
FORECASTING AND MANAGEMENT OF TECHNOLOGY

ALAN L. PORTER

Georgia Institute of Technology, Atlanta, Georgia

A. THOMAS ROPER and THOMAS W. MASON

Rose-Hulman Institute of Technology, Terre Haute, Indiana

FREDERICK A. ROSSINI

George Mason University, Fairfax, Virginia

JERRY BANKS

Georgia Institute of Technology, Atlanta, Georgia

with software and documentation by

BRADLEY J. WIEDERHOLT

Institute for Software Innovation, Alpharetta, Georgia



A WILEY-INTERSCIENCE PUBLICATION

JOHN WILEY & SONS, INC.

NEW YORK CHICHESTER BRISBANE TORONTO SINGAPORE

--- 8

MONITORING

OVERVIEW

Monitoring uses diverse sources of information to set the stage for forecasting, to feed information to the forecasting and planning processes, and to update and improve the projections upon which management decisions are made. This chapter presents the assumptions underlying monitoring, the various types of monitoring, and examples of the use of monitoring in industry. Special emphasis is placed on monitoring the context of technological changes as well as the changes themselves. Techniques to develop, interpret, and communicate monitoring results are presented, and suggestions are offered on its integration within an organization.

8.1 INTRODUCTION

Monitoring is to watch, observe, check, and keep up with developments, usually in a well-defined area of interest for a very specific purpose (Coates et al., 1986, p. 31).

This description captures the essence of this simple technique. Monitoring means scanning the appropriate environment for pertinent information. That information may pertain to a particular technology—*technological monitoring*—in which case one may want historical information on the technology's development, current information on the state of the art today, and/or information pointing directly to future prospects. Alternatively, monitoring may consider the context in

which technology develops—*contextual monitoring*, or *issues management*.¹ This approach derives from the assumption that technological change is foreshadowed by changes in other technologies and/or in the socioeconomic environment. Thus it should be possible to monitor signals in these environments, analyze them, and forecast technological development. For instance, concerns about carbon dioxide emissions promoting the greenhouse effect may presage the return to favor of nuclear power.

The last part of the monitoring definition emphasizes “specific purpose.” This emphasis is vital to avoid accumulating vast quantities of ill-focused information with no organizing principle. According to Coates et al. (1986, p. 31), possible objectives in monitoring include:

- Detecting scientific/technical or socioeconomic events important to your company
- Defining potential threats for the organization implied by those events
- Seeking opportunities for the organization implied by changes in the environment
- Alerting management to trends that are converging, diverging, speeding up, slowing down, or interacting.

These generic objectives—looking for events and trends that imply threat or opportunity—can be broken down further: identifying competing or supporting relationships, flagging critical needs, and ascertaining potential breakthroughs and their effect on the organization. Whatever the objective, careful definition of purpose and focus are essential to worthwhile monitoring.

Successful forecasting based on monitoring involves more than merely gathering data. The forecaster must sift the information for meaningful signals and envision their implications. This is best done through a systematic monitoring procedure that directs information search and interpretation. Finally, results of monitoring must be synthesized and communicated effectively to generate appropriate action. Indeed, substantive work may be dwarfed by the effort to orchestrate, package, and sell the information to achieve the necessary action (Coates et al., 1986).

Monitoring is one of the most useful techniques in forecasting. Applications depend on what the user needs. For instance, an Air Force research organization built a technological forecast for artificial intelligence largely on an expert/literature monitoring effort to systematize research funding plans (Reitman et al., 1985). Search Technology Inc. has used a monitoring system to help design a new tech-

¹*Issues Management* is defined by Coates et al. (1986) as “the organized activity of identifying emerging trends, concerns, or issues likely to affect an organization in the next few years and developing a wider and more positive range of organizational responses toward that future” (p. ix). Issues management can help companies devise positive, anticipatory (rather than merely reactive) responses to new technologies, possible governmental constraints, or potential confrontations. Issues management emphasizes contextual monitoring (law-making, social controversies, etc.), but it can include technological monitoring, too.

nology called the Designer's Associate (DA)—a computer-based system to assist cockpit designers—which was prepared for the Air Force. This required an assessment of what technological alternatives might be available to accomplish given DA functions in 1995. Monitoring contributed in the following ways (Porter, 1988):

- It ascertained technological capabilities as of 1988 to help “configure” the DA.
- It alerted management to possible new technological capabilities for 1995.
- It assessed the risk in designing a system based on those potential capabilities for 1995.

