AMS-512 Capital Markets and Portfolio Theory

Using Models

Robert J. Frey, Research Professor
Stony Brook University, Applied Mathematics and Statistics

frey@ams.sunysb.edu
http://www.ams.sunysb.edu/~frey

Understanding how to use models well is at the heart of the successful practice.

The Non-Mathematical in Mathematical Applications

In real world applications the usage of models depends on qualitative insights expressed mathematically. The structure of the model cannot be fully understood simply as formal mathematics.

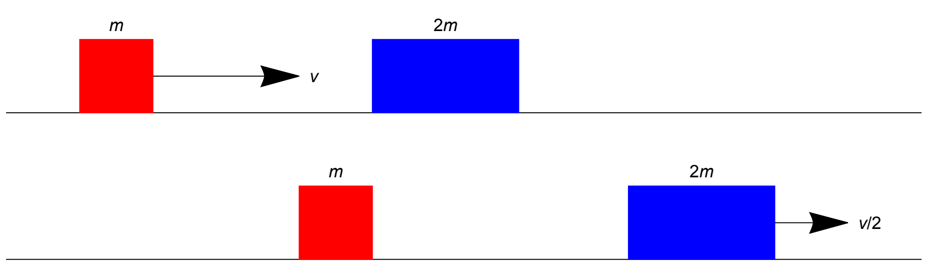
Examples

Conservation Laws in Physics

In a closed physical system momentum, mass times velocity, is conserved. Consider a simple two-body system on a line. As they collide the total momentum before the collision at time t and after the collision at time $t+\epsilon$ is conserved; hence, using m for mass and v for velocity we have

$$m_1\, v_1(t) + m_2\, v_2(t) = m_1\, v_1(t+\epsilon) + m_2\, v_2(t+\epsilon)$$

If we consider the case of an *elastic collision*, where there is a complete transfer of momentum, then a body with mass m and velocity v colliding with a stationary body of mass $2m$ will result in the first body stationary and the second moving off with velocity $v/2$.



Conservation of Momentum in Elastic Collision

The conservation of momentum is *stated* mathematically but it is not *derived* mathematically. It is a hypothesis supported by extensive and systematic experimentation and observation. Looking solely at the mathematics as if it contained the entirety of the model is—according to the Buddhist parable— akin to confusing a pointing finger with the object it is pointing at.

Arbitrage-Free Markets in Finance

The Black Scholes solution to the pricing of European options starts with the notion of a *replicating portfolio* $\Pi(t)$ in which the changes in the price of an option with price F is *delta hedged* by taking a short position in its underlying with price S. The hedge is adjusted from instant to instant such that in the replicating portfolio's the random dynamics of F and S are canceled out.

$$\Pi(t) = F(t) - \frac{\partial F}{\partial S} S(t)$$

Using Itö's lemma, we can express the dynamics of the replicating portfolio as a *deterministic* PDE:

$$d\Pi(t) = \left(\frac{\partial F}{\partial t} + \frac{1}{2}\sigma^2 S(t)^2 \frac{\partial^2 F}{\partial S^2}\right) dt$$

In an arbitrage-free market, a portfolio whose dynamics are deterministic will realize the risk free rate of return r. Thus, the replicating portfolio must also satisfy this condition.

$$d\Pi(t) = r\,\Pi(t)\,dt$$

Setting the right-hand sides of the two equations above equal to one another allows us to solve the resulting parabolic PDE in terms of the known parameters which describe the price dynamics of S and its relationship to F. However, there is no *mathematical* basis for the equality of the two expressions above for $d\Pi(t)$. One justifies that equivalence solely based upon the arbitrage free-assumption, which is external to the mathematics.

Is There No "Truth" Outside Mathematics?

We tend to think of the evidence-based approach as models being "proven" by data. This sort of thinking is dangerous. In mathematics, we are given formal systems supported by a set of axioms which we use to prove additional results. Such results—*in the context of that system*—are true. That is *never* the situation in real world applications. In applications even highly accurate mathematical models are always approximations.

However, it also would be a serious mistake to think of even mathematical facts as true in some universal sense. Mathematics is not a monolith. The fact is, there are different ways of building mathematical systems—all having their limitations. The idea that mathematics represents some sort of ultimate truth is wrong.

As Leopold Kronecker once wrote, "God made the integers; all else is the work of man." I would say that Kronecker was an optimist and narrow "integers" to "natural numbers".

Axiom of Choice

The *Axiom of Choice* (https://en.wikipedia.org/wiki/Axiom_of_choice) is essentially the assumption that that given a collection of non-empty bins, one can select an object from each bin. This holds even if the collection is infinite. While it may seem that this obvious, it turns out that in set theory the Axiom of Choice cannot be deduced from the other axioms. Thus, we can have different versions of set theory that appear perfectly reasonable.

Wikipedia gives the example, "Until the late 19th century, the axiom of choice was often used implicitly, although it had not yet been formally stated. For example, after having established that the set X contains only non-empty sets, a mathematician might have said 'let $F(S)$ be one of the members of S for all S in X.' In general, it is impossible to prove that F exists without the Axiom of Choice."

