1] [,2] [,3]
[1,]    1    2    3
[2,]    4    5    6
[3,]    7    8    9

> det(M1)
[1] 6.661338e-16

> qr(M1)$rank
[1] 2

> v1_p <- solve(qr(M1,LAPACK=TRUE),b1)
> v1_p
           [,1]
[1,] -0.1147976
[2,]  0.2295953
[3,]  0.2185357

> M1 %*% v1_p
     [,1]
[1,]    1
[2,]    2
```

Case 2: $s > n$

- When there are **more equations than unknowns**, then there are **no solutions** for at least some values of **b**.

- Because a "smaller" space is going into a "bigger" one, some vectors **b** in the target space **cannot be reached** from any vector in the "smaller" space. To get technical, from the Fundamental Theorem of Linear Transformations,

$$\dim(\mathrm{Im}\, M) = r = n - \dim(\ker M) \leq n < s$$

- Although there is no solution in general, some special points (in Im *M)* may yield solutions.

Case 2: $s > n$

- R example:

$$\begin{aligned} x + 2y &= -1 \\ 3x + 4y &= 1 \\ 5x + 6y &= 3 \\ 7x + 8y &= 5 \end{aligned}$$

  - **qr.solve(M,b)** gives **particular solution** if it exists
  - **Warning**: need to **check answer** since it also gives values when no solution exists(!)

Solution:

$$\mathbf{v}_p = \begin{pmatrix} 3 \\ -2 \end{pmatrix}$$

```
> M2 <- matrix(1:8,byrow=T,nrow=4)
> b2 <- matrix(c(-1,1,3,5),nrow=4)
> v2_p <- qr.solve(M2,b2)

> M2
     [,1] [,2]
[1,]    1    2
[2,]    3    4
[3,]    5    6
[4,]    7    8

> v2_p
     [,1]
[1,]    3
[2,]   -2

> b2 <- matrix(c(-1,1,3,6),nrow=4)
> v2_not <- qr.solve(M2,b2)
> v2_not
      [,1]
[1,]  3.50
[2,] -2.35

> M2 %*% v2_not
     [,1]
[1,] -1.2
[2,]  1.1
[3,]  3.4
[4,]  5.7
```

Case 3: $s < n$

- When there there are **more unknowns than equations**, then there are **multiple solutions**.

- Because a "bigger" space is going into a "smaller" one, some vectors must be mapped to zero. To get technical, from the Fundamental Theorem of Linear Transformations,

$$\dim(\ker M) = n - \dim(\operatorname{Im} M) \geq n - s > 0$$

- To any "particular solution" can be added any element of the kernel.

Case 3: *s* < *n*

- R example:

$$\begin{aligned} x + 2y + 3z + 4w &= 1 \\ 5x + 6y + 7z + 8w &= 1 \end{aligned}$$

  - **qr.solve(M,b)** gives a **particular solution**
  - ker(M) gives (sometimes inconvenient) basis for kernel

Solution: $\mathbf{v} = \mathbf{v}_p + c_1\mathbf{z}_1 + c_2\mathbf{z}_2$, where

$$\mathbf{v}_p = \begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \end{pmatrix},$$

$$\ker M = \operatorname{span}\left\{ \mathbf{z}_1 = \begin{pmatrix} 2 \\ -3 \\ 0 \\ 1 \end{pmatrix}, \mathbf{z}_2 = \begin{pmatrix} 1 \\ -2 \\ 1 \\ 0 \end{pmatrix} \right\}$$

```
> M3 <- matrix(1:8,byrow=T,nrow=2)
> b3 <- matrix(c(1,1),ncol=1)
> v3_p <- qr.solve(M3,b3)

> M3
     [,1] [,2] [,3] [,4]
[1,]    1    2    3    4
[2,]    5    6    7    8

> v3_p
     [,1]
[1,]   -1
[2,]    1
[3,]    0
[4,]    0

> ker(M3)
           [,1]       [,2]
[1,]  0.0000000 -0.5477226
[2,]  0.4082483  0.7302967
[3,] -0.8164966  0.1825742
[4,]  0.4082483 -0.3651484
```

Case 3: *s* < *n*

- R example:

$$\begin{aligned} x + 2y + 3z + 4w &= 1 \\ 5x + 6y + 7z + 8w &= 1 \end{aligned}$$

  - **qr.solve(M,b)** gives a **particular solution**
  - ker(M) gives (sometimes inconvenient) basis for kernel

```
ker <- function(M){  # adapted from nullspace
      r <- qr(M)$rank
      cols <- if (r>0) -(1:r) else (1:ncol(M))
      V <- eigen(t(M) %*% M)$vectors[,cols,drop=FALSE]
      if (length(V)==0)
        return(NULL)
      else return(V)
}

> ker(M3)
           [,1]       [,2]
[1,]  0.0000000 -0.5477226
[2,]  0.4082483  0.7302967
[3,] -0.8164966  0.1825742
[4,]  0.4082483 -0.3651484
```