8. *Concavity and Convexity*

To say nonlinear science is akin to saying non-elephant zoology.

—John von Neumann

We now introduce nonlinear models and nonlinear functions. Nonlinear functions can curve downward or upward, they can form S-shapes, they can kink, jump, and squiggle. In time, we cover all of these possibilities. We start here with models that rely on convexity and concavity. We show how growth and positive feedbacks produce convexity and how diminishing returns and negative feedbacks produce concavity. Most disciplines contain models of both types. Economic production models assume that delivery and inventory costs decrease with a firm's size, making profits per unit sold a convex function of a firm's size, which explains why Walmart earns such large profits.[1] Economic models of consumption assume that the utility (or value) is concave, that we enjoy the fifth piece of pizza less than the first. In ecosystems, when a new species invades and confronts no predators, its population grows at a constant rate, producing a convex function. As that population grows, it has less food. Fitness, as a function of population size, is therefore concave.

The chapter consists of three parts. The first part covers models of population growth and decay. The second part covers concavity. In it, we see how concavity implies risk aversion and a preference for variety. In the third part, we study a series of growth models from economics that combines concave functions and linear functions.

Convexity

Convex functions have an increasing slope: the function's value increases by a larger amount as we increase a variable's value. The number of possible pairs of people is a convex function of the group size. A group of three people includes three unique pairs. A group of four people includes six unique pairs, and a group of five includes ten unique pairs. Each increase in group size increases the number of pairs by a larger amount.

Similarly, each time a chef adds a new spice to his repertoire, he increases the number of spice combinations by a larger amount.

Our first model of convexity, the *exponential growth model,* describes the amount of a variable, often a population or a resource, as a function of its initial value, a growth rate, and the number of periods.

Exponential Growth Model

A value of a resource at time t, V_t, that has an initial value of V_0 and grows at a rate R can be written as follows:

$$V_t = V_0(1 + R)^t$$

This single-equation model plays central roles in finance, economics, demography, ecology, and technology. When applied to finance, the variable is money. Using the equation, we can calculate that a $1,000 bond paying 5% annual interest increases in value by $50 in year one and by more than $100 in year twenty. To draw clean inferences, we assume a constant growth rate. Given that assumption, we can manipulate the exponential growth equation to derive the *rule of 72*.

Rule of 72

If a variable grows by a percentage R (less than 15%) each period, then the following provides a good approximation:

Periods to Double ≈ image

The rule of 72 quantifies the cumulative effect of higher growth rates. In 1966, Zimbabwe had a per capita GDP of $2,000, twice that of Botswana. Over the next thirty-six years, Zimbabwe experienced little growth. Botswana, meanwhile, averaged 6% growth, meaning that Botswana's GDP doubled every twelve years. In thirty-six years, it doubled three times, an 8-fold increase. Thus, in 2004, Botswana's per capita GDP of $8,000 was four times that of Zimbabwe.

This same formula reveals why housing bubbles must end and technological progress need not. In 2002, home prices in the United States rose by 10%. That would imply a doubling every seven years. Had that trend continued for thirty-five more years, prices would have doubled five times—a 32-fold increase. A house costing $200,000 in 2002 would cost $6.4 million in 2037. Prices cannot rise at that rate. The bubble had to burst. In contrast, *Moore's law* states that the number of transistors that can fit on an integrated circuit doubles every two years. Moore's law has persisted because spending on research and development has generated a near constant rate of improvement.

Demographers apply the exponential growth model to human populations. A population that grows at 6% a year doubles in size in twelve years. In thirty-six years, it doubles three times, and in one hundred years, it doubles eight times (increasing 256-fold). In 1798 British economist Thomas Malthus noticed that the population was growing exponentially and wrote a model showing that if the economy's ability to produce food only increased linearly, then a crisis loomed. The short version goes as follows: Population was growing like 1, 2, 4, 8, 16, 32,.… Food production was growing like 1, 2, 3, 4, 5,.… Malthus foresaw disaster. Fortunately, birth rates fell, and the arrival of the Industrial Revolution increased productivity. Had nothing changed, Malthus would have been correct. But he ignored the potential for innovation—the focus of models later in this chapter. Innovation subverted the trend.