There are cogent arguments for and against the inclusion of the Axiom of Choice; specifically, it lets you prove that certain things or solutions exist without being able to specify how to construct them or, worse, that they exist but there is no way to construct them. For some mathematicians the Axiom of Choice is just intuitively obvious. For others, the idea that one can assert the existence of something without being able to construct it is dangerous and indeed using it can lead to results which are highly non-intuitive.

Gödel's Incompleteness Theorems

Gödel's Incompleteness Theorems (https://en.wikipedia.org/wiki/Gödel%27s_incompleteness_theorems) arose from his study of exactly what could and could not be proved within a consistent formal system of a certain minimal complexity. His results, as expressed in his two *Incompleteness Theorems*, had profound effects on mathematics.

Let $\mathcal{F}$ be a consistent formal system within which a certain amount of arithmetic can be carried out. The *First Incompleteness Theorem* states that $\mathcal{F}$ is incomplete; *i.e.*, there are statements in the language of $\mathcal{F}$ which cannot be proved or disproved in $\mathcal{F}$.

The *Second Incompleteness Theorem* deals with the consistency of $\mathcal{F}$, *i.e.*, it contains no contradiction. According to the second theorem $\mathcal{F}$ cannot prove that it itself is consistent.

The issues that Gödel's work uncover typically arise in self-referential statements. A linguistic example is, "I am a liar." If the statement is true, then it is false. If it is false, then it is true. The assertion that the person is or is not a liar can only be resolved by appealing to something outside that referenced by the sentence.

Another example: For a system to be true/perfect/complete in an absolute sense would require it to include itself. Thus, a perfect system can only exist if it includes itself as a proper subset, a ridiculous assertion. If we cannot assert completeness for any reasonably complex mathematical system, then we cannot do so for any model that relies on that system!

From a practical point of view the Incompleteness Theorems mean that any model one has will be incomplete. It must continually checked against reality and updated.

Importance of Known-To-Be-False Models

"Since all models are wrong the scientist cannot obtain a 'correct' one by excessive elaboration. On the contrary following William of Occam he should seek an economical description of natural phenomena. Just as the ability to devise simple but evocative models is the signature of the great scientist so over-elaboration and over-parameterization is often the mark of mediocrity."—George Box

To a model effectively one must understand:

- What problems is the model is meant to deal with,
- The nature of compromises the model makes, and
- The domain of application, outside of which the model is not useful.

(Possibly) Useful "Facts"

A moon revolves its planet.

We think of the planets revolving around the Sun and the Moon around the Earth. The fact is that two bodies in orbit revolve around their common center of gravity. Looking at the Solar System as a whole all of the bodies about rotate around their collective center of gravity in a chaotic manner, much different from the simple secondary-around-primary model. Moreover, the orbits are further perturbed by relativistic effects. It was precisely these discrepancies in the orbit of Mercury that represented the first validation of Relativity.

Nevertheless, for many applications, such as satellite orbits, a simple two-body model is more than adequate. From the point of view of teaching about the structure of the Solar System, a planets-around-Sun model provides a good approximation of what is going on because the majority, 98%, of the mass of the Solar System is contained in the Sun. The center of gravity of the Solar System is well inside the Sun. The Sun does have a slight but measurable wobble, however. See https://en.wikipedia.org/wiki/Orbit.

The idea that a satellite orbits the Earth is a useful one, but it is a terrible one for describing cases with comparable masses such the Pluto-Charon system.

A person in orbit is weightless.

The gravitational attraction between two bodies is proportional to the product of their masses and inversely proportional to the square of their distance. We can approximate the bodies as two point masses at their respective centers of mass. Thus, the weight experienced by an object is based on a distance to the center of the Earth, 3959 miles.

An object in a low Earth orbit, LEO, of about 200 miles weighs about $(3959 + 200)^{-2} / 3959^{-2}$ or 91% what it does on the surface of the Earth. For example, 160 pound person weighs about 145 pounds at this altitude. The reason individuals experience a “state of weightlessness” in orbit is because they are in free-fall, not because there is no gravity in space. The Wikipedia article above on orbits also covers this effect.

Financial markets are efficient; hence, one must hold the market portfolio.

The basic assumption of the Efficient Market Hypothesis (EMH) is that market prices reflect all known information. See https://en.wikipedia.org/wiki/Efficient-market_hypothesis. Thus, future moves are caused by new information and such moves are in principle unpredictable. One of the consequences of this assumption (with a few more tacked on) is that if one wants to hold a portfolio of stocks, then a portfolio reflecting current market weights is optimal. This is called passive investing.

While the fact that the S&P 500, and mutual funds and ETFs that track it, outperform most active managers may seem to confirm this, it turns out that this is the result of a simple tautology. The index represents in total the sum of the capital-weighted portfolios of the market’s participants; *i.e.*, it is the average. Roughly, speaking half of those participants will be below average and half will be above average.

Market participants, unlike indexes, do not operate without costs; hence, the average investor must underperform the index. The tracking portfolios trade very little so they are the least affected. The observation that a portfolio that tracks the index outperforms is therefore going to be true regardless of how efficient the market is. Claiming that active management is a failure on this basis is like claiming an educational system is failing because half of its students are below average.

