

Module 13 Forecasting, Simulation, and Monte Carlo Analysis

Learning Goal for Module 13

Many specialized software systems and platforms support simulation. Excel can serve as a preliminary method for exploring changes in a simulation model

After reviewing the concepts of projection and forecasting based on the regression model, this Module explores how to generate scenario values using the four probability distributions as an extension of the material in Module 7. Monte Carlo models and the use of data tables for common business and economic problems complete the discussion.

By the end of the Module, you will be able to:

- Formulate a decision problem in probabilistic terms
- Understand the structure of a simulation model
- Use regression to develop scenarios for future trends
- Learn how to generate draws from any distribution to simulate across a range of model parameters
- Explore model properties using systematic parameter variation

This module is under development and will not be including in the final exam

1. Introduction

Economic policy involves applying conjectures to an economic model of reality. Economists respond to questions such as “What will happen to sales (total revenue) if own price rises or if the price of a complement falls?” Similarly, one might be interested in the impact of an increase in deficit spending on the interest rate.

These “what-if” experiments create projections of the future course of one or more outcomes based on imagined changes in one or more “causal” variables. When causal (independent) variables change due to purposeful adjustment, these become policy variables. These can be at the household level (deciding to save more), firm level (increasing advertising), or at the government level, such as raising taxes or by a central bank engaging in open market operations to affect interest rates. Of course, many independent variables are not subject to manipulation and one task in any model building exercise is to identify those variables that are subject to human intervention and those that are immutable.

1.1. Projections and forecasting

Projection is the process of using the parameters of a model to extend independent variables to calculate the outputs (dependent variables) of a model. Such projections are conditional on hypothesized values of the inputs (independent variables), and we assume the parameters of the regression have no error. If one knows that Portage la Prairie is 66 km from the Perimeter on Highway 1, then driving at 100 km per hour

Module 13 Forecasting, Simulation, and Monte Carlo Analysis

means that you will arrive in approximately 40 minutes (39.6 to be exact). This is a simple projection based on a formula from physics.

Forecasting is a more general term than projection and explicitly incorporates uncertainty into predictions of future. The errors of the regression model are a case in point. Figure 1 shows the distinction between projection and regression-based forecasts, which have an increasing band of uncertainty as the forecast moves beyond that data used to estimate the regression.

Time series data are by far the most common bases for such projection, but cross-sectional data may also serve as a basis for projecting the results of one group to another. The term *generalizability* or *external validity* refers to the applicability of a model to other times, places, and people.

Projection tends to be mechanistic. A typical example might use a regression of GDP on independent variables such as interest rates to project future values of GDP based on hypothesized future values of interest. Such calculations will implicitly reflect the errors from the regression model and the errors made in the assumed future values of interest. Most statistical software packages have forecasting options built in. Excel does not have these options embedded, although add-on packages do exist.

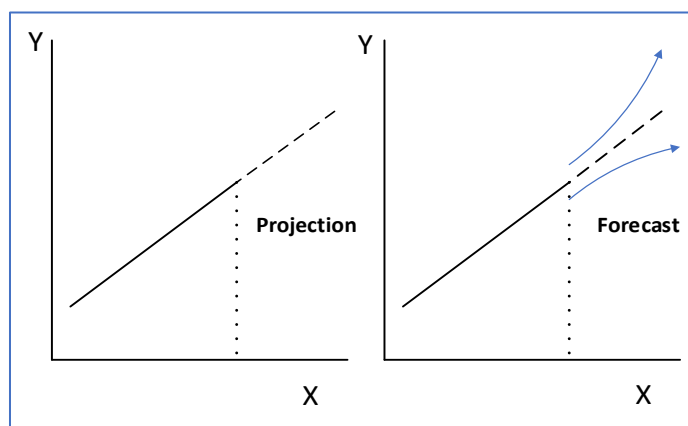


Figure 1: Projections and forecasts

1.2. Simulation and scenario design

Simulation and scenario design represents an evolution of forecasting used in financial economics, policy design, and academic research. For example, when applying for a mortgage, most financial institutions will use hypothetical future higher interest rates to ensure that the client can maintain the monthly payments. Since the Financial Crisis of 2009, governments, and central agencies, such as the International Monetary Fund, subject banks to regular “stress tests” to assess their capacity to withstand monetary shocks.

Simulation starts with an abstract, usually mathematical model. The basic demand and supply models of microeconomics show the implications of all manner of policy. Often these use simple geometry, but in Excel we can create more precision by specifying the exact mathematical forms and simulate the revenue generated by various levels of excise tax on cannabis.

Simulation comes in two forms – deterministic and stochastic.

- *Deterministic models* represent the approach used to this point in the course. Mathematical formulas (Gini, demand/supply, taxation, and net present value) support the calculation of

Module 13 Forecasting, Simulation, and Monte Carlo Analysis

consequences with certainty. Changing the model parameter through a range, termed “sensitivity testing,” is what a bank does when it evaluates an applicant’s fitness for a loan.

- *Stochastic models* reflect messy reality with probability functions replacing coefficients to yield not a single (point) forecast.

This Module starts by examining a basic forecasting structure based on the regression model. Then, a gambling model demonstrates the basic form of a probabilistic model that serves as the foundation for Monte Carlo methods.

2. Forecasting using regression

Simulation covers a range of activities in economics and business. A forecast using a trend line is the simplest simulation that assumes the future will track linearly based on the value of a_1 .

$$Y_t = a_0 + a_1 t \quad 13.1$$

where t represents any time unit. This model serves as a very elementary baseline, the critical problem being that time is not really a proper independent variable. Change over time represents other biological, social, and economic processes that should enter the model directly as independent variables.

$$Y_t = a_0 + a_1 X1_t + a_2 X2_t + u_t \quad 13.2$$

Often, dummy variables signifying seasonality and a lagged value of the dependent variable can improve the fit of the forecast, such as:

$$Y_t = a_0 + a_1 X1_t + a_2 Y_{t-1} + a_3 D1_t + a_5 D2_t \dots + a_{14} D11_t + u_t . \quad 13.3$$

The regression with independent variables other than time (t), dummy variables, or lagged values of the dependent variables suggest a causal relationship. Use the basic consumption function $C_t = a_0 + a_1 Y_t$ as an example. Now $C_{t+1} = a_0 + a_1 Y_{t+1}$ is a conditional forecast of C_t , based on values of Y_t supplied by the analyst. Simulation requires the projection of Y_t to Y_{t+1} , Y_{t+2} The simplest approach is just to project the trend line on Y_t , but that does not really improve on trend, as shown in equation 13.2. A more sophisticated model projects one or more growth rates for Y_t where each growth rate represents a “scenario.” This one explores a range of

$$C_t = a_0 + a_1 Y_t, \text{ for } t = 2, \dots, N, \text{ and for } a_1 = \{.5, .1, .15, .2\} \quad 13.4$$

Here, the projection yields four scenarios. The model from 13.3 also supports simulation using a range of values for a_1 and a_2 . Changing the coefficients on the dummy variables (a_3, \dots, a_{14}) has little value, as including these variables serves mostly to improve the estimates on $X1_t$ and Y_t . Note also that Y_{t-1} is now an independent (exogenous) variable, while Y_t remains as the dependent or endogenous variable.

$$\text{Retail}_t = a_0 + a_1 \text{Wages}_t + a_2 \text{Unempl}_t + u_t$$

Module 13 Forecasting, Simulation, and Monte Carlo Analysis

Finally, when a_1 and a_2 represent draws (samples) from a probability distribution, and we repeated select values, the resulting values of C_t also take on a probability distribution.

See [Example 13_1.xlsx](#)

See [Exercise 13_1.xlsx](#) and [Exercise 13_2.xlsx](#)

3. Excel as a probability modelling system

Excel offers data scientists a sound platform to explore a range of probability problems. Although other systems exist that support sophisticated modelling – R, Matlab, Mathematica – Excel’s wide implementation means that many will be able to understand what you are doing and collaborate in the analysis.