Monitoring also can serve those on the lookout for new products or new processes. Vanston (1985) has sampled such American corporate applications. For example:

- ALCOA watches for new market areas and examines threats to present markets.
- Kraft analyzes changes in home eating patterns to determine effects on new food products.
- Owens-Corning Fiberglas seeks new products and processes.
- Johnson Controls predicts competitors' entrance into new technical areas.

Exhibit 8.1 provides a case illustration of the implications of monitoring.

■ Exhibit 8.1 Whirlpool Beats the Competition (Based on Davis, 1973).

Whirlpool Corporation tracked developments in the chemical and textile industries (not their own business domains) to identify permanent press fabrics prior to their commercialization. They acted quickly on this information to generate the first washer and dryer permanent press cycles, beating their competition to market by about a year. This resulted in a substantial increase in market share. The major events involving permanent press products and their impact on Whirlpool are as follows:

Winter 1963-1964	Rumors in textile industry of new “delayed cure” process for resin application.
April 1964	Confirmation of rumor
May 1964	First glimpse of process
May-June 1964	Education of Whirlpool personnel about permanent press garments; forecast of doubling of dryer sales in a year
August 1964	“Development conference” arranged for vice-president, laundry, at large fiber manufacturer; permanent press garments shown at Whirlpool for first time
September 1964	Textile industry introduces permanent press

January 1964	Permanent press cycles on Whirlpool washers and dryers; first by appliance industry
March 1965	Research project for a "new concept ironer" dropped; monies diverted to other uses

In a quite different context, through its National Center for Science and Technology for Development, the Chinese State Science and Technology Commission maintains a significant technology forecasting effort. Monitoring new technologies available from developed nations, such as the United States and Japan, serves its immediate industrial needs. Monitoring for emerging technologies helps to guide China's own technological development efforts by avoiding dead ends and seeking realistic niches.

Policymakers also can be served by the monitoring of emerging trends and possible breakthroughs in technology as well as by contextual monitoring. The Congressional Clearinghouse for the Future provides this service for the U.S. Congress.

Monitoring also provides a base for technology forecasting using other techniques. It provides data and background for choosing and using forecasting tools. Section 8.2 considers the principles for sound monitoring; Section 8.3 offers a systematic monitoring process; Section 8.4 addresses implementation issues.

8.2 PRINCIPLES

This section describes the basic assumptions that support monitoring and the monitoring methods that are commonly applied.

8.2.1 Assumptions

Monitoring is an opportunistic technique. If patterned technological development is expected to continue, monitoring tracks this historical development as a basis for projection. If patterns are not as well-defined, monitoring relies on current information on the state-of-the-art and expert assessments of future advances. Such advances may be incremental or breakthroughs—monitoring can tackle either. In some cases, monitoring contributes more to current awareness than to true forecasts, but it is an extremely robust technique as long as the information gathered is cautiously interpreted.

Both technological and contextual monitoring must be tailored to fit the developmental stage of the technology, as well as to fit user needs. Table 8.1 presents a version of the "suspect" linear model of change (recall the innovation process models discussed in Chapter 2)—not to argue for linearity, but to suggest what information may be most salient to monitoring at various stages of a technology's development.