Also, if everyone held the supposedly optimal market portfolio, then no one would trade. How then does the market come to reflect the impact of new information? Clearly, someone must trade. Without the trading of active managers the market would (even more) poorly reflect economic reality and their actions “raise all boats”. This doesn’t mean that for many investors holding a passive market portfolio is a bad idea. It just doesn’t necessarily have anything to do with the EMH.

Historical Insights

- Wittgenstein’s Ladder - https://en.wikipedia.org/wiki/Wittgenstein%27s_ladder.
- Lies to Children - https://en.wikipedia.org/wiki/Lie-to-children
- Expedient Means (upāya kaushalya उपाय कौशल्य) - https://en.wikipedia.org/wiki/Upaya
- Toy Models - https://en.wikipedia.org/wiki/Spherical_cow
- Fermi Problems - https://en.wikipedia.org/wiki/Fermi_problem

Examples of Simple Models

- What Does the Model Do?
- What Does the Model Leave Out or Misrepresent?
- How Can the Model Can Be Used?

Ballistic Trajectories

What Does the Model Do?

Velocity is a vector representing the first derivative of displacement with respect to time and can be expressed using the standard basis as

$$\mathbf{v} = v_x\,\mathbf{i} + v_y\,\mathbf{j}$$

The magnitude of the velocity is the scalar quantity

$$v = |\,\mathbf{v}\,| = \sqrt{v_x{}^2 + v_y{}^2}$$

Acceleration is the first derivative of velocity with respect to time. There is no horizontal acceleration, but gravity, represented by g, acts downward.

$$a_x = 0$$
$$a_y = -g$$

Thus, velocity as a function of time. Let v be the initial velocity and ϑ denote its angle to the horizontal, then velocity for $0 \le t \le t_{\max}$ can be expressed as

$$v_x(t) = v\cos\vartheta$$
$$v_y(t) = v\sin\vartheta - g\,t$$

Position is then just the the velocity integrated with respect to time; *i.e.*, $x(t)\,\mathbf{i} + y(t)\,\mathbf{j} = \int v_x\,\mathbf{i} + v_y(t)\,\mathbf{j}\,dt$.

$$x(t) = v\,t\cos\vartheta$$
$$y(t) = v\,t\sin\vartheta - \frac{1}{2}\,g\,t^2$$

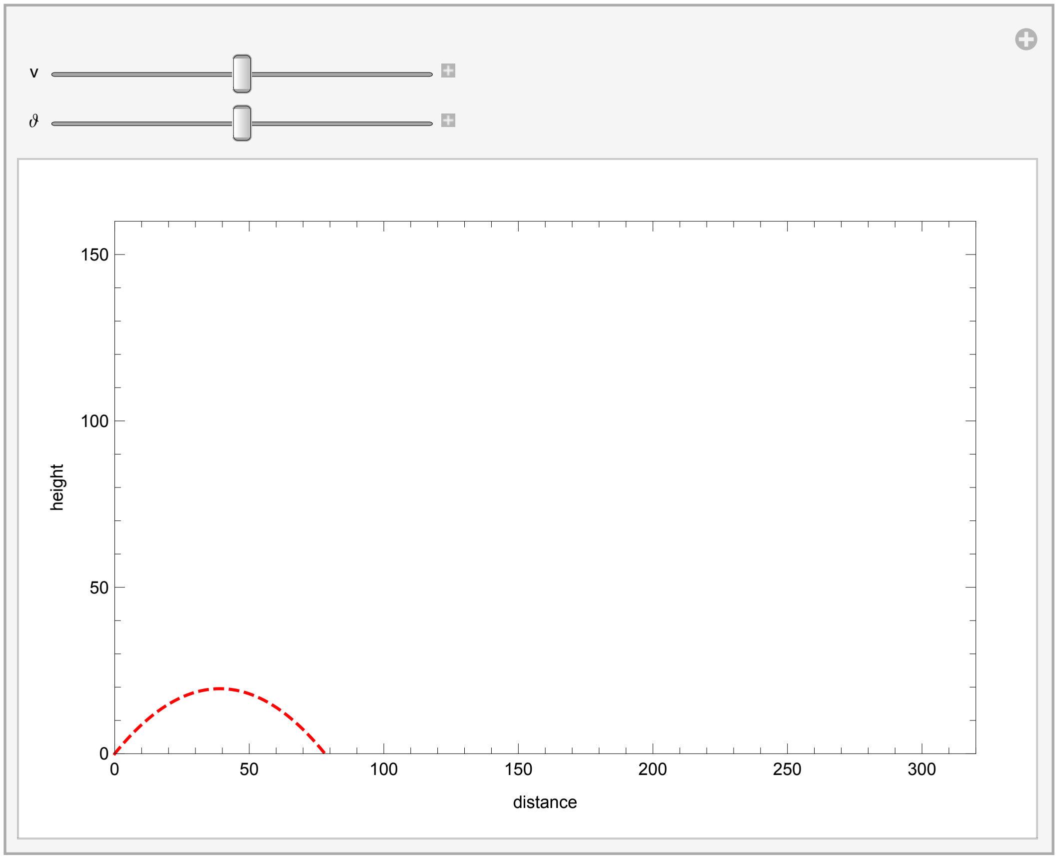
The total time in flight is computed for the value of $t > 0$ such that $y(t_{\max}) = 0$.

$$(v\sin\vartheta - g\,t)\,t = 0 \Longrightarrow t_{\max} = \frac{2\,v\sin\vartheta}{g}$$

The total horizontal distance covered is $x(t_{\max})$.

$$x_{\max} = (v\cos\vartheta)\,t_{\max} = \frac{v^2\sin 2\,\vartheta}{g}$$

Out[221]=



What Does the Model Leave Out or Misrepresent?