A **probability model** specifies a rule that depends on

- a probability rule or generator; and
- outcomes related to the probability rule.

The simplest probability rule or generator is a coin toss, which, when coupled with a gain or loss associated with each outcome, creates a **stochastic process**.

Imagine you are playing a coin toss game. When a head comes up, you get \$1, and when a tail occurs, you lose \$1.

We can model this using Excel. The text provides three examples that demonstrate the function **=RAND()** that forms the basis for simple related probability models –

See

[Simple Gambling Game.xlsx](#)

[Coin Toss Model 1.xlsx](#)

[Coin Toss Model-Experiment.xlsx](#)

These models show how gambling can produce “runs.” So, when someone’s cumulative wins keep increasing, they are likely to say that they are on a lucky streak, when you know that they are equally likely to lose all their money just as quickly. In fact, luck has nothing to do with it, probability rules!

	A	B	C	D	E
		Rand()	Test for Head/Tail	Head = +1, Tail=-1	Cumulative Winnings
1	Trial				
2	1	0.935357	T	-1	-1
3	2	0.255058	H	1	0
4	3	0.950461	T	-1	-1
5	4	0.361637	H	1	0
6	5	0.448226	H	1	1
7	6	0.207381	H	1	2
8	7	0.930727	T	-1	1
9	8	0.861331	T	-1	0
10	9	0.697351	T	-1	-1
11	10	0.911884	T	-1	-2

Figure 2: Simple Game

To review, the **=RAND()** function produces a number between 0 and 1 based on the uniform distribution. Any number between 0 and 1 has an equal chance of selection, which means we can use this number with the formula **=IF(cell <.30,H,T)**. (See Figure 3, **next page**). A second **=IF** sets the value for head =+1 and tails =-1 (we could combine these two statements). Finally, the last column cumulates the winnings, resulting in a plot of the entire game. A game comprises one or more trials or coin tosses.

Module 13 Forecasting, Simulation, and Monte Carlo Analysis

Turn automatic calculation off (under File -> Options-> Formulas)). Toggle F9 to play a game.

	A	B	C	D	E
1	Trial	Rand()	Test for Head/Tail	Head = +1, Tail=-1	Cumulative Winnings
2	1	=RAND()	=IF(B2<0.5,"H","T")	=IF(C2="H",1,-1)	=D2
3	2	=RAND()	=IF(B3<0.5,"H","T")	=IF(C3="H",1,-1)	=E2+D3
4	3	=RAND()	=IF(B4<0.5,"H","T")	=IF(C4="H",1,-1)	=E3+D4
5	4	=RAND()	=IF(B5<0.5,"H","T")	=IF(C5="H",1,-1)	=E4+D5
6	5	=RAND()	=IF(B6<0.5,"H","T")	=IF(C6="H",1,-1)	=E5+D6
7	6	=RAND()	=IF(B7<0.5,"H","T")	=IF(C7="H",1,-1)	=E6+D7
8	7	=RAND()	=IF(B8<0.5,"H","T")	=IF(C8="H",1,-1)	=E7+D8
9	8	=RAND()	=IF(B9<0.5,"H","T")	=IF(C9="H",1,-1)	=E8+D9
10	9	=RAND()	=IF(B10<0.5,"H","T")	=IF(C10="H",1,-1)	=E9+D10
11	10	=RAND()	=IF(B11<0.5,"H","T")	=IF(C11="H",1,-1)	=E10+D11
12					
13					

Figure 3: Code for Simple Game

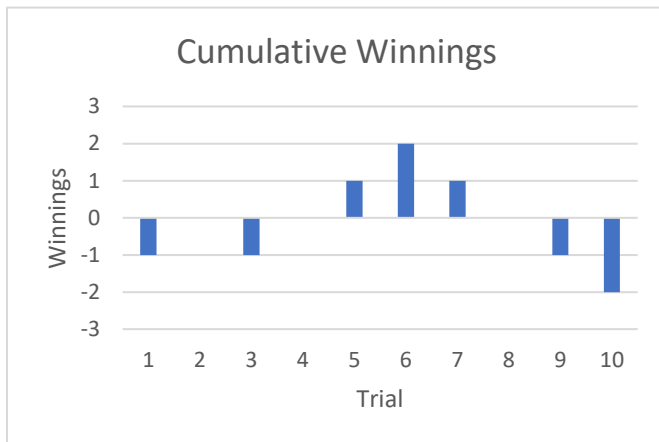


Figure 4: Simple Game - Trial 1

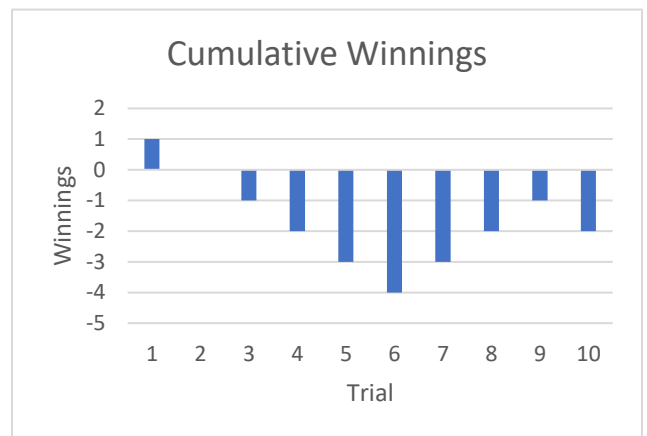


Figure 5: Simple Game – Trial 2

Coin Toss Model.xlsx elaborates on the simple model by determining the aggregate winnings for 11 games. This model represents a Monte Carlo Analysis for the simple gambling model.

See [Exercise 13_3.xlsx](#)

Excel probability models typically use the triangle, normal, lognormal, and Poisson distributions. More exotic models employ special probability distributions, such as the Weibull.

Video: [Some Excel tricks to support Monte Carlo Modelling](#)

3.1. Additional distributions for Monte Carlo Analysis: Weibull

Module 5 has presented the uniform, normal, lognormal, and Poisson distributions. The Weibull (“weebull”) distribution finds wide application in manufacturing and health. It is a tool to predict the time

Module 13 Forecasting, Simulation, and Monte Carlo Analysis

to failure in manufacturing or the time to death (or recovery) in health. It is a very flexible distribution, taking on many shapes, which is why it is so popular in Monte Carlo Analysis.

Weibull

Though the mathematics of the Weibull appear formidable, it has simple forms. For the two-parameter version (there are one- and three-parameter versions), here is the probability density function:

$$f(x) = \frac{\beta}{\alpha} \left(\frac{x}{\alpha} \right)^{\beta-1} e^{-\left(\frac{x}{\alpha}\right)^\beta}, \quad 13.5$$

where $\beta (>0)$ is the shape parameter and $\alpha (>0)$ the scale factor. The cumulative distribution function (cdf) is

$$F(x) = 1 - e^{-\left(\frac{x}{\alpha}\right)^\beta}. \quad 13.6$$

Varying the parameters will produce the variety of shapes that make this distribution so useful. Using the time to failure concept, $\beta < 1$ indicates that the failure rate (proportion of manufactures that fail) falls, meaning that products fail prematurely and early; $\beta = 1$ suggests that failure is constant with time, and it is impossible to predict the point when failure will occur; and $\beta > 1$ is that failure rate increases, as with products subject to wear and tear. The expression $1/\alpha$ is the failure rate. The average time to failure is

$$\alpha \Gamma\left(1 + \frac{1}{\beta}\right) \quad 13.7$$

(Γ is the gamma function, but this is as complex as things need to be for this presentation.)

Of most interest is its expression in Excel.

