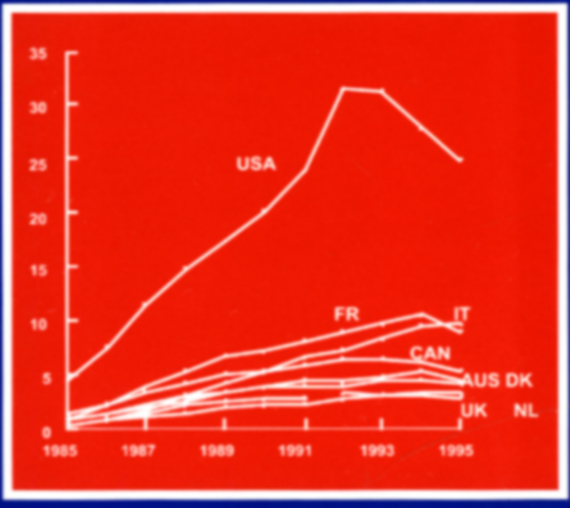
WILEY SERIES IN PROBABILITY AND STATISTICS

Empirical Model Building

Data, Models, and Reality

Second Edition



James R. Thompson

WILEY

This page intentionally left blank

Empirical Model Building

WILEY SERIES IN PROBABILITY AND STATISTICS

Established by WALTER A. SHEWHART and SAMUEL S. WILKS

A complete list of the titles in this series appears at the end of this volume.

Empirical Model Building

Data, Models, and Reality

Second Edition

JAMES R. THOMPSON

WILEY

A JOHN WILEY & SONS, INC., PUBLICATION

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

Library of Congress Cataloging-in-Publication Data:

Thompson, James R. (James Robert), 1938–
Empirical model building : data, models, and reality / James R. Thompson. — 2nd ed.
p. cm — (Wiley series in probability and statistics ; 794)
Includes bibliographical references.
ISBN 978-0-470-46703-9
1. Experimental design. 2. Mathematical models. 3. Mathematical statistics. I. Title.
QA279.T49 2011
519.5'7—dc23 2011013569

Printed in the United States of America

ePDF ISBN: 978-1-118-10962-5
oBook ISBN: 978-1-118-10965-6

10 9 8 7 6 5 4 3 2 1

To: Marian Rejewski and Herman Kahn,
Master Empirical Model Builders

Marian Rejewski cracked the Enigma Code in 1932,
sparing the United Kingdom strangulation by
German submarines in World War II.

The writings of Herman Kahn provided the basis for the
Reagan–Kohl Pershing II strategy which brought down the
"evil empire" without firing a shot.

This page intentionally left blank

Contents

This page intentionally left blank

Preface

That mathematics is not a science because it exits in a Platonic world of abstraction was well argued by the late John W. Tukey. Statistics, on the other hand, deals with data from the real world. Hence statistics can be claimed to be a science, to the extent that its practitioners focus on data analysis including model inferences based on data. Many scholars have believed that Tukey (himself a brilliant topologist) made a mistake in taking statistics out from under the umbrella of mathematics. Indeed, some members of the departments of statistics seldom, if ever, look at a set of data, contenting themselves with elegant data-free mathematical structures. Many a named chair has been awarded to a "statistician" who contented himself/herself with developing tests (seldom used in actual data analysis), and then proceeding to prove the asymptotic optimality of such tests under idealized conditions.

The father of Exploratory Data Analysis, John Tukey, took the position that individuals who avoid data from the real world, be they ever so elegant mathematically, are not practicing statistics but mathematics. Tukey went further and argued that those who did only apply to data standardized tests of models of other scientists were confining themselves unnecessarily to "confirmatory data analysis." He wanted statisticians to be more original than that. It was better if data analysis were done in an exploratory fashion. In other words, he wanted statisticians to be key players in science. They should examine data without relying too much on existing theories, and try to make inferences about real world systems.

Statistics is older than mathematics. Moses did carry out a census of the number of warriors in the Jewish tribes in 1452 BC. He made inferences from his census about the logistics of caring for his people and the military conquest of Canaan. Furthermore, I would submit that Thucydides, who wrote a fairly objective account of the Peloponnesian War between 431 BC and 411 BC (the war continued until 404 BC, but Thucydides apparently did not live to see the end of it), should be considered the father of time series analysis, an important subcategory of statistics.