There are many simplifications. Among them: The mass is a point mass that instantaneously appears with its initial velocity. Air resistance and the curvature of the Earth are not considered. To get even more nit-picky, we have used Newtonian mechanics, ignoring relativistic effects.

How Can the Model Be Used?

Although the model leaves out air resistance, its results clearly place an upper limit on the distance traveled given an initial velocity and firing angle. If the object is dense, then air resistance may not have a material effect. It also demonstrates that the maximum distance is achieved by a firing angle of $\pi/4$ radians or 45°. For distances less than the maximum, there are clearly two trajectories, a low and high with very different transit times. Unless the distance traveled is extreme, then the issue of the curvature of the Earth can be ignored.

M/M/1 Queues

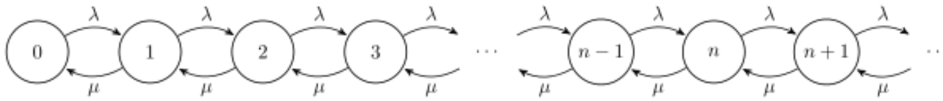
Situations in which customers arrive for service queuing up if the system is busy cover a wide range of situations from traditional customer service systems, to traffic congestion, planes taking off and landing at airports, and computer and networking systems.

What Does the Model Do?

An M/M/1 queuing system is a single-server system in which customers arrive at the server at rate λ and are served at rate μ. Both the arrival and service processes are Poisson processes. If the server is busy, then customers queue up and wait for service, which is performed by a first-come-first-served (FCFS) protocol. The utilization of the server is $\rho = \lambda/\mu$.

`Source : Wikipedia`

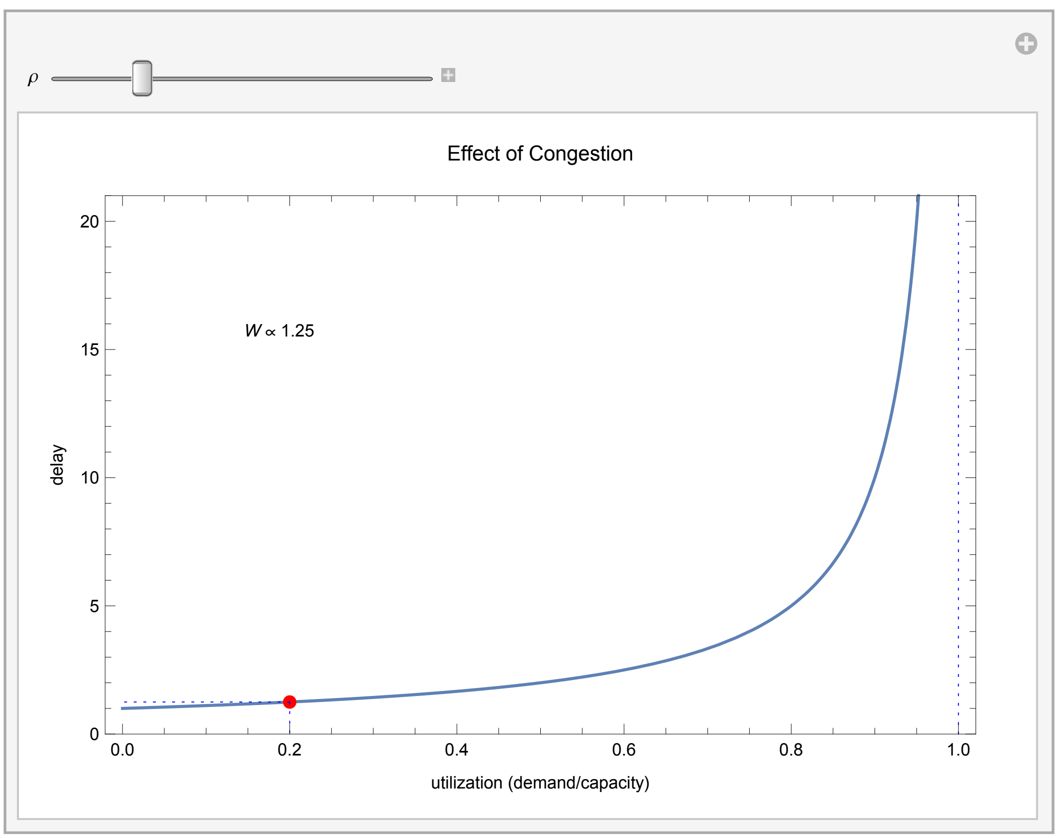
The M/M/1 queueing system can be modeled as a continuous-time Markov process as illustrated below. At steady-state the average total time-in-system service time W is proportional to $(1-\rho)^{-1}$; *i.e.*, W is inversely proportional to the residual capacity.



`Source : Wikipedia`

M/M/1's properties make it easy to solve, but many of its properties generalize to other queuing problems. This is especially the case in heavy traffic situations where the precise nature of the arrival and service processes become less important. This is even true when the service system contains multiple service processes since under heavy traffic the service system is usually near capacity and behaves in aggregate much like a single fast server.