=WEIBULL(x,β,α,TRUE/FALSE), where, like all other probability functions, TRUE will produce the cdf and FALSE the pdf.

Example: In an experimental treatment for cancer, clinical trials suggest that the distribution is Weibull with $\alpha = 200$ days, and $\beta = .6$. What is the probability that a patient will survive to a year?

$$=WEIBULL(365,.6, 200, TRUE) = .7618.$$

See Weibull.xlsx

Module 13 Forecasting, Simulation, and Monte Carlo Analysis

The mean time to death requires that formula 13.7 or $1000\Gamma(1+1/.6) = 200*\text{EXP}(\text{GAMMALN}(1+1/.6))$ or 300.91 days.

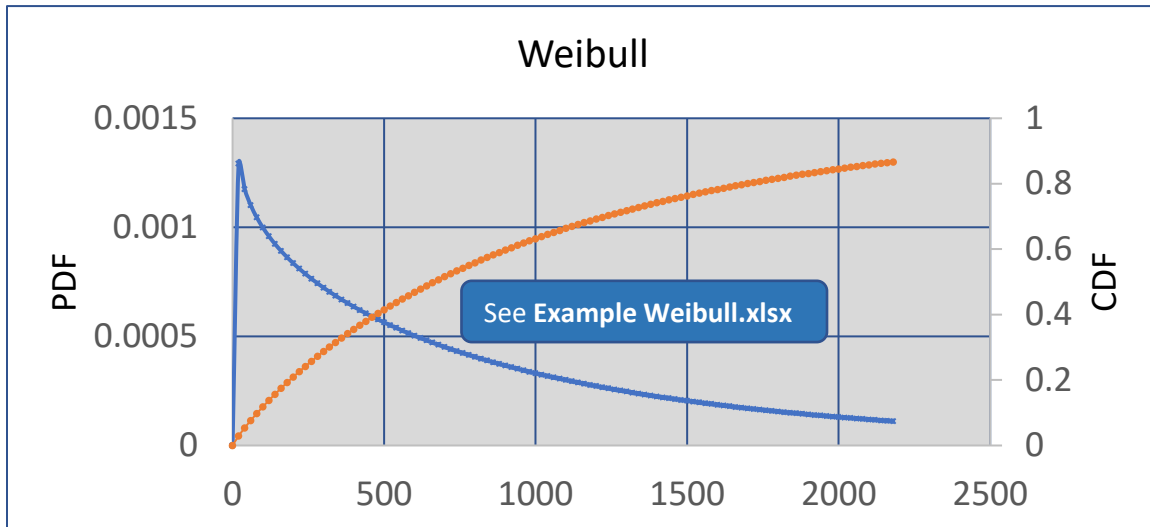


Figure 6: Weibull distribution, $B=.6$, $a=200$

The Excel implementation only offers the two-parameter version of the Weibull, which limits the pdf to various forms of skew right (sharp peak on the left).

Example Crypto Mining: You have decided to become a Crypto miner and need 1,000 computers to remain in continuous operation for one year or 365 days before you replace them all. A manufacturer claims that a Weibull distribution with a scale parameter of 400 and a shape factor of 4 characterizes the failure time for its new Epsilon computer line. A competitor claims its new Omega computer has a failure time also with a Weibull distribution with a shape factor of 600 and a scale parameter of 2. If an Epsilon costs \$800 and an Omega costs \$1,000, what purchase strategy would you advise?

See [Exercise 13_4.xlsx](#) and [Exercise 13_5.xlsx](#)

See [Example Crypto Mining.xlsx](#) and [Interactive Weibull.xlsx](#)

Video: [Check Distribution](#)

Module 13 Forecasting, Simulation, and Monte Carlo Analysis

3.2. Where do distributions come from?

Consider this a discussion of the “facts of life” applied to probability distributions.

1. The mathematical functions for the various distributions used in economic analytics have two essential features. First, the probability density function $f(x)$ lies between 0 and 1, for the specified range of x . The probability of any event must lie between 0 (cannot occur) to 1 (will certainly occur). For some distributions, such as the normal, the range of x is $-\infty < x < \infty$. For other distributions, such as the Weibull, x varies from 0 to ∞ , or the unitary (uniform) distribution, which has a range from a to b , defined for specific situations.
2. Data collected by observing industrial processes, such as medical/health records, client information, or sales creates the information base for thinking. Regression and other statistical methods identify the parameters of various linear and non-linear equations. The statistics used for such distribution estimation lie beyond the level of this text but estimating the form of the distributions used in a policy simulation is an important first step. While the parameters (mean, variance, shape factor) are hypothetical for the examples used here one would estimate these from real data collected from multiple sources.

Portraying data in histograms offers the first insights into how the shape of a distribution might fit the data collected. Since the distribution has a mathematical form, one can use regression and other statistical techniques to determine whether the data fit the model. Repeated data collection and statistical testing will produce increasing confidence that specific distribution characterize particular situations.

Example: The normal distribution characterizes many human attributes. Intelligence is the most common attribute associated with the normal distribution.

Example: Natural resources often have lognormal distributions. Few large oil reserves exist. Most of the reserves are smaller with a skewed distribution.

See: Normal Distribution.xlsx

See Interactive Weibull.xlsx

See Check Norm.inv

The simplest Monte Carlo simulations use repeated draws from the uniform or normal distribution. Other common distributions and their Excel implementation appear in Table 1.

Table 1: Common distributions in Monte Carlo Analysis		
Distribution	PDF $f(x)$	Excel expression
Exponential	$\exp(-\lambda x)$	=LN[RAND()]/ λ
Uniform	$1/X$	=RAND()* X also = =RANDBETWEEN(upper, lower)

Module 13 Forecasting, Simulation, and Monte Carlo Analysis

Table 1: Common distributions in Monte Carlo Analysis		
Distribution	PDF f(x)	Excel expression
Normal	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-u}{\sigma}\right)^2}$	=NORM.INV(Rand(),μ,σ) μ= mean σ = SD
Lognormal	$\frac{1}{\sigma x\sqrt{2\pi}} \exp\left\{-\frac{(\ln x - u)^2}{2\sigma^2}\right\}$	= LOG.INV(RAND(),μ,σ) μ= mean σ = SD
Weibull	$\frac{\beta}{\alpha} \left(\frac{x}{\alpha}\right)^{\beta-1} \exp\left\{-\left(\frac{x}{\alpha}\right)^\beta\right\}$	=α{-LN(1-RAND())}^(1/β) α = scale parameter β = shape parameter

3.3. Data tables and scenario design

Imagine you decided to open vending machines to sell PPE (personal protective equipment). To start, you will only sell two types of masks. You will price the good-quality masks at 10 masks for \$20, which nets you \$12 profit per set. The better-quality mask has a price of 5 masks for \$20 and nets you \$14 per set. You are uncertain about the profit range because you are unsure what customers will choose when at the vending machine. A type of “what-if” analysis using **data tables** allows you to explore the profitability based on different ratios of good to better masks sold.

This set-up shows the titles (grey), the action cell (red), inputs (blue), intermediate calculations (yellow), and the result of the “what-if” calculations (green). The data table is the area A8:B12.

	A	B	C	D
		Number of mask packages in a vending machine	% sold of "better" masks	
1				
2		100	0%	
3			Masks of each type sold	Unit profit
4		Better mask	0	\$16
5		Good mask	100	\$12
6				\$1,200
7	Share of better masks	1200		
8	60%	1440		
9	70%	1480		
10	80%	1520		
11	90%	1560		
12	100%	1600		
13				
14				

and the result of the “what-if” calculations (green). The data table is the area A8:B12.

This extends to two dimensions, where both the share of mask quality and the unit profit may vary.