Centuries later, geostrategist Herman Kahn (who was a major player in Monte Carlo simulation) argued from historical patterns and extrapolations how best to overcome the Soviet threat to the Free World. He was going further than Thucydides in that he was not only talking about qualitative facts in the past, but was extrapolating into an unknown future what would be the likely results of various strategies first to contain and then to destroy politically the Soviet Union. In other words, he engaged in Extrapolatory Data Analysis. The Reagan–Kohl Pershing II strategy was one almost taken right out of the pages of Kahn's numerous books.

Yet many scholars would argue that neither Thucydides nor Kahn could be considered statisticians.

In my own professional experience, I have had the good fortune of working in the building of practical models in such fields as oncology, tall building construction, manufacturing, epidemiology, military strategy, remote sensing, public policy, market analysis, nonparametric data-based estimation of functions, and others. Seldom have I been able to sit inside the box of standardized theories, statistical or otherwise. I always had to climb outside the box and use data to build appropriate models for dealing with the problem at hand.

The purpose of this book is not only to add to the arsenal of tools for using data to build appropriate models, but to give the reader insights as to how he/she can take his/her own direction for empirical model building. From my own experience in obtaining a bachelor's degree in engineering from Vanderbilt and a doctorate in mathematics from Princeton, I can say that I seldom use directly the tools learned in my university education. But it is also true that an understanding of those mechanisms has been of significant help in my career in data analysis and model building.

In Chapter 1, I consider topics in growth, including population growth, tax policy, and modeling tumor growth and its chemotherapeutic control. Parts of this chapter are simple from a mathematical standpoint, but the entire chapter has important implications for optimization and control. The so-called Malthusian theory is an example of a model that makes logical sense, absent data, but is seriously flawed when compared with reality. It is also noted that Malthus's ideas led to Social Darwinism and the unnecessary horror of the Irish potato famine.

Chapter 2 starts with an attempt to use data from the Old Testament to make inferences about the growth of the Jewish population starting with 1750 BC and ending with the fall of the Kingdom of David in 975 BC. Then we examine John Graunt's creation of the life table analysis of the demographic effects of the London plague of the sixteenth and seventeenth centuries. Then we delve into the data-based combat modeling started by Georg von Reisswitz and carried through with great success by Lanchester. Although easily computerized and data-modifiable, the U.S. Department of Defense, since the time of Robert McNamara has opted for non-data-based, noninteractive computer games such as Castforem. The work of the late Monte Carlo innovator, Herman Kahn, working without federal support in creating the strategy used by Reagan and Kohl to bring down the USSR, is also discussed.

Models for coping with contagious epidemics are presented in Chapter 3. We start with the laws of Moses concerning leprosy. We then proceed to John Snow's halting of the East End cholera epidemic of the mid-1800s. Each of these epidemics was more or less successfully dealt with sociologically. We then suggest that the AIDS epidemic that has killed over 600,000 Americans could have been prevented in the United States if only the simple expedient of closing the gay bathhouses had been followed. We note that sociological control of epidemics should not be neglected. The United States has much the highest number of AIDS cases per hundred thousand in the First World (over ten times that in the United Kingdom). The failure of the Centers for Disease Control to shut down the gay bathhouses is shown to be a plausible explanation of why AIDS continues to thrive in the United States. An argument is made that the United States may be "country zero" for the First World epidemic.

Chapter 4 deals with the bootstrap work of Julian Simon, Bradley Efron, and Peter Bruce. This computer-intensive resampling technique has revolutionized statistics. The classic zea mays data set from Fisher's *The Design of Experiments* is reanalyzed using various bootstrapping techniques. Then we deal with some rather unusual problems successfully handled by bootstrap techniques.

In Chapter 5 we show the importance of simulation in solving differential equations which do not admit of closed-form solutions (i.e., most of them). Particularly for partial differential equations in dimensions higher than three, simulation becomes virtually our only recourse.

I have often been approached by clients who wanted me to increase the size of their data sets by statistical means. That is generally not doable. However, the SIMDAT algorithm can build a continuous nonparametric density estimation base of pseudo-data around and within the neighborhood of an existing set which avoids such anomalies as suggesting that ammunition be stored in a section of a tank where there was no data but is shown from the nonparametric density estimator approach to be very vulnerable indeed. For many problems it is easy to write down the plausible axioms which have generated a data set. However, it is rarely the case that these axioms lead to a ready representation of the likelihood function. The problem is that the axioms are written in the forward direction, but the likelihood requires a backward look. The SIMEST algorithm allows a temporally forward approach for dealing with the estimation of the underlying parameters. SIMDAT and SIMEST are developed in Chapter 6.