Out[222]=



What Does the Model Leave Out or Misrepresent?

Poisson processes are memoryless, *i.e.*, the arrival and service times are exponentially distributed. In general this is not true. The arrival and service rates often are themselves non-stationary. Many systems have service protocols other than FCFS and have additional features such as balking where customers either leave or are turned away when the system is overloaded. Systems also typically are multi-server. Finally, often one is interested in transient behavior, not steady state. All of these will affect the behavior of the system.

How Can the Model Be Used?

Despite the model's simplicity, it can be productively applied to many complex situations.

- The most important insight is that, unless one can carefully schedule arrivals and services with complete certainty, attempting to achieve 100% utilization of a service facility leads to disaster.
- The fundamental non-linear relationship between utilization and total time in system will hold for a wide variety of queuing systems. This effect, known as Little's law, is an important theoretical result.
- Even when arrivals and service can be carefully scheduled, if one does not leave a certain amount of slack in the system, then any disruption will cause out-of-control delays to occur. This can be seen in airports where relatively small delays in arrivals or brief service interruptions can take days to recover from.
- In cases where demand for service varies over the day, using average interarrival times is dangerous. During high-demand periods when $\lambda > \mu$ the waiting lines will steadily increase without bound and will only clear when the demand falls below the capacity of the system. Rush-hour and other forms of traffic congestion, *e.g.*, rubber necking at accidents or inclement weather, are good examples of this.
- Given that time in system is inversely proportion to utilization, this informs one's intuition about how changes in demand or system capacity will affect a system with known performance. Often complex situations consisting of multiple servers, multiple stages of service, and networks of queuing systems can be analyzed to first-order using the basic insights gained.

Dividend Discount Models

The price of an asset can be modeled as the present value of its future cash flows. In the case of stocks this forms the basis for so-called dividend discount models (DDM). Companies often don't pay out their entire profits as dividends and a related measure such as free cash flow is often substituted.

What Does the Model Do?

The simplest form of the DDM assumes that cash flows occur at fixed intervals with the next one c occurring one period hence. Further, the growth rate in cash flow is a constant g and the rate of return for discounting is a constant r. Thus, the solution to the DDM simplifies to the infinite series below, which converges for $r < g$.

$$S = \sum_{t=1}^{\infty} c \frac{(1+g)^{t-1}}{(1+r)^t} = \frac{c}{r-g}, \quad r > g \geq 0$$

Despite its simplicity the model derives the price of a financial asset in terms of the time-dependent economic benefits it yields. This is certainly a rational basis for such estimation. It also tells us that, for example, the stock price of an early stage growth company that over the near term is not expected to show a profit can only be justified by assumptions to what r and g will settle into in the longer term. If some investors had performed such an analysis, then the late 90s tech bubble and its subsequent damaging crash may have been more moderate and perhaps not even have occurred.

What Does the Model Leave Out or Misrepresent?

Certainly, the time-dependent rates $r(t)$ and $g(t)$ are random variables, not constants. While the various forces which effect the cash flow can be subsumed into the growth function, cash flow can be defined in a number of ways, *e.g.*, dividends or free cash flow. The value of the discounting rate $r(t)$ is far from straightforward as we would expect riskier investments to demand a higher rate when compared against the background of other rates for the same time period. The model also does not take into account events which can cause major economic dislocations. Examples of such would be revolutions, depressions, and wars.

How Can the Model Be Used?

Thomas Piketty in his bestselling *Capital in the Twenty-First Century* claimed that because $r > g$ in the long run this could not help but to generate increasing wealth inequality. (See “Why Economists Disagree with Piketty’s ‘r - g’ Hypothesis”.) This meant that policy makers need to implement active wealth distribution policies. The arguments that have gone back and forth have centered on whether this model is being misapplied by Piketty. The cited article is an interesting case study on how models can be applied and misapplied and why it is not always easy to tell the difference.