Video: [Data Tables](#)

See Data Tables.xlsx

Module 13 Forecasting, Simulation, and Monte Carlo Analysis

4. Epidemiology

Epidemiology is the study of disease patterns and progression. Infectious diseases such as COVID have a pattern of spread by population attributes and biological factors. COVID has triggered interest in epidemiological models, the simplest of which is the SIRD model (Susceptible – Infected – Recovered... Dead) shown in Figure 8. The idea behind the model is that, out of a total population N , the susceptible S

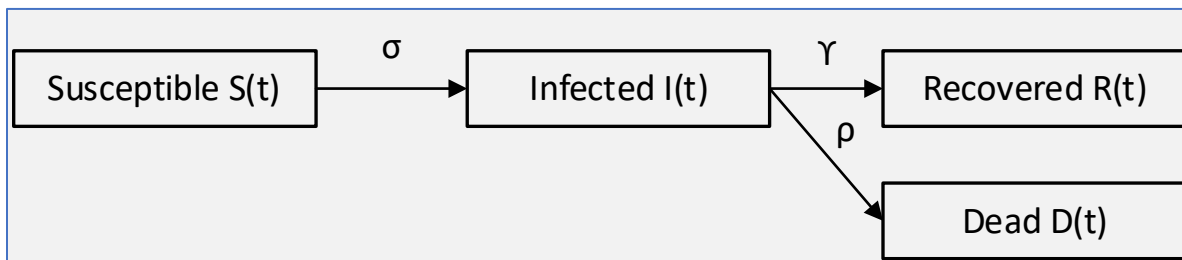


Figure 7: SIRD model

become infected I , then either recover or die. The model is dynamic, changing with time, making each of these values a function of t . Each of the model parameters forms part of a *rate of change* function (differential equation), showing the change from one period to the next:

Referring to Figure 8

- σ is the transmission rate (the rate at which those who are susceptible become infected)
- Υ is the recovery rate
- ρ is the death rate

First, the accounting identity stated that all components sum to the total population. (Let us ignore deaths for the moment.)

$$S(t) + I(t) + R(t) + D(t) = N$$

Like many models in economics, this accounting identity is the glue holding the model together. Recall the $Q_d = Q_s$ identity in microeconomics and $Y=C+I+G+(X-M)$ from macroeconomics.

Simulation models start from conditions. $S(0)$ is the number of susceptible people at $t=0$, $I(0)$ the number infected at $t=0$, and $R(0)$ the number recovered at $t=0$. What is interesting is how $S(t)$, $I(t)$ and $R(t)$ change over time, which is simply the first derivative of each elements $S(t)$, $I(t)$, and $R(t)$.

These are the four change equations of interest:

Module 13 Forecasting, Simulation, and Monte Carlo Analysis

$$\frac{dS}{dt} = -\sigma SI \quad \dots 1$$

$$\frac{dI}{dt} = \sigma SI - \gamma I \quad \dots 2$$

$$\frac{dR}{dt} = \gamma I \quad \dots 3$$

$$\frac{dD}{dt} = \rho \gamma I \quad \dots 4$$

The change in susceptible (dS/dt) is negative since it depends on how the infected and susceptibles interact. The term $S*I$ is the simplest description of this interaction.

The change in the numbers infected is (dI/dt), and this is growth in the numbers who become infected less than those who recover.

The change in recovered dR/dt is the proportion of infected who recover.

The change in the numbers who die is the death rate (ρ) times the numbers who recover (γI).

Notice how the three parameters (σ , γ , and ρ) enter the model. The recovery rate γ appear in three equations, σ in two, and ρ in only one. Work through the interpretation of each equation to ensure you understand the logic.

These differential equations require simultaneous solutions. A differential equation is simply a rate (dy/dx) that is equal to a specific formula, such as $dy/dx = x+y$. Euler's ("oilers") method is a basic numerical calculation process that avoids the need to solve the differential equation explicitly. (See Annex for a summary of Euler's method.)

See SIRD.xlsx

Equation 2 is of particular interest. It has two components $\sigma S * I$ and $-\gamma I$. Early in a pandemic, the first term is larger than the second, and dI/dt is greater than 0 ... and the infection curve rises. But as the number of recovered grows, fewer of the susceptibles become infected, and the second term dominates, making the rate of change negative and the infection curve falls. If we write Equation 2 to be at $t=0$, the result is

$$\frac{dI}{dt_0} = \sigma S_0 * I_0 - \gamma I_0$$

On the upside of the inflection curve, this is greater than 1, and on the downside, it is less than 1. The point where the numbers becoming infected equals the numbers recovering occurs when $dI/dt = 0$.

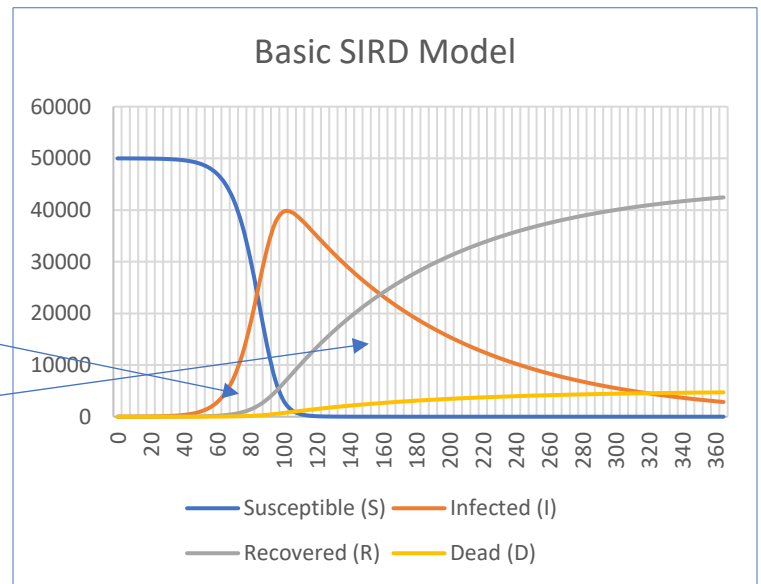


Figure 8: SIRD model

Module 13 Forecasting, Simulation, and Monte Carlo Analysis

$$\sigma S_0 * I_0 - \gamma I_0 = 0$$

$$\sigma S_0 - \gamma = 0$$

At the start of infection (the disease is spreading), more people are becoming infected than recover, or

$$\sigma S_0 - \gamma > 0, \text{ and when the disease is receding, } \sigma S_0 - \gamma < 0. \text{ This is the same as asking } \frac{\sigma S_0}{\gamma} < \text{ or } > 1.$$

The ratio $\frac{\sigma S_0}{\gamma}$ is the R_0 or “R – naught.” When it is greater than 1, the disease is spreading, and when less than 1, the disease is receding.

An extension of the SIRD model occurs with two additions: the addition of those exposed between the susceptible and infected, and the potential for those who recover to become reinfected. The common cold and influenza are two examples. This model does not include death as an outcome, but it could. The corona virus that causes colds has no vaccine (yet) and each year virologists need to anticipate the strains of influenza that will circulate to create the cocktail vaccine.

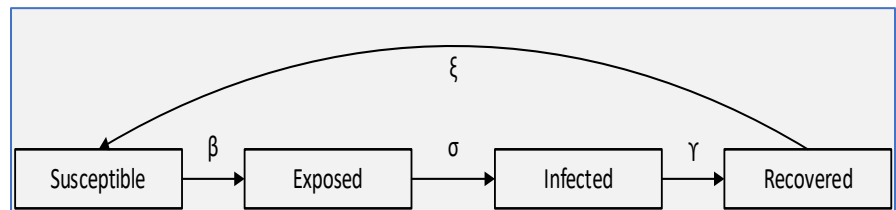


Figure 9 SEIR model

SEIR model (Susceptible –

Exposed - Infected - Recovered). The logic model and the associated parameters appear in Figure 11. Some diseases confer prolonged immunity on those infected, but others (such as malaria) do not, which creates the SEIR \bar{S} model, as show in the loop-back. The duration of immunity conferred by COVID remains unknown at this time.