Chapter 7 is a brief survey of the exploratory data analysis paradigm of John Tukey. He viewed statistics not just as a device by which models developed by nonstatisticians could be confirmed or rejected on the basis of data. He wanted statisticians to be on the cutting edge of discovery. He noted that exploration of data could be used to infer structures and effect inferences and extrapolations. EDA has greatly expanded the creativity horizons for statisticians as generalists across the scientific spectrum.

Chapter 8 is devoted to what I consider to be shortlived fads. Fuzzy logic and catastrophe theory have been shown to be inadequate tools. Chaos theory is not such a hot topic as formerly, but it still has its followers. The fads tend to build anti-Aristotelian structures not really sensitive to data. If we build a mathematical model that is not to be stressed by data, then we have entered the realm of postmodernism where everyone gets to create his/her own reality. Simply showing a mathematical structure does not indicate that structure conforms to anything in the real world. Nevertheless, I show that even if one does look at some of the artificial models of chaos theory, the addition of a slight amount of noise can frequently bring the chaotic model to something which does conform to real-world data bases.

Some professors of statistics believe that Bayesian data analysis is the only way to go. Bayesian theory has a lot to be said for it. For example, it gets around the claim of Karl Popper that statistics can only demolish hypotheses, never confirm them. Over the years, the use of noninformative prior distributions has seemingly weakened a major raison d'être of Bayesian analysis. In Chapter 9, the author attempts to give a data-based exposition of Bayesian theory, including the EM algorithm, data augmentation, and the Gibbs sampler.

There used to be surveys performed to decide who the most important living statistician was. Sometimes John Tukey would come in first. At other times it would be Edward Deming, the developer of Statistical Process Control. In Chapter 10 we go beyond the normal low-dimensional analysis advocated by Deming to show how higher-dimensional control charts can be constructed and how nonparametric as well as parametric tests can be used.

In Chapter 11 we investigate procedures where optimization may be readily implemented in the real world, which is generally noisy. We particularly emphasize algorithms developed by the statisticians Nelder and Mead which are amazingly useful in this age of swift computing. Lawera and Thompson built upon the piecewise quadratic optimization technique used in the rotatable experimental designs of the statisticians Box and Hunter.

Chapter 12 shows how no lesser a person than the author of *The Declaration of Independence,* Thomas Jefferson, persuaded President Washington to veto the rather reasonable and transparent allocation rule of Alexander Hamilton for the allocation of congressmen across the various states in favor of Jefferson's rule favoring the large population states (at that time Jefferson's Virginia was the largest). This first use of the Presidential veto is simply an example of the fact (pointed out by Jefferson earlier) that one should take the statements of politicians with a very large grain of salt. We show the basis of the utility theory of Bernoulli as well as that of Morgenstern and von Neumann. Finally, we present the Nobel Prize winning Impossibility Theorem of Kenneth Arrow which demonstrates the fact that group decisions which make everybody happy can never be constructed.

Chapter 13 is a brief practicum in sampling theory. The author has experience consulting in this area and believes that a great deal can be achieved by the use of transparent and relatively simple strategies.

In Chapter 14 it is shown how efficient market theory including capital market theory and the derivative approaches which proceed from the Nobel Prize winning work of Black, Scholes, and Merton are inconsistent with market data. The efficient market hypothesis has dominated finance, as taught in schools of business, for decades, much to the disadvantage of investors and society as a whole. As an alternative, it is demonstrated how computer-intensive and momentum-based strategies may be created which significantly best the market cap Index Fund strategies that proceed from capital market theory. Serious doubt is cast on the practice of the vending of uncovered call options. Most importantly, this chapter attempts to show young investors how they can develop their own strategies for purchasing stocks in an age where wars of choice (based on information later declared to be false), bad Federal Reserve policy, and the financing of houses to persons unable to pay off the mortgages have produced conditions where inflation becomes almost a certainty. The author has no magic rule for making the reader rich, but he gives the kind of information which is assuredly useful in showing him/her how to plan for using the market as a vehicle for obtaining a measure of financial security.

This work was supported in part by the Army Research Office (Durham) (W911NF-04-1-0354) *Some Topics in Applied Stochastic Modeling, Risk Analysis and Computer Simulation.* I would like to thank my mentor John Tukey and my colleagues Katherine Ensor, Scott Baggett, John Dobelman, Ed Williams, David Scott, Chapman Findlay, Jacek Koronacki, Webster West, Martin Lawera, Marc Elliott, Otto Schwalb, William Wojciechowski, Steven Boswell, Rachel MacKenzie, Neely Atkinson, Barry Brown and Ricardo Affinito. At John Wiley & Sons, I would like to thank my editor, Stephen Quigley.