The exponential growth model can be applied to the growth of species as well, and not just to rabbits. When you acquire a bacterial infection, tiny bacteria reproduce at incredible rates. Bacteria in human sinuses grow at around 4% a minute. By applying the rule of 72, we can calculate they double every twenty minutes. In a single day, each initial bacterial cell spawns over a billion offspring.[2] Their growth stops when the physical constraint of your sinuses leaves them no room. Food constraints, predators, and lack of space all reduce growth. Some species, such as deer in suburban America or the hippos brought to Colombia by drug lord Pablo Escobar, encounter few constraints on growth and their population grows rapidly, though not at bacterial rates.[3]

A convex function with a positive slope increases at an increasing value. A convex function with a negative slope becomes less steep. A convex function with an initially large negative slope will flatten. That is true for the equation in the *half-life model*, which captures decomposition, depreciation, and forgetting.

In the model, every *H* periods half of the quantity decays. Hence, *H* is known as the *half-life* for that process. For some physical processes, the half-life is constant. All organic matter contains two forms of carbon: an unstable isotope, carbon-14, and a stable isotope, carbon-12. In living organic matter, these isotopes are present in a constant ratio. When an organism dies, the carbon-14 in its body starts to decompose with a half-life of 5,734 years. The amount of carbon-12, on the other hand, does not change. Willard Libby, a physical chemist, realized that by measuring the ratio of carbon-14 to carbon-12, one can estimate the age of a fossil or artifact, a technique known as radiocarbon dating. Paleontologists apply radiocarbon dating to the remains of dinosaurs, woolly mammoths, and prehistoric fish. Archeologists use it to adjudicate claims of authenticity. The remains of Ötzi the Iceman, discovered in the Italian Alps, were estimated to be five thousand years old. The Shroud of Turin, first displayed in 1357 and claimed to be Christ's burial shroud, was found to date from the fourteenth century and not the time of Christ.

Half-Life Model

If every *H* periods half of the remaining quantity decays, then after *t* periods the following holds:

Proportion Remaining ≈ image

A novel application of the half-life model comes from psychology. Early psychological studies showed that people forget information at a near-constant rate. Our half-life of remembering depends on the salience of the event.[4] In 2016, the film *Spotlight* won the Academy Award for Best Picture. If people's memory of Oscar winners has a half-life of two years, in 2018, image of people will have remembered that fact, but by 2026, only image will recall it. The recollection of any particular event varies across people. Tom McCarthy, who directed and cowrote *Spotlight,* will likely never forget the year he won the Academy Award.

Concave Functions

Concave functions are the opposite of convex functions. Concave functions have slopes that decrease. Concave functions with positive slopes exhibit *diminishing returns*: the added value of each extra thing diminishes as we have more of that thing. Our utility or value from almost all goods exhibits diminishing returns. The more leisure, money, ice cream, or even time spent with loved ones, the less we value having more of it. Evidence for this can be found in the fact that the more we consume of just about anything, including chocolate, the less we enjoy it and the less we are willing to pay for it.[5]

Diminishing returns can explain a variety of phenomena, including why long-distance relationships are often so happy. If you see your partner just a few hours each month, every additional minute is wonderful. After a month of uninterrupted togetherness, the slope of the happiness curve flattens, and those few extra moments matter less.[6] It explains why developers invite people for free weekend visits to their beachfront condominiums. During a short weekend, you cannot get enough time on the beach. You are inclined to buy. After ten days on the beach, though, you may become bored.

When we assume concavity, we imply a *preference for diversity* and *risk aversion*. To show the former requires a concave function with multiple arguments. If our happiness is concave and increasing in both leisure and money, we prefer some leisure and some money to all leisure and no money or all money and no leisure. Risk aversion means a preference for a sure thing over a lottery. A risk-averse person prefers a certain payoff of $100 to a lottery that pays $200 half of the time and nothing the other half of the time. A risk-averse person prefers a double-dip ice cream cone to having either no ice cream or an unwieldy four-scooper.