Video: [SIRD Model](#)

The SEIR model has three parameters (four for the SEIRS) model:

- β determines the rate at which the disease spreads, or the probability of transmitting the disease from an infected person to a susceptible person
- σ is the incubation rate or the rate at which people become infected, where $1/\sigma$ or D is the average duration of the infection
- γ is the recovery rate and equals $1/D$
- ξ is the rate at which susceptible individuals become reinfected

This basic epidemiological structure supports a range of enhancements, especially the inclusion of behavioural concepts, such as mask mandates, lockdowns, and social distancing. The core challenge is developing the empirical basis for the parameters, especially with new diseases and the effects of policy, such as mask mandates, where compliance increases the impression. Many models proceed using judgment for the parameters, and simulations that result in fancy graphs can quickly become fanciful.

Module 13 Forecasting, Simulation, and Monte Carlo Analysis

5. Decision tree modelling

Complex decisions involve *risk* (known probabilities, such as a coin toss) and/or *uncertainty*, where probabilities are not known. Insurance works with probabilities of certain adverse events to create a balance between the fees collected across a large group who do not claim for the event and the small number who receive large payouts. Insurance mitigates the risk of adverse events, but it cannot manage uncertainty, which explains the existence of “acts of God” clauses that suspend the obligation to pay.

A decision tree uses a discrete step process for making choices when risks exist for costs and benefits. Each “node” represents an intermediate decision among two or more options. Any node without further options becomes an end state. Figure 12 presents a simple overview of the structure of a decision on a car purchase. Some consumers may start well into the decision tree, such as those who insist the next vehicle must be electric. The first node would be electric, followed by branches for car or truck, and then other branches to reflect additional choices within the set. When each branch has costs and benefits (losses and gains), a cost-benefit table becomes possible. When branches have probabilities, a Monte Carlo simulation becomes possible.

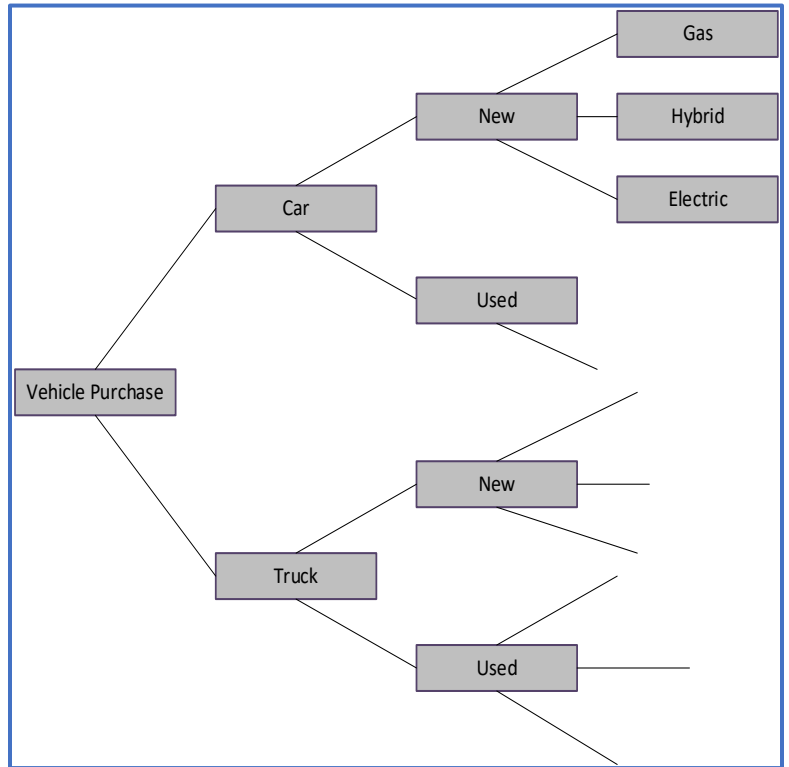


Figure 10: Decision tree

Module 13 Forecasting, Simulation, and Monte Carlo Analysis

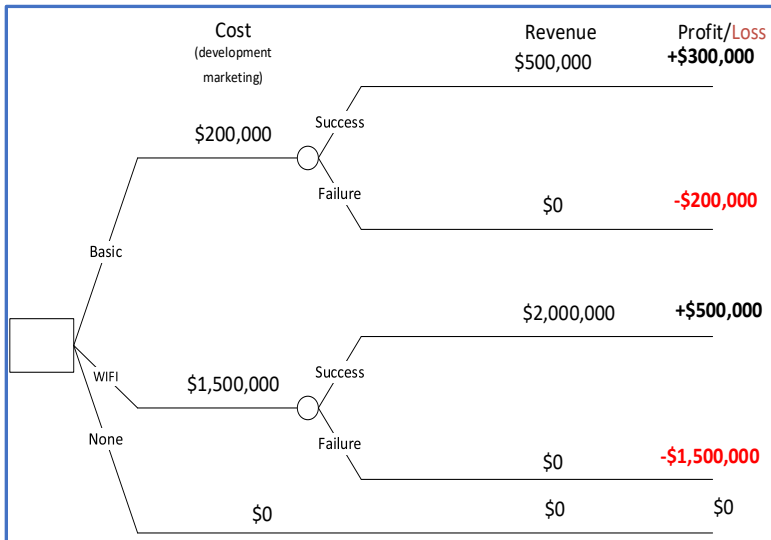


Figure 11: Product development for WI-FI

Decision tree analysis has some format conventions. Consider a decision on whether to manufacture two types of home security products. One is a basic perimeter alarm (with monitoring) and the other a WIFI perimeter and movement system that connects to a phone. Each has a development cost, and both have different sales expectations. The decision tree reveals the results of each of the three decisions. If each option has a probability (binomial), each of the three options has an expected return. Of course, the do-nothing option has a \$0 expected return.

Example: In Figure 13, assume the development of the Basic option has a 75% chance of success (25% chance of failure), while the WIFI option has a 25% probability of success. What is the best option? The expected value of the basic development option is +\$175,000 or $(.75 * \$300,000 + .25 * -\$200,000)$ and the expected value of the WIFI development option is -\$875,000 or $(.25 * 500,000 + .75 * -\$1,500,000)$. So, the investment in the WIFI option should not proceed.

Note that the Profit/Loss equals the Revenue less the Development/Marketing cost.

Module 13 Forecasting, Simulation, and Monte Carlo Analysis

5.1. Vaccination model

Disease progression represents a common application of decision trees. The core idea is that interventions in the form of diagnosis – treatment and prevention – represent methods of altering the course of disease.

Figure 14 shows how prevention in the form of cancer screening alters the course of disease. The benefit-cost model reflects the trade-off between the costs of screening (direct, as well as side effects from the screening) and the avoided costs of prolonged cancer treatment if the disease diagnosis occurs at a later stage. The benefit of screening is identifying cancer at an early, more easily treated, and less costly stage. The cost-benefit equation is comparing the costs of screening a large population,

only a few of whom will have cancer, against the reduced costs of treating fewer seriously ill patients, most of whom will have shortened lives. The increase in quality life years accruing to those treated at initial stages is a component of the benefit from screening.