Finally and most importantly, I wish to thank my wife, Professor Ewa Thompson, for her continuing love and encouragement.

JAMES THOMPSON

Houston, Texas
Easter 2011

Chapter 1

Models of Growth and Decay

1.1 A Simple Pension and Annuity Plan

It is almost always the case that everything we plan for is, in actuality, an approximation to what might happen. The distinction between "science" and "engineering" frequently has to do with the difference between models of Newtonian precision and approximations to a reality only partially understood. The fact is, of course, that Newton's laws are themselves approximations. It is true that we can much more accurately design an authomobile than we can plan an economy. However, though the author absolutely believes in objective reality, he understands that he is unlikely to find it in any physical or social system. As St. Paul said, "we see through a glass darkly,"(1 Corintians 13:12, King James Version). Of course some visions are cloudier than others. But, truth be told, if science is the study of precise reality and engineering is the study of approximations to that reality, then almost every scientist is actually an engineer. Frequently, physical scientists look down on the social sciences because the models in the physical sciences are more precise and more accurate. But, in reality, we are most of us empirical modelers doing the best we can, making logical inferences based on data, to understand what is going on and what will happen.

One easy introduction to the subject of growth models is obtained by considering an accounting situation, because such systems are relatively well defined. A pension and annuity plan is generally axiomatized clearly.

As a practical matter, pension funds in the United States are becoming much less generous than was the case in years past. A major reason for the distress of companies such as General Motors was the large liability built up

[0] *Empirical Model Building: Data, Modeling, and Reality,* Second Edition. James R. Thompson ©2011 John Wiley & Sons, Inc. Published 2011 by John Wiley & Sons.

over years to provide pensions and health care plans for their workers. For many small companies, there now simply are no pension funds beyond the federally mandated Social Security. A company is obliged to pay slightly more than 7% of a worker's salary into Social Security. The worker must match this. Insofar as the company's contribution is concerned, it is tax deductible. The worker's portion is not tax deductible. So an average worker is faced with the necessity of a "payroll tax" for income that is put somewhere in the maze of federal funds and is frequently spent out as though it were ready cash money. This is a kind of double whammy. In a sense, the worker is taxed on money that she does not receive. Then, when she starts to collect Social Security, a middle class employee is taxed a third time on 85% of monies received.

Typically, many employers are very concerned about the welfare of their workers. We recall that in 1908 Henry Ford instituted the 40-hour week and the minimum daily wage of $5. He also started at the same time a profit-sharing plan for his workers. He provided, at low mortgage rates, simple two-bedroom houses for his workers. We recall that this was 100 years ago when the majority of workers were not so blessed with so kindly an employer.

Unions tended to develop adversarial attitudes toward managers, whom they felt cared little about the welfare of the workers. Wage structures and benefit plans began to be increasingly divorced from economic reality. A plan that deferred some wage benefits into future retirement funds might be very attractive to a manager who was making promises that he would not have to meet in his own professional lifetime.

We will consider below a possible minimal pension fund. It is referred to as an Individual Retirement Account (IRA). Properly structured, the contribution of the worker and that of the employer are both tax deductible. At the time of retirement, the worker will receive monthly payments according to a mutually agreed upon plan. On these payments, he or she will pay taxes as though the annuity payments were ordinary income.

Suppose that we set up a pension plan for an individual who starts working for a firm at age N_1, retiring at age N_2. The starting salary is S_1 at age N_1 and will increase at rate α annually. The employer and employee both contribute a fraction $\beta/2$ of the salary each year to the employee's pension fund. The fund is invested at a fixed annual rate of return γ. We wish to find the value of the employee's pension fund at retirement. Furthermore, we wish to know the size of the employee's pension checks if he invests his pension capital at retirement in a life annuity (i.e., one that pays only during the employee's lifetime, with no benefit to his heirs at time of death). Let the expected death age given survival until N_2 be denoted by N_3.

Many investigators find it convenient to consider first a pilot study with concrete values instead of algebraic symbols. For example, we might try $S_1 = \$40,000; N_1 = 21; N_2 = 65; \alpha = 0.02; \beta = 0.0705; \gamma = 0.05; N_3 = 75$. The values of α and γ are rather low. The value of β is roughly half the value used at present in the U.S. Social Security System. The value of N_2 is the same as the present regular Social Security retirement age of 65. No allowance is made for annual losses from taxation, since pension plans in the United States generally leave the deposited capital untaxed until the employee begins his or her annuity payments (although the employee contributions to Social Security are taxed at full rate).