Figure 8.1 shows why concavity implies risk aversion. The figure plots happiness for values for three outcomes: a high outcome (*H*), a low outcome (L), and the mean of those outcomes (M). Given the downward-shaped curve, happiness at the mean outcome exceeds the average happiness of the low outcome and the high outcome. The opposite holds for convex functions. Convexity implies *risk-loving*: we prefer the extremes to the average. The amount of a stock you can buy is a convex function of its price. Therefore, buyers of stocks prefer price volatility. If prices go up and down, buyers end up with more stocks than if prices stay constant.[7]

Economic Growth Models

We next construct a series of economic growth models. These models reveal the causes of growth and can explain and predict growth patterns across countries. They can also guide actions such as increasing the savings rate. To lay the foundation for our study of growth models, we introduce a standard economic production model in which output depends on labor and physical capital. Empirical evidence and logic support concavity of output in both labor and capital. Holding the amount of capital fixed, labor should be worth less as more is added. Similarly, adding more machines or computers adds less value given a fixed number of workers. Logic also suggests that output should be linear in scale. Doubling both the number of workers and the amount of capital should double output. A broom-making company with sixty workers and one factory that builds a second factory and hires sixty additional workers should double its output. The *Cobb-Douglas model*, one of the most widely used models in economics, includes both properties. Output is concave in labor and capital and linear in scale. This model can be applied to capture production by single firm or by an entire economy.[8]

image

Figure 8.1: Risk Aversion: Value(Mean) > Mean of the Values

Cobb-Douglas Model

Given L workers and K units capital, the total output equals:

Output = Constant · L^{a} $K^{(1-a)}$

where a is a real number between 0 and 1 capturing the relative importance of labor.

We use the Cobb-Douglas model to construct models of economic growth. To simplify, we assume 10,000 workers in the economy and ignore wages and prices, allowing us to focus on how the number of machines affects total output. We can then connect investment in capital to growth. To make the model as simple as possible, we assume that output takes the form of a single commodity, coconuts. The coconuts provide flesh and rich milk for food. However, the coconuts grow high in trees, so the workers require machines to pick them. We then make the very unrealistic assumption that the machines are constructed from coconuts. This simplifies the model but maintains the key trade-off between consumption today and investment in the future As a special case of the Cobb-Douglas model, we write output as the square root of the number of workers times the square root of the number of machines.

image

If the economy has one machine, output equals 100 tons. If people consume all 100 tons of coconuts, they invest in no new machines. Output will be unchanged in the next year. The economy exhibits no growth. If they invest 1 ton of coconuts to build a second machine, output increases to 141 tons, a 41% growth rate. If they build a third machine, output grows to 173 tons.[9] Through a constant investment, the economy grows at a decreasing rate. Output is a concave function.

Simple Growth Model

Production Function: $O(t)$ = 100image

Investment Rule: $I(t) = s \cdot O(t)$

Consumption-Investment Equation: $O(t) = C(t) + I(t)$

Investment-Depreciation Equation: $M(t + 1) = M(t) + I(t) - d \cdot M(t)$

$O(t)$ = output, $M(t)$ = machines, $I(t)$ = investment, $C(t)$ = consumption, s = savings rate, and d = depreciation rate

Now that we have the basic idea of how investment drives growth, we can construct a more elaborate model that includes an investment rule. We can write investment as *savings rate* times output and assume a fixed *depreciation rate* on the machines, such as that the number of machines that are no longer useful at the end of the year equals a fixed proportion of the number of machines. We can then write the total number or machines in the next year as last year's machines plus the investment in new machines minus the machines lost to depreciation. The complete *simple growth model* consists of four equations.