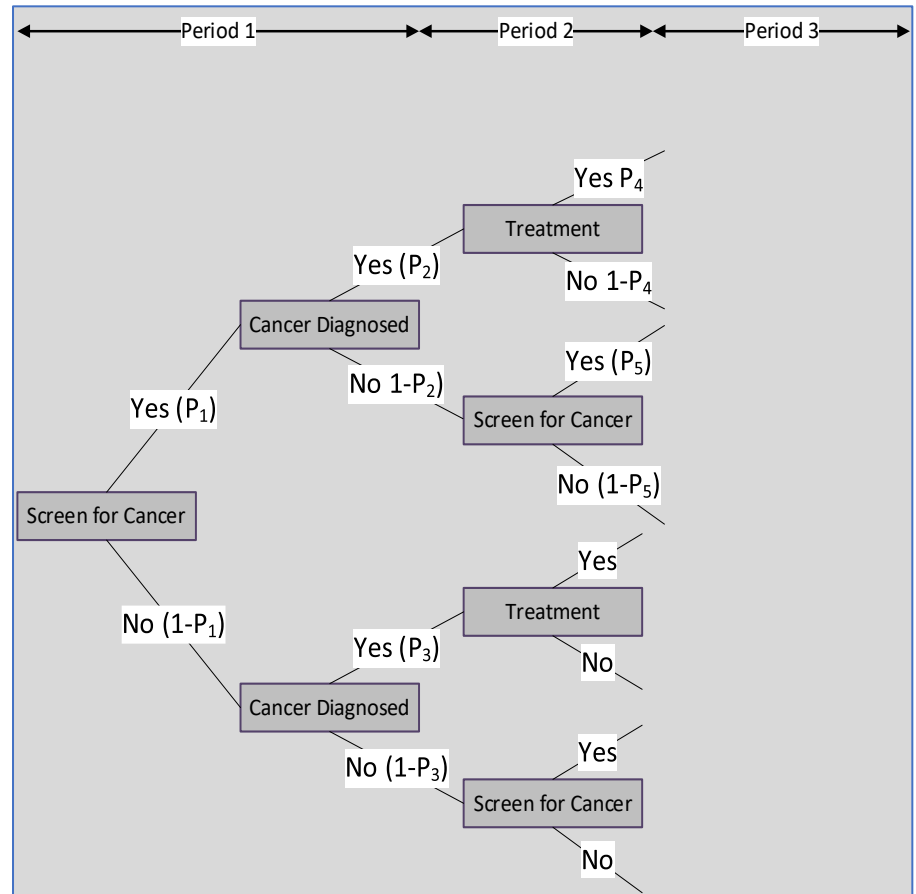


Figure 12: Cancer screening

In this example, the model cycles through three time periods; in applications, modelling may extend up to 30 periods. Each node has a binary choice, where the binomial distribution governs the choices. In theory, one can model multiple choice nodes through a sequence of binary choices, but this does not work. The binary choice structure is a limitation of decision trees. Finally, determining the probabilities requires a very robust data source. In healthcare insurance databases and where government centralises healthcare information, good estimates of probabilities become possible as consumers make decisions. Large sample surveys attempt to mimic the actual decisions through specialized surveys using discrete choice methods, where questions replicate the purchase decision.

Module 13 Forecasting, Simulation, and Monte Carlo Analysis

The vaccine decision model¹ shows the effect of a vaccination program on the expected value of the program administered over several years, assuming the probabilities in each year are independent. This creates the simplest decision structure. Typically, economists treat expected values as certain; once we know the risk (probability) of each return, the expected value results. If, however, risks are from a distribution, uncertainty enters, and the net expected benefit becomes contingent on the nature of the probabilities. Depending on the size of the pool of individuals and the nature of the distributions, the expected values may come quite close to the prediction from a standard model – or not.

Figure 15 presents the model, showing decisions (noted by a box) and contingency selection driven by the binomial distribution (shown by circles). The grey boxes show the costs, and the blue circles provide the probability that an epidemic begins. This appears complex. You lay out the decision tree starting in time = 0, the start of Year 1, where the choice is to have a vaccine program. Then the process unfolds over two years, as explained below. But to solve the decision, one uses backward induction, moving from right to left.

This model focusses on the costs of the program and assumes that everyone receives the vaccination if there is a program. The model does not consider vaccine hesitancy. The purpose of the example is to study the costs of vaccination over a two-year period.

¹ Adapted from Boardman, A.E., Greenberg, D.H., Vining, A.R., and Weimer, D.L., *Cost Benefit Analysis: Concepts and Practice*, 4th ed. Pearson, New Jersey, 2011.

Module 13 Forecasting, Simulation, and Monte Carlo Analysis

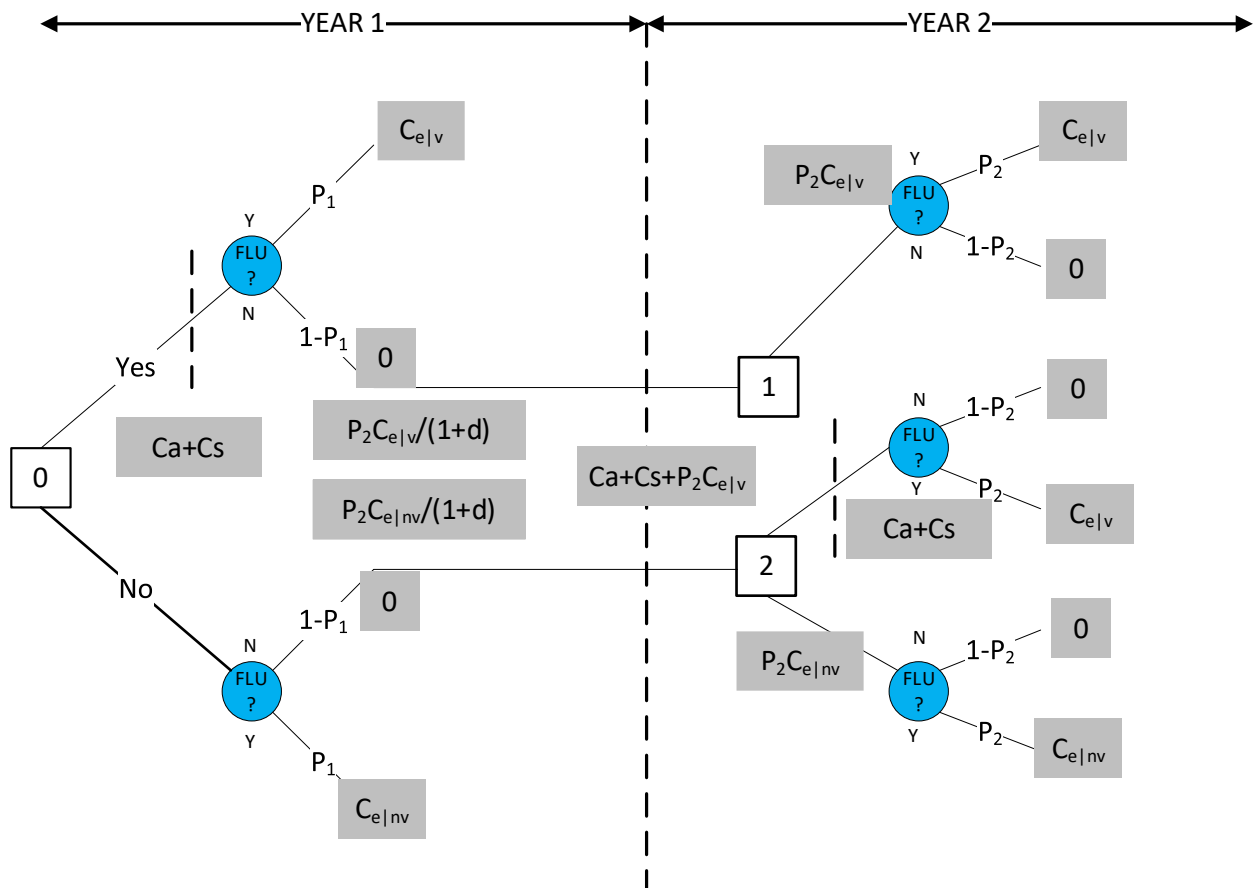


Figure 13 Flu Vaccine Decision Tree

To start at node 0 (the start of Year 1), a vaccine program either exists or does not. If yes, it incurs costs of C_a (administrative) and C_s , the costs of any adverse effects among the vaccinated. The vertical dashed line signals implementation of the vaccine program. Moving up and to the right, if the flu occurs (probability P_1), the costs ($C_{e|v}$) are the costs associated with the flu *given vaccination*. The population retains immunity for Year 1 – a strong assumption – and the process stops with no program needed in Year 2. If the flu does not occur, costs are 0, the model moves to Year 2 and decision node 1.