First, we note that at the end of the first year the employee will have approximately

$$P(1) = \beta S(1) = (0.0705)\$40,000 = \$2820 \tag{1.1}$$

invested in the pension plan. This will only be an approximation, because most salaried employees have their pension fund increments invested monthly rather than at the end of the year. We shall use the approximation for the pilot study.

At the end of the second year, the employee will have approximately

$$P(2) = \beta S(2) + (1+\gamma)P(1) = (0.0705)S(2) + 1.05(\$2820), \tag{1.2}$$

where

$$S(2) = (1+\alpha)S(1) = 1.02S(1) = 1.02(\$20,000) = \$20,400.$$

Thus, we have that $P(2) = \$5837$. Similarly,

$$P(3) = \beta S(3) + (1+\gamma)P(2) = (0.0705)S(3) + 1.05(\$5837), \tag{1.3}$$

where

$$S(3) = (1+\alpha)S(2) = 1.02S(2) = 1.02(\$20,400) = \$20,808.$$

So $P(3) = \$9063s$. By this point, we see how things are going well enough to leave the pilot study and set up the recurrence relations that solve the problem with general parameter values. Clearly, the key equations are

$$S(j+1) = (1+\alpha)S(j) \tag{1.4}$$

and

$$P(j+1) = \beta S(j) + (1+\gamma)P(j), \quad \text{for} \quad j = 1,2,\ldots,N_2 - N_1. \tag{1.5}$$

Moreover, at this point, it is easy to take account of the fact that pension increments are paid monthly via

$$S(j+1) = (1+\alpha)S(j), \quad \text{for} \quad j = 1,2,\ldots,N_2 - N_1, \tag{1.6}$$

and

$$P_j(i+1) = \frac{\beta}{12}S(j) + \left(1 + \frac{\gamma}{12}\right)P_j(i), \quad \text{for} \quad i = 0,1,2,\ldots,11, \tag{1.7}$$

N_1	=	starting age of employment
N_2	=	retirement age
S	=	starting salary
P	=	starting principal
α	=	annual rate of increase of salary
β	=	fraction of salary contributed by employee
γ	=	annual rate of increase of principal in fund
Year	=	N_1
Month	=	1
$^{**}P$	=	$\frac{\beta}{6}S + \left(1 + \frac{\gamma}{12}\right)P$
Month	=	Month + 1
Is Month	=	13?
If "no"	go to**	
If "yes"	continue	
Year	=	Year + 1
S	=	$S(1+\alpha)$
Month	=	1
Is Year	=	$N_2 + 1$?
If "no"		go to**
If "yes"	continue	
	Return P	

Figure 1.1. Subroutine annuity (N_1,N_2,S,α,β,γ).

where $P_{j+1}(0) = P_j(12)$. This system can readily be programmed on a handheld calculator or microprocessor using the simple flowchart in Figure 1.1. We find that the total stake of the employee at age 65 is a respectable \$450,298 (recall that we have used an interest rate and a salary incrementation consistent with a low inflation economy). We now know how much the employee will have in his pension account at the time of retirement. We wish to decide what the fair monthly payment will be if he invests the principal $P(N_2)$ in a life annuity. Let us suppose that actuarily he has a life expectancy of N_3 given that he retires at age N_2. To do this, we first compute the value to which his principal would grow by age N_3 if he simply invested it in an account paying at the prevailing interest of γ. But this is easily done by using the preceding routine with $S = 0$ and setting N_1 equal to the retirement age N_2 and N_2 equal to the expected time of death N_3. So we determine that using this strategy, the principal at the expected time of death is computed to be $P(N_3)$.

The monthly payments of the life annuity should be such that if they are immediately invested in an account paying at rate γ, then the total accrued principal at age N_3 will be $P(N_3)$. Let us suppose a guess as to this payment

$$\begin{aligned} X &= \text{guess as to fair monthly return} \\ P(N_2) &= \text{principal at retirement} \\ \gamma &= \text{annual rate of return on principal} \\ N_2 &= \text{retirement age} \\ N_3 &= \text{expected age at death given survival until age } N_2 \\ P(N_3) &= P(N_2)\left(1+\frac{\gamma}{12}\right)^{N_3-N_2} \end{aligned}$$

** Call Annuity $(N_2, N_3, X, 0, 6, \gamma)$
Compare P with desired monthly payout
Make new guess for X and return to**

Figure 1.2. Program trial and error.

is X. Then we may determine the total principal at age N_3 by using the flowchart in Figure 1.1 using $S = X, \alpha = 0, \beta = 6, \gamma = \gamma, N_1 =$ (retirement age) $N_2, N_2 =$ (expected age at death)N_3. We can then find the fair value of monthly payment by trial and error using the program pwreviously flowcharted in Figure 1.1.