Black-Scholes Option Pricing Models

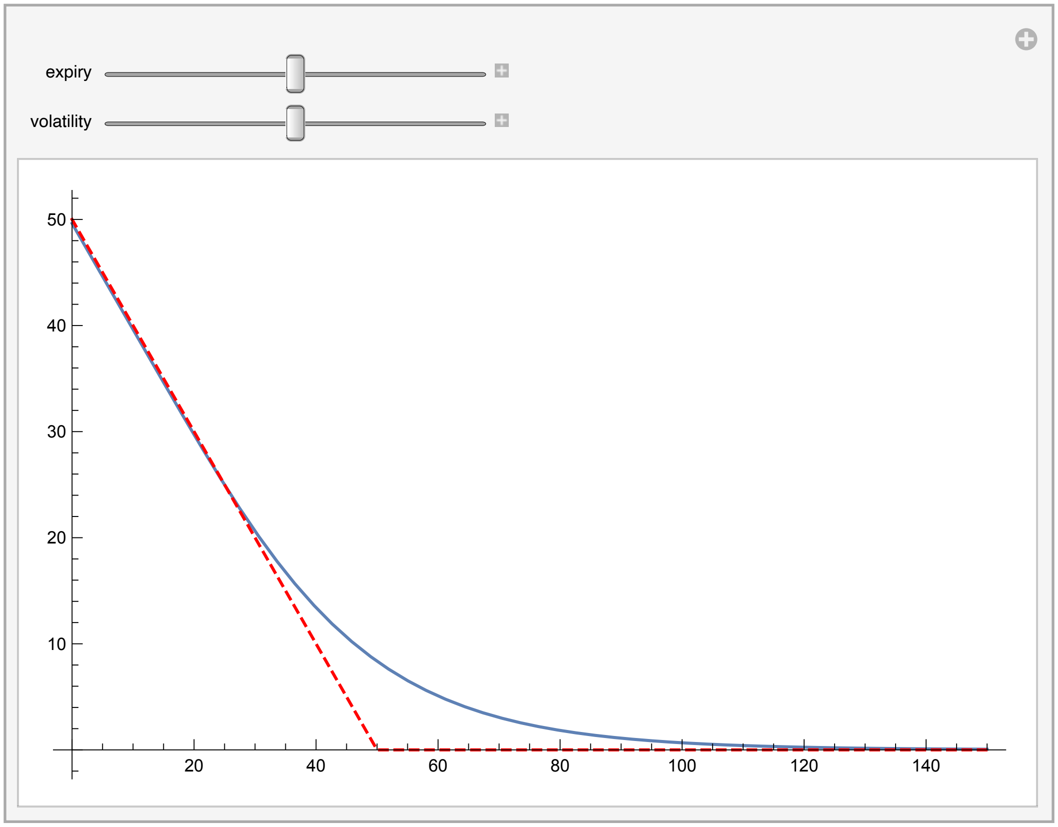
The Black-Scholes options pricing model is used extensively throughout finance. It forms the starting point for a wide array of models of increasing complexity.

What Does the Model Do?

An overview of Black-Scholes has already been covered. More detail can be found in https://en.wikipedia.org/wiki/Black-Scholes_model.

Perhaps the most important insight is that the price of the option is based on current market conditions. A option holder’s opinion as to whether the price of the underlying is going up or down is irrelevant (although the decision whether to hold a given option position certainly is not). Even if one does not find the delta hedging argument convincing, equivalent results can be derived using alternative formulations (http://www.fooledbyrandomness.com/PCParity.pdf).

Out[223]=



In[224]:=
```
FinancialDerivative[{"European", "Put"},
 {"StrikePrice" → 50.00, "Expiration" → 0.25, "Value" → 4.91},
 {"InterestRate" → 0.01, "CurrentPrice" → 50.0}, "ImpliedVolatility"]
```

Out[224]= 0.500462

What Does the Model Leave Out or Misrepresent?

In its simplest form, the Black-Scholes model assumes that the parameters are constant, while they are clearly stochastic processes changing over time. One area in which the basic model is frequently extended is to assume volatility is a random variable. This tends to increase the likelihood of tail events which is consistent with real assets; however, there is no universally agreed model for *stochastic volatility*. Also, real markets are not always continuous, prices can make significant discrete jumps from one price to another, *e.g.*, overnight or during liquidity crises. The arbitrage argument do not factor in liquidity and transaction costs.

How Can the Model Be Used?

In the real world one do not solve for the price of a given option. One observes the market price and infers an *implied volatility* consistent with that price. One way to think of implied volatility as a coordinate transformation—from price-space to volatility-space—where comparisons of value can be more intuitively applied. This is similar in spirit to evaluating and comparing bonds by their *yields*. This transformation allows practitioners to make common sense adjustments based on the effect these unmodeled effects have on options pricing.

Thus, traders and other practitioners who work with options think in terms of implied volatility. The deficiencies in the Black-Scholes model are compensated by systematic adjustments to implied volatility, *e.g.*, the volatility smile which tracks changes in implied volatility at different strike prices (https://en.wikipedia.org/wiki/Volatility_-smile), the volatility surface which tracks changes in implied volatility at different strike prices and expiries, and the evolution of the volatility surface over time.

Portfolio Optimization

Harry Markowitz's mean-variance optimization has become a standard method of building portfolios which attempt to balance overall reward and risk. Although the model has several challenges associated with its implementation and typically underestimates tail risk, it nevertheless represents a clear framework for making decisions about asset allocation.

What Does the Model Do?

Consider the simplest form of the mean-variance (Markowitz) model where $0 \le \lambda \le \infty$ is a trade-off parameter balancing reward measured as expected return and risk measured as the variance of return. The QP is then

$$\mathcal{M} = \min_{x} \left\{ \frac{1}{2} \mathbf{x}^T \mathbf{\Sigma} \mathbf{x} - \lambda \boldsymbol{\mu}^T \mathbf{x} \right\}$$

Assuming a positive definite covariance matrix, setting the gradient to zero finds the minimum

$$\nabla \mathcal{M} = \frac{1}{2} \mathbf{x}^T \mathbf{\Sigma} \mathbf{x} - \lambda \boldsymbol{\mu}^T \mathbf{x} = \mathbf{0}$$

If we assume the covariance matrix is of full rank, solving for $\mathbf{x}$ to proportionality is

$$\mathbf{x} = \lambda \mathbf{\Sigma}^{-1} \boldsymbol{\mu}$$

As a special case, if the investments are uncorrelated, this further simplifies to

$$x_i \propto \frac{\mu_i}{{\sigma_i}^2}$$

What Does the Model Leave Out or Misrepresent?