Some have tried to develop a monitoring framework based on technological characteristics. In particular, there have been attempts to distinguish functional and structural measures. For instance, Knight (1985) differentiates main structural units of the computer (memory, computation, control, input-output) and key functional

TABLE 8.1 Monitoring Information as a Function of Technological Development

Stage of Development	Key Issues	Information
Scientific discovery	Scientific uncertainty	Enduring interest within the concerned segment of the scientific community Sporadic interest in public media
Applied research & development	Technological uncertainty	Information may not appear in the open literature, depending on the developers Trade magazines may be a prime source of what information is useful "Inside" experts may be critical to the process of verifying information Concern about possible impacts of the technology may arise in scientific or public media
Initial applications	Economic uncertainty	Widely available from diverse sources Market assessment of key concern Impact assessment may become a serious issue Policy development and assessment issues
Widespread adoption	Social uncertainty	Technology evaluation Policy assessment Renewed emphasis on technological development, seeking to project successor technologies Market assessment to determine potential, niches, and competitors

attributes (computing power, cost, and reliability). He then cumulates qualitative and quantitative measures on each of these over extended time periods. His own assessment is that the approach does a good job of depicting the evolution of the computing industry, but is not detailed enough to isolate specific technological changes. This approach is appealing, but there are concerns about its application. For instance, over the long term, structural frameworks may change (for example, with the advent of personal computers or neural computing) and functional parameters may evolve (for instance, increases in computing power may result in qualitative changes in criteria, or new applications such as artificial intelligence may rely on nonmathematical processing power). It makes the most sense to consider function and structure, or other organizing principles, in light of the scope, duration, and purposes of the monitoring effort.

8.2.2 Types of Monitoring

Table 8.2 poses a series of overarching choices that must be made about any monitoring process. The basic focus, whether one is monitoring technology or contextual factors, has been discussed in Section 8.1.

The first choice is the time frame. In some cases, a technology or a particular context is being studied in an exploratory or preimplementation mode (for instance,

TABLE 8.2 The Overarching Choices of Monitoring*Time Frame*

Preimplementation vs. Imminent Decision vs. Postdecision

Monitoring Process

One-time study vs. Ongoing monitoring

Focus

Technological monitoring vs. Contextual monitoring

Breadth

Macro vs. Micro

Purpose

Choosing vs. Forecasting

Developmental Stage

Invention vs. Innovation

Established vs. Emerging technology

to help decide if the company wants to pursue that technology, enter a certain market, etc.). In this instance, monitoring is likely to have limited resources and uncertain management commitment; it is also likely to require strong selling to convey an effective punch line. In contrast, imminent decision monitoring suggests an attentive and anxious audience whose specific concerns will be directly addressed in the monitoring report. The postdecision time frame implies evaluative monitoring (that is, determining what happened by focusing on actual events directly involving the organization). This may be required by outside authorities (as a way to ensure that negative effects are mitigated) or may be undertaken to help the organization manage a specific technology. In any event, this postdecision monitoring clearly differs from preimplementation monitoring.

The second choice is whether to undertake a one-time study or to set up an ongoing monitoring system. The former must ensure that it does not narrow the monitoring focus prematurely and miss significant influences. The latter has the advantage of compiling contextual information over a long period that can be used to form more precise inquiries (see Exhibit 8.2). The monitoring system must have a clear focus and precise objectives to avoid becoming irrelevant to the organization.

Technological monitoring approaches will differ depending on the level of focus. Macro technology monitoring could encompass a whole spectrum (such as the information technologies, as per Exhibit 8.2). Alternately, micromonitoring might focus on one technology (the computer) or a particular type of that technology (personal computers). Another choice would be to focus on systems (office automation), individual devices (personal computers), or components (microprocessor chips). Whatever level is appropriate for the study, it is incumbent upon the

■ **Exhibit 8.2 A “Two-Stomach” Information Technology Monitoring System (Based on Neste, 1988).**

Search Technology implemented an ongoing monitoring system to keep abreast of changing information technologies in early 1989 (Neste, 1988). Essential criteria for this system included:

- Minimal effort required for dispersed participants to contribute information to the system
- Convenient access to information for users with quite different needs
- Low cost
- Ongoing system that facilitates generation of up-to-date reports as needed

One technical staff member takes responsibility to maintain the system and evaluate its performance (observe usage and suggest improvements). One clerical employee helps input information by copying documents, filing, and performing other related tasks. Six or so technical staff members each take lead responsibility for several technologies. They process the information as needed and generate suitable syntheses and forecasts. Another 20 or so technical people contribute to the system as they find materials pertaining to the targeted 20 (changing over time) technologies. Users include project managers designing a particular technological system; management personnel proposing to provide monitoring information per se to clients or drawing on the system for general background information; and technical personnel writing articles or reports who want information on one or more of these technologies.