We note that if the pensioner invests his principal at retirement into a fund paying 5% interest compounded monthly with all dividends reinvested, then at the expected death date he would have at that time a total principal of $741,650.

$$P(N_3) = \$450,298\left(1+\frac{0.05}{12}\right)^{(75-65)12} = \$741,650. \tag{1.8}$$

Now on a monthly basis, the pensioner should receive an amount such that if he invested each month's payment at 5%, then at age 75 his stake would have increased from $0 to $741,650. As a first guess, let us try a monthly payout of $5000. Using the program in Figure 1.2, we find a stake of $776,414. Since this is a bit on the high side, we next try a payout rate of $4,500—producing a stake at age 75 of $737,593. Flailing around in an eyeballing mode gets us to within one dollar of $P(N_3)$ in a number of iterations highly dependent on the intuition of the user. Our number of iterations was nine. The equitable monthly payout rate is $4,777.

It should be noted in passing that we have not taken into account the effect of inflation. What would the value of the first monthly payment be in today's dollars if the inflation rate has progressed at the rate of 3% per year? The value of the first months's payout is then

$$\$4,777 \times (.97)^{41} = \$1,370$$

If the inflation rate should grow to 8%, then, the pensioner's first month check is in current dollars $583. Social Security, on the other hand, is (supposedly) indexed for inflation. As John Bogle [2] has pointed out and as we shall demonstrate in Chapter 14, those who seek risk-free investment in fixed-rate bonds have failed to note that the actual value of a bond at maturity is an unknown because of inflation. The wise investor should put some of his or her annuity investment in stocks, since stocks, in a sense, self-adjust for inflation. Moreover, the current Social Security System has other benefits that the employee could elect to have incorporated into his payout plan, for example, survivorship benefits to a surviving spouse, and disability. Thus, our employer should really look further than bank returns. He needs to find something which is responsive to the cost of living. We will demonstrate later, in Chapter 14, how this might be achieved.

These additional add-ons would not cost sufficiently to lower the fair monthly payout below, say, $3000 per month. And we recall that these are dollars in a low inflation economy. Considering that there is some doubt that an individual entering the current job market will ever receive anything from Social Security at retirement, a certain amount of indignation on the part of the young is perhaps in order. Furthermore, the proposed private alternative to Social Security would allow the investments by the employee and his employer in the plan to be utilized as capital for investment in American industry, increasing employment as well as productivity.

The method of trial and error is perhaps the oldest of the general algorithmic approaches for problem solving. It is highly interactive; that is, the user makes a guess (in the above case X) as to the appropriate "input" (control variable, answer, etc.) that he feeds into a "black box." An output result is spewed out of the black box and compared to the desideratum—in the earlier problem, P and $P(N_3)$.

We note that the above example is one in which we know the workings of the black box well. That is, we have a model for reality that seems to be precise. And little wonder, for annuity is a manmade entity and one should be able to grasp the reality of its workings far better than, say, daily maximum temperatures in Houston forecast 3 years into the future.

In actuality, however, even the pension fund example is highly dependent on a series of questionable assumptions. For example, the interest figures used assume negligible inflation—a fair assumption for the mid-1980s but a terrible one for the late 1970s. Objections as to the assumption that the pension fund will be soundly managed are not too relevant, since most such funds are broadly invested, essentially a random selection from the market. Objections as to the uncertainty of employment of the employee in his current company are also irrelevant, since it is assumed that the vesting of such a fund is instantaneous, so that the employee loses no equity when he changes jobs. The plan suggested here is simply the kind of private IRA arrangement used so effectively by the Japanese as both a vehicle of retirement security and capital formation. Such a plan can reasonably be designed to track the performance of the overall economy. But the assumptions as to the specific yields of the plan will almost certainly be violated in practice. At a later time, we shall cover the subject of *scenario analysis*, in which the investigator frankly admits he does not fully understand the black box's workings very well and examines a number of reasonable sets of scenarios (hypotheses) and observes what happens in each. At this point, we need only mention that in reality we are always in a scenario analysis situation. We always see through a glass darkly.