Portfolio optimization problem normally have a set of constraints; however, such constraints may be viewed as perturbations of the solution. Consider, for example, a problem with linear equality constraints.

$$\mathcal{M} = \min_{x} \left\{ \frac{1}{2} \mathbf{x}^T \mathbf{\Sigma} \mathbf{x} - \lambda \boldsymbol{\mu}^T \mathbf{x} \;\middle|\; \mathbf{A} \mathbf{x} = \mathbf{b} \right\}$$

If we price out the constraints using a vector of Lagrange multipliers ζ, then it is easy to show that a necessary (though not sufficient) condition for optimality (https://en.wikipedia.org/wiki/Quadratic_programming#Equality-_constraints) is

$$\mathbf{x} = \mathbf{\Sigma}^{-1} \left(\lambda \boldsymbol{\mu} - \mathbf{A}^T \zeta \right)$$

which is similar to an unconstrained case where the return vector $\boldsymbol{\mu}$ has been perturbed to achieve feasibility. Hence, the inversion of the covariance is still a key element of the solution.

If there are N assets in the investment universe, then the estimation of covariance involves $N(N+1)/2$ unique parameters; *e.g.*, $N = 500 \Longrightarrow 125{,}250$ parameters. One needs a large number of observations to estimate covariance properly. Even then, the assumption that covariance is stationary is not supported by actual return histories.

The fact that the model represents risk in terms of portfolio variance is based on the assumption of multivariate Normal returns; another assumption not justified by real returns. This reliance on variance means that the optimization will be blind to certain risks that are not well approximated by it.

How Can the Model Be Used?

As noted under the previous heading, the unconstrained case can give insight into the solution of more realistic problems. Establishing the role of the covariance matrix is an important insight, because it clearly shows that its efficient statistical estimation and the numerical stability of its inversion is key to portfolio selection. In evaluating the effect of statistical uncertainly upon portfolio selection, it also makes clear that one's focus needs to be on the sampling distribution of the inverse covariance matrix, not the covariance matrix itself.

While variance is an incomplete measure of risk, it is nevertheless an important component of it. It can be argued that as a portfolio's variance increases its exposures to certain other forms of risk also tend to increase. Thus, it is can be a reasonable strategy to limit analysis to the set of mean-variance efficient portfolios when extending analysis to other measures of risk.

A Model Spectrum

We will cover a spectrum of model types based on how they are exercised to gain insight into a situation or make predictions of a system's behaviors. It is in no way claimed to be complete. It certainly isn't the only way models can be classified. For example, we don't delineate deterministic from stochastic models which is another way that models can be categorized.

Where the model of a given problem fits in the model depends upon *both* the inherent characteristics of the problem and one's own state of knowledge—or perhaps more appropriately our state of ignorance—about how to solve the model as posed.

Closed Form

Some problems admit a *closed form solution*. Their dynamics or evolution can be modeled with a single equation or system of equations without the necessity of computing intermediate results. The ballistic trajectory model above is an example of a model with a closed form solution. The state of the missile at any point is time can be computed directly without integrating instant-to-instant along the path.

Analytical/Numeric

Simulation

Computationally Irreducible

Chaotic Systems

Aphorisms

We will delve into aphorisms in a little more depth than the cases above because aphorisms, as the simplest and most general form of modeling, best illustrate the nuances involved in using models effectively.

Definition

aphorism - (noun) - a pithy observation that contains a general truth, often expressed as a metaphor or simile

Generality Has Its Limitations

Here is a list of seemingly contradictory aphorisms, most drawn from Carl Sagan's *Demon Haunted World*.

- Detail Oriented *vs*. Big Picture
  - "The devil is in the details."
  - "Don't sweat the small stuff."
- Action *vs*. Communication
  - "Actions speak louder than words."
  - "The pen is mightier than the sword."
- Collaboration *vs*. Independence
  - "Two heads are better than one."
  - "Too many cooks spoil the broth."
- Action *vs*. Caution
  - "He who hesitates is lost."
  - "Fools rush in where angels fear to tread."

Applying Aphorisms

Almost everyone has had the experience of having a generally sound idea or statement trivialized by irrelevant nit picking. It's often a go-to move for Internet trolls. Aphorisms aren't theorems, but they are a form of model whose purpose is to focus attention on an issue or to encourage a question about a problem.

An aphorism is a prescription, but implicit in that prescription is a diagnosis. The apparent contradictions in the aphorisms above are illusory. While one may think this is obvious, unfortunately a good deal of rather stupid discourse is undertaken criticizing "general truths" because they are not *universal* truths and misapplying models to situations where they do not apply because someone told us they were "right".

A silly example of this is the well-known nonsense that science has proven that bubble bees cannot fly. This comes from applying equations meant to model *fixed-wing* aircraft weighing hundreds of *thousands of kilograms* to a tiny insect with *moving wings* weighing about *ten centigrams*.

Consequences

Know the Domain and Understand the Limitations

As with the case of "proving bumble bees can't fly", it is important to understand what domain a model is designed for. Even within that domain, it will have limitations that may trap the careless. It is always important to never confuse the map with the territory. Many people, who would otherwise claim to know better, have treated model results as if they were reality. A great number of serious mistakes have been made because of model blindness.