Initiation required that management approve the system, that the target technologies be identified, that the six staff members agree to take responsibility for particular technologies, and that all staff be alerted to the new system. Staff prepared pilot forecasts on certain key technologies and generated initial technological maps to help identify what other technologies should be monitored. Management and project users were interviewed to determine their needs. Staff were briefed on how to contribute information at a meeting and through a memo, and a system manual was prepared. A special bulletin board alerted staff to key items.

The physical setup was based on a dual filing system—the two stomachs. One folder for each technology contained all raw (undigested) materials, including copied articles, book references (indicating where the book could be found), clippings, and other such material. A second folder for each technology contained monitoring reports prepared by Search Technology staff on an as-needed basis (every six months or so) plus important supporting sources (well digested). When a report was prepared, that input folder was emptied, and only vital materials were maintained in the report folder. System materials were kept in a designated set of easily accessible filing cabinets. ■

analyst to consider how changes at other levels might impact the focal technology. For example, the person monitoring office automation cannot ignore improvements in computer storage that might substitute for networking arrangements.

Sometimes monitoring serves more than just forecasting purposes; it may also help in choosing a future course. As mentioned, the Chinese monitor technology developments in other advanced nations. The state of Japanese or Western technology provides an excellent menu from which to choose their own development programs. Obviously, monitoring existing technologies greatly reduces the uncertainty, but it must also consider technology transfer issues, including cultural and political factors.

Invention and innovation are concepts that denote the issues attendant to different stages of development (see Table 8.1). However, these concepts also represent a choice of monitoring emphasis. For certain uses, interest may be invention-oriented—that is, focused on how to reduce technological uncertainty (for instance, it may be necessary to determine technological capabilities by a given date). For innovation-oriented uses, the key is really economics—Will a technology succeed in some market? Invention focuses on critical technical “milestones,” their prerequisites, and the likelihood of success by a given date. Innovation, on the other hand, focuses on whether a technology will meet needs of some users at an attractive price and/or in a way that is better than competing technologies. It is often easier to predict the emergence of technical capabilities (particularly when improvements come as a stream of incremental gains rather than discrete breakthroughs) than to predict the date of the actual innovation (that is, success in some market). These two types of technological monitoring emphasize different information sources (e.g., engineering versus market knowledge).

The monitoring of an established technology can draw upon historical pattern information concerning generally agreed-upon parameters. Martino (1987) offers a process to monitor established technologies that uses precursors (see Exhibit 8.3). In contrast, the monitoring of an emerging technology must draw on more diffuse information (also likely to imply more diffuse sources) with a less well-defined target. The emerging technology is likely to be more sensitive to contextual influences, and the time frame for it is likely to be longer, with attendant uncertainty in any forecasts.

Motorola devised a systematic monitoring/forecasting approach that used a technology roadmap process to sharply differentiate between emerging and established technologies (see Chapter 3). A very different approach is to track emerging technologies in general. For example, the National Institute for Standards and Technology (NIST, formerly the National Bureau of Standards) identified 12 emerging technologies for the 1990s (Technology Administration, 1990):

Emerging Materials

1. Advanced materials
2. Superconductors

■ Exhibit 8.3A Precursors of Technological Change

Martino (1987) illustrates this approach by compiling entries from the Engineering Index (and its on-line Compendex data base) for selected technologies from 1970 to 1987. The resulting qualitative histories can be used to anticipate coming changes. Martino recommends watching for