Having admitted that even in this idealized situation our model is only an approximation to reality, we observe that a wide variety of algorithms exist for solving the problem posed. Usually, if we can come up with a realistic mathematical axiomatization of the problem, we have done the most significant part of the work. The trial-and-error approach has many advantages and is not to be despised. However, its relative slowness may be somewhat inefficient. In the present problem, we are attempting to pick X so that P is close to $P(N_3)$. It is not hard to design an automated algorithm that behaves very much like the human mind for achieving this goal. For example, in Figure 1.3, we consider a plot of $G = P - P(N_3)$ versus X. Suppose that we have computed G for two values X_{n-1} and X_n. We may then use as our next guess X_{n+1}, the intercept on the X axis of the line joining $[X_{n-1}, G(X_{n-1}\,])$ and $[(X_n, G(X_n)]$.

Using the program in Figure 1.4, we use as our first guess a monthly output of 0, which naturally produces a stake at 75 years of age of 0. As our second guess, we use $X_2 = \$450,298/(7\times12)$. With these two starting values of X_n and G_n, the program converges to a value that gives $P = 741,650$ to within one dollar in three iterations. The equitable payout rate obtained by the secant method is, of course, the same as that obtained by trial and error—namely, \$4,777.

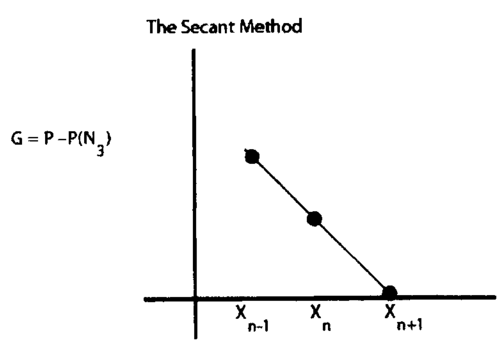


Figure 1.3. The secant method

$$P(N_2) = \text{principal at retirement}$$
$$\gamma = \text{annual rate of return on principal}$$
$$N_2 = \text{retirement age}$$
$$N_3 = \text{expected age at death given survival until age } N_2$$
$$P(N_3) = P(N_2)\left(1+\tfrac{\gamma}{12}\right)^{12(N_3-N_2)}$$
$$X_{n-1} = 0$$
$$G_{n-1} = -P(N_3)$$
$$X_n = \frac{P(N_3)}{12(N_3-N_2)}$$
$$\text{*Call Annuity } (N_2, N_3, X_n, 0, 6, \gamma)$$
$$G_n P - P(N_3)$$
$$\text{Slope} = \,= (G_n - G_{n-1})/(X_n - X_{n-1})$$
$$X_{n+1} = X_n - \frac{G_n}{\text{Slope}}$$
$$\text{Call Annuity } (N_2, N_3, X_{n+1}, 0, 6, \gamma)$$
$$G_{n+1} = P - P(N_3)$$
$$\text{Is } G_{n+1} < 1?$$
$$\text{If yes, print } X_{n+1} \text{ and stop}$$
$$\text{If no continue}$$
$$X_{n-1} = X_n$$
$$G_{n+1} = G_n$$
$$X_n = X_{n+1}$$
$$G_n = G_{n+1}$$
$$\text{Go to *}$$

Figure 1.4. Program secant method

1.2 Income Tax Bracket Creep and the Quiet Revolution of 1980

Many, with some justification, believe that our leaders in Washington have something approaching a proprietary interest in the wealth of America's citizens. That works up to a point, but beyond some hard-to-establish threshold taxes can cause revolutions. The American Revolution is a case in point. The high tariffs on exported cotton and imported textiles led, in large measure, to the American Civil War. The discussion here concerns a quiet revolution that removed President Carter from office by the election of 1980.

Facing inflation of more than 10% and interest rates that reached 20%, President Carter gave a speech on July 15, 1979, in which he blamed the problem on a "general malaise" of the American people. This was heady stuff, for most Americans did not feel they were part of such a general shiftlessness. The General Malaise Speech, as it came to be called, flew in the face of the famous maxim of the long dead economist and sociologist Vilfredo Pareto. One form of Pareto's Maxim is that the catastrophically many failures are not due to a general malaise but to a small number of assignable causes [[11, p.10].

It seemed to many that perhaps President Carter himself, rather than the general population, had allowed things to go very wrong with the nation's economy. The election of Ronald Reagan in November of 1980 had as its most important accomplishment the abolition of the Soviet Union. But the electorate was not dreaming of such a result. They were concerned about the deterioration in their standards of living. We will demonstrate below how the tax system contributed to the economic problem, which was the real reason for Carter's defeat.