Fit + Errors

Clearly, we mean errors in the sense of deviations of the real system and a model of it. Even though errors are often characterized as noise, they are as much a part of the model as the fit. Too often a modeler will fit a model and throw away or over-summarize the errors. *The important thing to remember is that given that any model is incomplete, its errors must contain information*. The assumption that errors are noise is always false. Of course, the errors-as-noise compromise is often made in using a model. This must be done deliberately, never blindly.

Obscurity of “Better” Models

Complex models can also be more obscure. It can be difficult to develop intuition and insights into how system's behaviors change with changes in assumptions and parameters. Even when a complex model is necessary, using it along with a simpler model that properly captures key elements of behavior can keep one grounded in reality and help to avoid confusing the map for the territory. As a model gets more and more abstract or detailed, it is more likely to need some form of “sanity check” to keep it tied down to reality.

Build a Toolkit

Models have their limitations; therefore, one must have a toolkit of models, not only for different problems, but addressing the same problem from multiple directions. Often models can be used to check one another. A complex and difficult to run model may be initialized with an approximate solution generated by a simpler model.

Need for a Meta-Model

The fact that one inevitably needs a number of different models means one must also have an overall approach on how to use the tools in that kit. This is a process that is both art and science, and experience will be an important teacher.

Three Questions

Remember the three questions we need to ask of any model:

- What Does the Model Do?
- What Does the Model Leave Out or Misrepresent?
- How Can the Model Can Be Used?

Matching one’s answers to those questions to the problem at hand is a good place to start.

The Observe-Orient-Decide-Act Loop

The OODA Loop was developed as a general decision framework for fighter pilots where actions must be made on sound decisions under conditions of high stress and tight time constraints. (See https://en.wikipedia.org/wiki/OODA_loop.)

An Incomplete List of Desiderata

There are many approaches to designing and using models. Below is a set of characteristics that we have found useful. They look at the problem from different and often competing perspectives and are interdependent in their effects on one’s model design.

Fidelity

The objective of any model is to create a construct—whether physical or mental—that in some way reproduces the behavior of reality. It helps one to understand or predict some aspects of reality. Models are meant in that sense to be representative, to mirror in their behavior the behavior of the system being modeled.

It would be wonderful if we could make a model better simply by making it more complex. Unfortunately, most the reasons behind the other desiderata below address why that isn't so.

Simplicity

Detailed and complex models may sometimes more closely represent the system being studied. Unfortunately, there are also possible disadvantages:

- The more parameters a model has, the greater the danger of over-fitting. The fact that a model fits the history becomes less and less comforting as the number of parameters increases. Experience shows that out-of-sample behavior suffers. Errors in implementation are also harder to detect in complex models.
- A simple back-of-the-envelope model is extremely useful in the early stages of analysis to test initial ideas and focus attention even when a more complex model is needed in the final solution.

Simplicity, what is sometimes called parsimony, is an important consideration and there are obvious natural trade-offs between fidelity and simplicity. As Einstein is reputed to have said, "Everything should be made as simple as possible, but not simpler."

Stability

One important characteristic is stability—cross-temporal generality. Models are driven by observations. When those observations are drawn too narrowly, the stability of any model based on them will suffer. Often the stability of a model can be improved by carefully modeling the source of the instability.

For example, it is well known that many financial time series display a time-varying variance, a condition known as *heteroskedasticity*. GARCH models (https://en.wikipedia.org/wiki/Autoregressive_conditional_heteroskedasticity) deal with this issue by introducing a second time series model to handle volatility changes. In effect, instead of modeling a scalar x_t it models the 2-vector $\left(x_t, \sigma_t{}^2\right)$. If GARCH is successful, then while the model in x_t is non-stationary, increasing the state space to $\left(x_t, \sigma_t{}^2\right)$ may give us a stationary model. The price one pays for this is increased model complexity. However, as will be covered below, broadening the applicability of a model can sometimes actually make it less complex.

Commonality

Similar to stability, another important characteristic is commonality—cross-sectional generality. If one feels the needs to fit models to a number of different related situations, then even though the models individually may be simpler, the total number of parameters may be quite large and the amount of data available to fit each model may be small.

Usability

Models must be usable. A model meant to guide a high frequency trading system that takes an hour to run is not practical. The data used to drive the model must be available and the costs of collecting and maintaining them must be balanced against the model's expected benefits.

Rationality

All models are shadows of the past. Some may be based on careful deep reasoning from basic principles; others on purely empirical fits to available data. Regardless, all are tied back to past experience. While one often makes a model design decision based on observed patterns in the data, having some cause-and-effect insight into why a particular model form make sense is important.

One danger of looking at many different models against the same set of data is that simply by chance some of them will fit. The web site http://www.tylervigen.com/spurious-correlations gives several amusing examples of this effect. For example, there is a high degree of yearly correspondence between the number of movies starring

Nicholas Cage and the number of people drowning in pools.

Next Step: Go back.

The material above looks at a profoundly difficult problem from a number of different perspectives which are all interdependent. The best thing a reader can do at this point is go back and review each section above with the additional insights gained from the others.