If we assume the economy has 100 machines, a savings rate of 20%, and a depreciation rate of 10%, output equals 1,000 tons of coconuts, consumption equals 800 tons, and new investment equals 200 machines. A total of 10 machines will be lost to depreciation, leaving 290 machines at the start of the new year. Similar calculations show that in the second year, outcome will equal 1,702 tons and in the third year it will equal almost 2,500 tons.[10] In the first three years, output increases at an increasing rate. This initial convexity is a result of the small number of machines in the first few years implies almost no effect of depreciation. Over time the number of machines grows and depreciation starts to matter making output concave. In the long run it ceases altogether, as shown in figure 8.2. By analyzing the model we can see why. Investment is linear in output: the number of new machines added grows linearly with output. Output is concave in the number of machines, so as the economy grows, investment will also be concave in the number of machines. Depreciation, though, is linear in the number of machines, and eventually the linear depreciation catches up with the concave increases in production.

image

Figure 8.2: Output in the Basic Growth Model for One Hundred Years

In the *long-run equilibrium* of the economy the number of new machines created by investment equals the number lost to depreciation. In our model, the equilibrium occurs when economy has 40,000 machines and produces 20,000 tons of coconuts. At that point, the economy invests 20% or 4,000 coconuts, in new machines and loses exactly that many machines to depreciation (10% of the 40,000). Thus, the number of new machines lost to depreciation equals the number of new machines created through investment and growth stops.[11]

The Solow* Growth Model

We now construct a more general model that is a simplification of the *Solow growth model* (thus the asterisk). We replace machines with physical capital and include labor as a variable. We also add a technology parameter that increases output linearly. Innovations increase this parameter. As in the previous model, the long-run equilibrium occurs when investment equals depreciation. Here, though, the equilibrium-level output depends on the amount of labor and on the technology parameter, as well as the savings and depreciation rates.[12]

Solow* Growth Model

Total output in the economy is given by the following equation:

Output = Aimage

where L denotes the amount of labor, K denotes the amount of physical capital, and A represents the level of technology. The long-run equilibrium output, O^*, is given by the equation[13]

image

Long-run equilibrium output increases in the amount of labor, the growth in technology, and the growth in the savings rate. It decreases with a rise in the depreciation rate. None of these results is surprising. More workers, better technology, and more savings increases output, and faster depreciation reduces output. The fact that output increases linearly with labor and savings is less intuitive. Labor produces diminishing returns, so without working through the model, we might expect long-run output to be concave in the amount of labor. However, as the amount of labor increases, so too does output, which in turn increases investment, leading to more output. The positive feedback from investment exactly offsets the decreasing returns. Last, equilibrium output is convex with the depreciation rate. Lowering the depreciation by 20% increases output by 25%.

Finally, long-run equilibrium output increases as the square of the technological improvements. Innovation therefore increases output more than linearly. We can use the model to show why. If we start with an economy in a long-run equilibrium and increase the technology parameter by 50%, output increases by 50%, and so too does investment. Investment then exceeds depreciation, so the economy continues to grow. Investment continues to outpace depreciation until the economy has grown another 50%, at which point the capital lost to depreciation offsets investment. These calculations reveal that innovation has two effects, creating an *innovation multiplier*. First, innovations directly increase outputs. Second, they indirectly lead to more capital investments creating an additional increase in output. Innovation, therefore, is the key to sustaining growth.[14]

These increases in output do not occur instantaneously. When a breakthrough occurs, the technology parameter changes slowly. The direct effects unfold over time. Old physical capital must be replaced by new physical capital with the better technology. A company’s computers do not get faster when technology changes; they get faster when technology changes and the company buys new computers. The second-order increase that results from the increased investment in physical capital takes place over an even longer time frame. Lags between technology and its effects on growth can imply that an innovation produces growth over a period of decades. Trains were invented in the early 1800s. The Gilded Age did not begin until the latter part of that century, a gap of over fifty years. The internet boom took place three decades after the creation of the ARPANET.[15]

Why Nations Succeed and Fail

We can apply our growth models to big policy questions such as whether backward countries can catch up, why some countries succeed and some fail, and the role of government in promoting growth. Those investigations show the value and limits of our models. We can begin with the ability of low-GDP countries to achieve fast growth. The models show that building up capital can produce fast growth, as will investing in technology. A backward country with less physical capital that could jump to the technological frontier with new capital outlays could experience incredible growth.[16]

The necessity of innovation for long-term growth, as shown in the second model, implies the limits of one-time imports of new technology. Continued growth requires innovation. Thus, when the Soviets dismantled German factories and rebuilt them in the Soviet Union following World War II, they could produce short-term growth, so much so that on November 18, 1956, Soviet premier Nikita Khrushchev, speaking at the Polish embassy in Moscow to ambassadors from Western nations, proclaimed, “Mi vas pokhoronim!,” or “We will be present at your funeral!” They did not. They failed to do so because the Soviet Union did not innovate.[17] They limited freedom and stifled entrepreneurs.