Moving back to node 0, if no vaccination and no flu, the costs are 0 with a probability of $1-P_1$ and the process moves to Year 2 and node 2. If the flu occurs, the costs are $C_{e|nv}$, the costs of the flu *given that everyone is unvaccinated*. Since contracting the flu offers immunity for two years, the process stops.

At node 1, the process repeats. Since everyone has received a vaccination in Year 1, if the flu occurs (probability P_2), the costs are $C_{e|v}$. If the flu does not occur, the costs are 0.

At node 2, things are a little more complicated, but it repeats node 0. If the vaccine program now starts, (moving up and to the right), the implementation costs are C_a+C_s . If the flu occurs (with probability P_2),

Module 13 Forecasting, Simulation, and Monte Carlo Analysis

the costs are $C_e|v$. If no flu, there are no costs. If the vaccine program is not implemented at node 2, the tree moves down and to the right. Should the flu occur (with probability P_2), the costs are $C_e|nv$; if the flu does not occur, costs are 0.

To solve the total costs, start from the upper right, where the costs are $C_e|v$, conditioned by the probability or $P_2 * C_e|v$ plus the costs of no flu or $(1-P_2) * 0 = 0$. Cumulating the costs associated with node 2, the costs are $P_2 * C_e|v + (1-P_2) * 0$ plus the set-up costs of the program $C_a + C_s$. The total becomes $C_a + C_s + P_2 * C_e|v$. Finally, from the lower left, the costs are $P_2 * C_e|nv + (1-P_2) * 0$.

If at node 2, the costs of the program ($C_a + C_s + P_2 C_e|nv$) are greater than not implementing it, $C_e|nv$, then the cheaper option is to not implement the program.

$$C_a + C_s + P_2 * C_e|v > P_2 * C_e|nv$$

$$C_a + C_s > P_2(C_e|nv - C_e|v)?$$

If we assume the second equation, then it is cheaper to not implement the vaccine program in Year 2. The expected cost of the program at node 0 is

$$E(Cv) = C_a + C_s + P_1 * C_e|v + (1 - P_1) * P_2 * C_e|v / (1 - d). \text{ (Red path in Figure 14.)}$$

The expected cost of not implementing the program appears as

$$E(Cnv) = P_1 * C_e|nv + (1 - P_1) * P_2 C_e|nv / (1 + d). \text{ (Yellow path in Figure 14.)}$$

Module 13 Forecasting, Simulation, and Monte Carlo Analysis

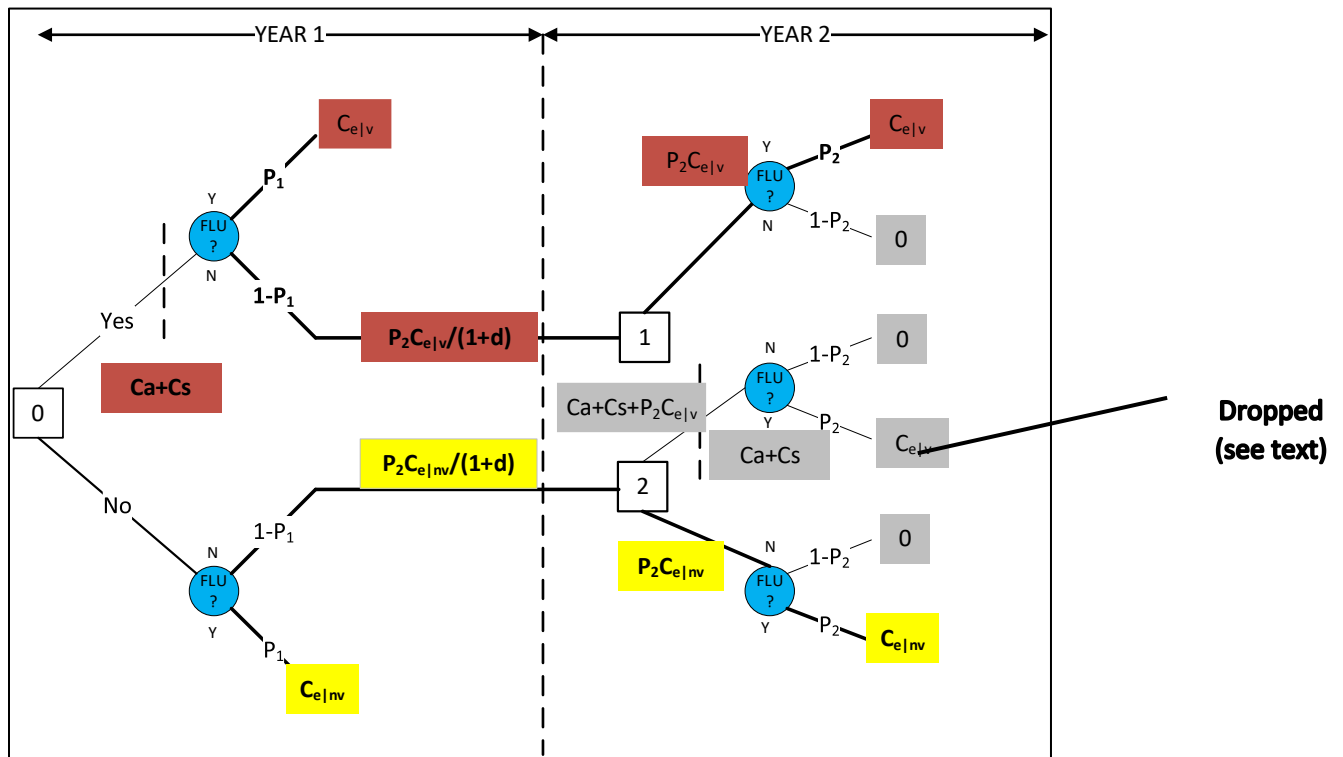


Figure 14: Vaccine model solved

Now the model needs parameters, which appear in the spreadsheet

Video: [Vaccine Model](#)

6. Conclusion

Simulation falls into two categories - deterministic (forecasting, scenario design, and sensitivity analysis) and probabilistic (commonly called Monte Carlo analysis). Using the probability functions in Excel supports an important extension for economic analytics. Many mathematical functions support simulations. The differential equations and decision tree models are just two structures available.

Module 13 Forecasting, Simulation, and Monte Carlo Analysis

Annex: Euler's method for solving differential equations

A differential equation relates a derivative (written as $\frac{dy}{dx}$ or $f'(x)$) as a function of x and other variables.

$$\frac{dy}{dx} = 2x$$

$$\frac{dy}{dx} = x^2 + 2$$

The usual "analytic" solution is to use integrals or anti-derivatives to find:

$$y = x^2$$

$$y = \frac{x^3}{3} + 2x$$

Then, when we need to find the value of Y_1 at some point X_1 , it is just a matter of inserting in the solved function.

Using the first example, Figure 15 shows the function $y = x^2$ and the value of $\frac{dy}{dx} = 2x$ for values of x .