- Incomplete inventions that need other elements before they can be deployed economically
- Development of performance-improving supporting technologies needed by a basic technology
- Development of cost-reducing supporting technologies needed by a basic technology
- Development of complementary technologies that mesh with the basic technology to make it useful (such as sensors for automated controllers)
- Use of a technological advance in a prestige application before transition to general use (such as transition of technology from aviation to automotive uses—early uses in racing cars)
- An incentive for use, such as reduced cost or elimination of externalities (for instance, fuel economy or reduction of automotive emissions)
- Different origins for leading indicators (for instance, motor car manufacturer advances are nearer-term indicators of coming automotive change than are advances coming from parts suppliers)

Martino provides the accompanying chronology of plastic auto body shells to illustrate one such precursor trail. (His report also provides another trail of improvements in steel auto body shells to “counterattack” the threat posed by plastics.)

Sometimes, a leading indicator relationship can be found. Martino illustrates this relationship using the development of new aluminum alloys, followed by their first application in an aircraft, as an example. In that case, a mathematical relationship (by using regression, for example) may be devised to anticipate when to expect the application of new alloys (average lag is 4.2 years). Martino suggests treating such leading indicator relationships by devising probability distributions. In the alloy–aircraft case, he uses the method of maximum entropy (Tribus, 1969) with only the mean specified (to yield a geometric distribution) or with the standard deviation of the lag specified (to yield a Gaussian distribution truncated on the lower end—lag of zero). In this case, the Gaussian distribution is most informative, predicting virtually no chance of application of a new alloy in less than three years and almost certain availability within six years. Such information would be quite helpful to an aircraft manufacturer planning materials for a new plane. ■

■ Exhibit 8.3B Events in the History of Plastic Auto Body Shells

1978

Ford built demonstration Graphite Fiber Reinforced Plastic car; test results reported in 1983.

ICI America introduces resins and urethane thickeners for producing of structural grade sheet molding compound (SMC). Actual field use still needed to demonstrate usefulness for auto applications.

1980

Ford Motor Co. reports design, development, manufacture of lightweight, one-piece SMC hood for Econoline Van.

Laminates of steel sheet with plastic core have same formability as steel, lower cost, lighter weight.

1981

About 100 kg polymers now used in medium-sized car.

Reports of several experiments with plastic for body panels, interior components.

1982

FIAT builds demonstration car with steel skeleton, plastic panels for body shell. Skeleton has only 800 welds as compared with 3,000 typical for all-steel car. More resistant to corrosion, better formability for low drag shape, saved 20% weight over all-steel car.

1983

1984 model Fiero bolt-on body allows $\pm .005$ tolerances in finish and assembly of bodies; precision is equal to or better than steel.

1984

Production of Pontiac Fiero, steel "space frame" with all-plastic body skin. Flexible plastic "Bexloy" used in rear. Proves plastic can match surface finish of steel. Special painting arrangements; separate lines for space frame, for plastic panels, for flexible front and rear fascias. Painting panels before mounting reduces paint repair work.

Plastics use in autos, triples that of a decade earlier, now 10% of car total weight.

1985

Finite-element methods used to design steel chassis which carries all torsional loads, no stress on plastic body.

Successful in-plant coloring of ABS body plastic, reducing inventory costs by eliminating need to store colored plastic.

1986

2,000-ton plastic molding machine produced for auto plants.

Ford Sierra plastic grille panel is body-color painted and in same plane as painted metal panels, demonstrating that plastic can be used in conjunction with steel.

Source: Martino, 1987. Reprinted by permission of Elsevier Science Publishing Co., Inc. ■

Emerging Electronics and Information Systems

3. Advanced semiconductor devices
4. Digital imaging technology
5. High-density data storage
6. High-performance computing
7. Optoelectronics

Emerging Manufacturing Systems

8. Artificial intelligence
9. Flexible computer-integrated manufacturing
10. Sensor technology

Emerging Life-Science Applications

11. Biotechnology
12. Medical devices and diagnostics

In contextual monitoring, macro and micro distinctions are also made. At the macro level, various entities (typically governmental bodies) compile indicators—ideally, time series data—on various socioeconomic factors of general interest (such as educational achievement test scores, unemployment rates). Private-sector firms may also compile such data, monitor literature, and/or prevail upon experts to identify contextual trends of note. For instance, the Roper Organization produces “The Public Pulse”—a report on American preferences. One issue (Roper’s, 1987) offered 31 major trends shaping the future of American business. Two examples from that report are:

The Culture of Convenience—the fast-rising number of two-income households lies behind the takeoff of the “convenience industry,” projected to continue. . . . this also explains the tremendous potential of prepared take-out foods and of appliances like microwave ovens. . . .