The models also show how extraction and corruption, the taking of output from the economy for government use, will reduce growth through reduced savings. Cross-country comparisons of growth rates support both findings: reducing extraction and corruption and promoting innovation enhance growth. Achieving those aims requires a strong but limited central government that promotes pluralism. The strong center establishes property rights and rule of law. Pluralism prevents capture by the elite, who often prefer the status quo and may not embrace innovation, which can be destructive.

As an example of destructive innovation, consider the website Craigslist, which posts for-sale and help-wanted ads. In the early 2000s, Craigslist contributed to the loss of hundreds of thousands of newspapers jobs in the United States. At that time, Craigslist itself employed only a few dozen workers. Though people lost jobs, Craigslist made the economy more efficient by increasing the technology parameter. In a less pluralistic society, the newspaper industry might have lobbied the government to stop Craigslist. Doing so would have slowed growth.

Japanese Chinese Economic Dominance

Linear model + rule of 72: From 1960 to 1970 Japan's GDP grew at a 10% annual rate. A linear projection of continued 10% increases would result in a doubling of the Japanese economy every seven years (using the rule of 72). In 1970, Japanese per capita GDP was approximately $2,000 in current US dollars. Had that trend continued, by 2012 per capita GDP would have doubled six times, resulting in a per capita GDP of $128,000.

Growth model: This model explains Japanese growth as due to investments in physical capital. The model predicts *concave* growth rates over time. The growth model predicts that as Japan's GDP approached that of the United States and Europe, its growth rate should decrease to the historical cross-country average of 1–2%.[18] The evidence supports this. From 1970 to 1990 Japan's GDP grew at around 4% annually. From 1990 to 2017, it grew at 1% or less.

Chinese growth: China's GDP grew at nearly a 10% rate from 1990 to 2010. In 2016, the per capita GDP in China reached approximately $8,000, and as predicted by the growth model, growth has slowed, with GDP growing at closer to 6% from 2013 to 2017. In China as well, sustained 10% growth rates run afoul of the rule of 72. If Chinese economic growth averaged 10% for the next century, per capita GDP would exceed $100 million.

It's a Nonlinear World After All

We construct nonlinear models because few phenomena of interest are linear. In this chapter we saw how diminishing and increasing returns are common features of economic, physical, biological, and social phenomena. We also saw some of the implications of including curvature in our models.

Most important, perhaps, we saw how functional forms structure our thinking and then how fitting functional forms to data allows us to make precise statements. Scientists can compute the age of artifacts using carbon-14 data. Economists can estimate the long-term effects of small increases in growth.

A central takeaway from this chapter is that intuition becomes insufficient once we include nonlinearities. Intuition tells us the direction of effects: growth is increased by a rise in savings, an increase in labor, and technological innovation. Models reveal the shape and form of those effects. Savings, as we would expect, have a linear effect. Increases in labor do as well in the long run, even though the model assumes short-run diminishing returns. Increases in innovation produce a multiplier effect: we get the square of those effects. The first increase is the direct effect of the innovation. The second increase in output arises from the increase in capital.

Insights such as these become clear with the help of models. Without models, we can usually infer what goes up and what goes down, but we lack understanding of the shape of functional relationships. As a result, we often make linear extrapolations—China's economy will soon take over the world. With models, we can better think through the logic that produces nonlinear effects. The set of nonlinear functions is enormous. The concave and convex models we covered in this chapter represent but a small dip in that vast sea. If we hope to improve our capacity to reason, explain, and act in a complex world, we need an even deeper dive into nonlinear phenomena.