Beginning in 1981, there were several changes in the U.S. income tax laws. The major reason advanced for these modifications in the tax regulations was something called "bracket creep." This is a phenomenon of the progressive income tax which causes an individual whose income increases at the same rate as inflation to fall further and further behind as time progresses. There was much resistance on the part of many politicians to this indexing of taxes to the inflation rate, since the existing tax laws guaranteed a 1.6% increase in federal revenues for every 1% increase in the inflation rate. Another problem that was addressed by the tax changes in the early 1980s was the fact that a professional couple living together in the unmarried state typically paid a few thousand dollars less in taxes than if they were married. Those who felt the need for indexing and some relief from the "marriage tax" had carried out several "if this goes on" type scenario analyses. We consider one such below. All the figures below use typical salary rates for 1980 and an inflation rate a bit below that experienced at that time. The tax brackets are those of the 1980 IRS tables.

Let us consider the case of John Ricenik who accepts a position with a company that translates into a taxable income of $20,000. Let us project John's earning profile in the case where both inflation and his salary increase at an annual rate of 7%. First of all, we see that John's income will grow annually according to the formula

$$\text{income} = 20,000(1.07)^{\text{year} - 1980}. \tag{1.9}$$

The tax required to be paid in any given year is easily determined from Table 1.1. Inflation, on an annual basis, can be taken care of by expressing all after tax amounts in 1980 dollars according to the formula

$$\text{value in 1980 dollars} = \frac{\text{nominal amount}}{1.07^{\text{ year } - 1980}}. \tag{1.10}$$

To determine John's after tax profile, we need to examine the 1980 tax tables for single taxpayers.

We will see that that John is not holding his own against inflation despite of the fact that his salary is increasing at the same rate as inflation. This is due to the fact that his marginal increases in salary are being taxed at rates higher than the average rate for the total tax on his earnings. The purpose of indexing is to see that the boundaries for the rate changes increase at the annual rate of inflation.

We can readily compute the 6-year horizon table for John Ricenik's after tax income in 1980 dollars (Table 1.1).

Table 1.1. Rates for Single Taxpayers.

If taxable income is not over \$2,300...		0	
Over	But not over		of amount over
\$2,300	\$3,400	14 %	\$2,300
\$3,400	\$4,400	\$154 + 16 %	\$3,400
\$4,400	\$6,500	\$314 + 18 %	\$4,400
\$6,500	\$8,500	\$692 + 19 %	\$6,500
\$8,500	\$10,800	\$1,072 + 21 %	\$8,500
\$10,800	\$12,900	\$1,555 + 25 %	\$10,800
\$12,900	\$15,000	\$2,059 + 26 %	\$12,900
\$15,000	\$18,200	\$2,605 + 30 %	\$15,000
\$18,200	\$23,500	\$3,565 + 34 %	\$18,200
\$23,500	\$28,800	\$5,367 + 39 %	\$23,500
\$28,800	\$34,100	\$7,434 + 44 %	\$28,800
\$34,100	\$41,500	\$9,766 + 49 %	\$34,100
\$41,500	\$55,300	\$13,392 + 55 %	\$41,500
\$55,300	\$81,800	\$20,982 + 63 %	\$55,300
\$81,800	\$108,300	\$37,677 + 68 %	\$81,800
\$108,300	...	\$55,697 + 70 %	\$108,300

Table 1.2. After-Tax Income.

Year	Nominal	Tax	Nominal After Tax Income	After-Tax Income (1980 dollars)
1980	\$20,000	\$4,177	\$15,823	\$15,823
1981	\$21,400	\$4,653	\$16,747	\$15,651
1982	\$22,898	\$5,162	\$17,736	\$15,491
1983	\$24,501	\$5,757	\$18,744	\$15,300
1984	\$26,216	\$6,426	\$19,790	\$15,098
1985	\$28,051	\$7,142	\$20,909	\$14,908

Let us now investigate the "marriage tax." Suppose that John Ricenik marries his classmate, Mary Weenie, who has the same earnings projections as does John—that is, 7% growth in both salary increments and inflation. You might suppose that computing the after tax income of the Ricenik family is trivial. All one has to do is to double the figures in Table 1.1. This is, in fact, the case if John and Mary live together without being legally married.

There is another table that applies to John and Mary if they are legally husband and wife. See Table 1.3.