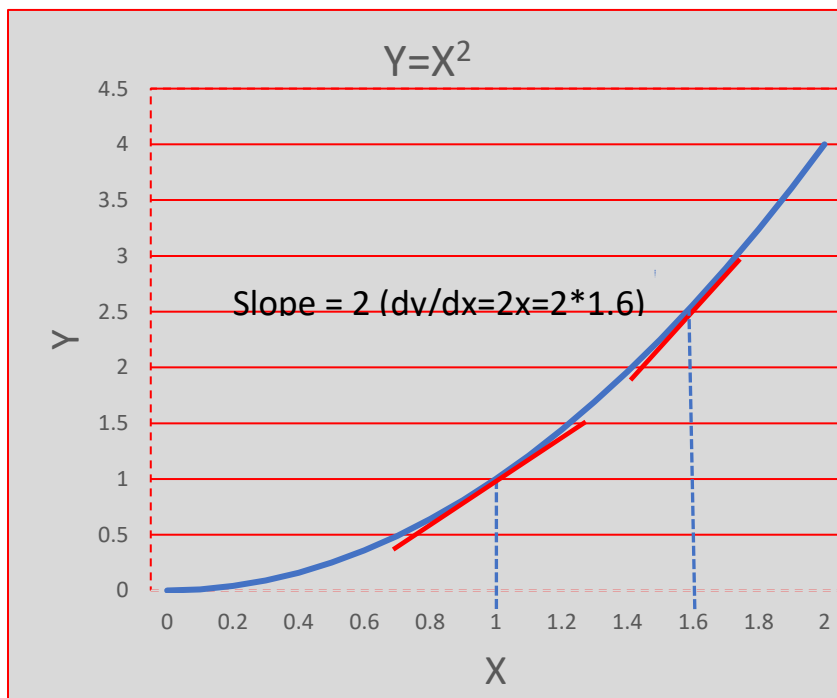


Figure 15: Tangents approximating curvature

At $X=1$, $dy/dx=2$ and at $X=1.65$, $dy/dx=3$. Note how the slope increases with X . Euler's insight was that, by taking successive values of the slope, one can trace the curve of $Y=X^2$ using a series of tangent lines. Each tangent line is the slope of $Y=X^2$ at a particular value of X .

Why not always use the analytical approach, since solving $dy/dx=2x$ to reach $Y=2$ seems simple enough? First, some differential equations do

Module 13 Forecasting, Simulation, and Monte Carlo Analysis

not have an analytic solution and numerical approximation is the only choice. Second, simulation modelling requires repeated calculation, as parameters vary to express a variety of behaviour assumptions. For example, $dy/dx=2x$ is easy to solve, but in a simulation model, the analyst may vary parameter (“2”) across a wide range, and it is inconvenient to recalculate the value of the tangent manually with each iteration of x .

Note that numerical solutions to differential equations are approximations. The discrepancy between the numerical estimated value of Y and the analytic (true) value depends on the “step” length.

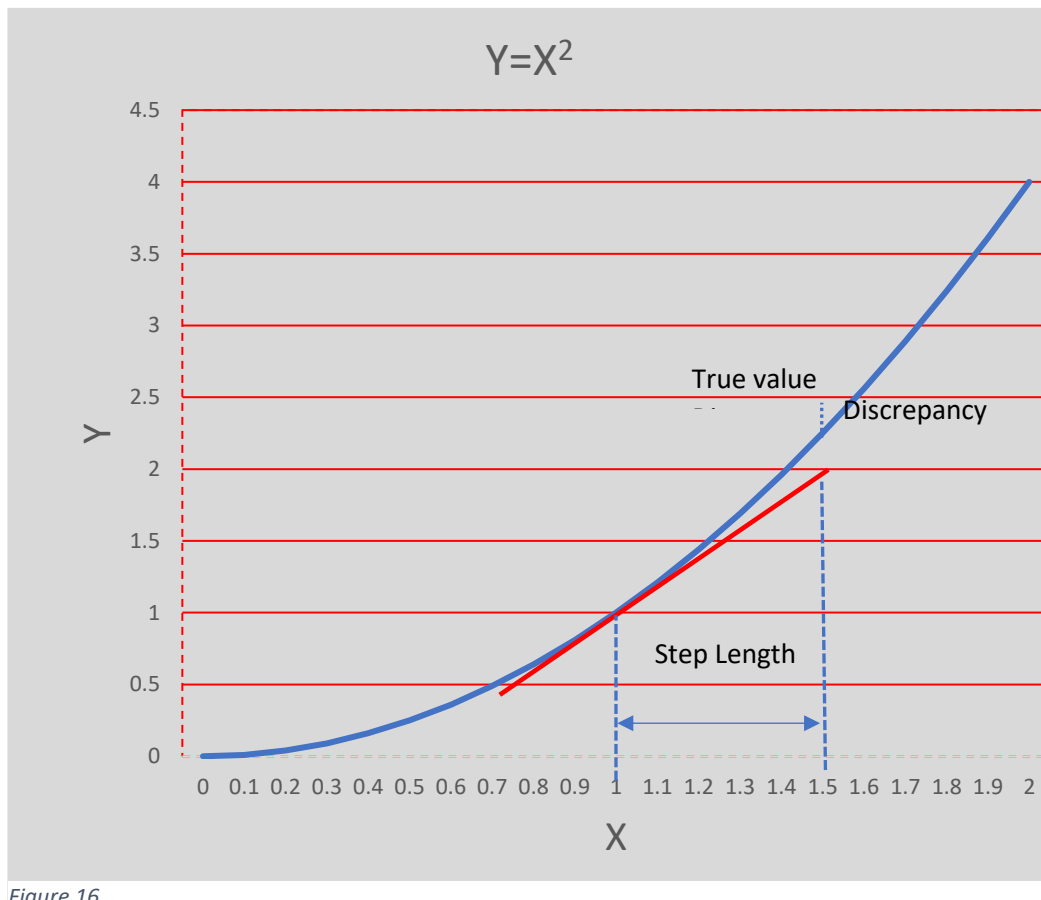


Figure 16

Module 13 Forecasting, Simulation, and Monte Carlo Analysis

The formal Euler's method uses the formula $Y_{1est} = Y_0 + h * f'(X_1)$. Referring to Figure 2, $Y_0 = 1$, $h = .5$, and $f'(X_0) = 2$, which makes $Y_1 = 2$. [$Y_{1est} = Y_0 + h * f'(X_1) = 1 + .5 * 2$]. Note that the true value of Y_1 is $1.5^2 = 2.25$. All of this appears in Table 2, which shows the calculations with a much smaller step size of .1.

Euler.xlsx

Trial	X	Y _{est}	Y _{actual}	Discrepancy
0	1	1	1	0
1	1.1	1.20	1.21	0.01
2	1.2	1.42	1.44	0.02
3	1.3	1.66	1.69	0.03
4	1.4	1.92	1.96	0.04
5	1.5	2.20	2.25	0.05
6	1.6	2.50	2.56	0.06

The SIRD pandemic simulation has three differential equations

$$\frac{dS}{dt} = -\sigma SI \quad \dots 1$$

$$\frac{dI}{dt} = \sigma SI - \gamma I \quad \dots 2$$

$$\frac{dR}{dt} = \gamma I \quad \dots 3$$

$$\frac{dD}{dt} = \rho \gamma I \quad \dots 4$$

Table 2: Excel calculation for curve approximation

Here the "x" variable is time (t) and note that it does not appear on the righthand side of any of these equations. Referring to SIRD.xlsx, here is the code for equation 1.

Module 13 Forecasting, Simulation, and Monte Carlo Analysis

Basic SIR Model					
Time - Days (t)	Susceptible (S)	Infected (I)	Recovered (R)	Dead (D)	
0	49990	5	0		
5	49986	8.49925	0.2475	0.0	
10	49979.9	14.4	0.7		
15	49969.0	24.6	1.4		
20	49950.6	41.7	2.6		

The step length is five days. We start at $t=0$ when $S=49990$, $I = 5$, and $R = 0$. J_9 is the transmission rate σ . The change dS/dt uses Euler's method. Note that equation 1 is negative, which means the curve is dipping down. Subtract the change from the previous value and repeat.

SIRD.xlsx