Permanent Damage to Nuclear Energy Industry—growing environmental concerns, combined with the Chernobyl disaster, have retarded further developments in nuclear energy for the foreseeable future. . . . Public worries and opposition to new nuclear plants are high and rising. . . . The nuclear power industry probably cannot recover in the short term. . . .

It appears that the first sample projection is holding true in the 1990s, but the second one could be overturned by other environmental concerns (such as acid rain and the greenhouse effect).

Micro-level contextual monitoring focuses on those elements of the environment perceived to impinge more directly on an organization’s interests (or a technology of interest). For instance, computer firms monitor announcements, patents, and rumors concerning their competitors’ activities.

Contextual monitoring can address a vast array of possible changes. At times it may be useful to check coverage of all domains likely to influence the organization as regards its monitoring objectives. The following list will generally provide a comprehensive catalogue of 10 areas for investigation:

- Technological (scientific)—the 12 NIST technologies to watch
- Health—AIDS rates, life span, nutritional levels
- Institutional (legal, organizational)—governmental regulation and agency power, expansion of multinational corporations, industrial concentration in a sector, public participation
- Social (behavioral)—advent of the Information Age, changing educational levels, urbanization, birth rates, expanding elderly population, minority interests
- Political—governmental stability, national and state party power, organizational politics
- Economic—recession versus prosperity, union strength, job displacements, service sector growth
- Cultural (values)—changing work ethic, mass media, cultural homogenization, rising aspirations, consumerism, women's movement, acceptance of fractured families
- International (development)—world economy, changing patterns of competition, financial instabilities, OPEC
- Ecological—greenhouse effect, aquifer pollution, resource shortages, nuclear waste disposal
- Security—changing patterns of hostility, evolving alliances, military funding, SDI

8.3 MONITORING STEPS

Given the diversity of possible forms of monitoring, any suggested set of steps must be carefully adapted to specific needs. However, the following steps are generally useful:

1. Determine the monitoring objectives and focus
2. Describe the technology and map the pertinent context
3. Adapt an appropriate monitoring strategy
4. Interpret and communicate results

8.3.1 Monitoring Objectives and Focus

Sections 8.1 and 8.2 have discussed a range of monitoring information types, possible uses, and prospective users. The first task of the would-be monitor is to de-

termine which elements of these ranges fit the situation. Specific objectives should be spelled out (monitoring has a tendency to generate overwhelming amounts of information unless objectives are kept strictly in view), and a focus must be established. To undertake technological or contextual monitoring, you need to resolve the choices posed in Table 8.2.

Objectives and focus should be discussed by those engaged in the monitoring effort and with potential users. Agreement should be reached on issues such as scope, temporal extent, personnel and financial resources, timetables, whether the effort will be ongoing, and interest in trends or breakthroughs.

8.3.2 Technology Description and Contextual Mapping

As discussed in Chapter 4, you must describe a technology before you forecast it. Monitoring is, however, likely to generate an iterative situation. It begins with a presumptive definition/description of the technology. Monitoring enriches understanding of the technology, possibly leading to its redefinition in terms of level (for instance, addressing microprocessors instead of microcomputers), the pertinent technological system, and the critical parameters or milestones (critical qualitative or quantitative improvements) to monitor.

Mapping, whether graphic or text, can help identify relationships among technologies (see Figure 8.1 for selected information technologies perceived pertinent to computer systems). Likewise, it can help identify vital contextual socioeconomic influences (see the discussion of technology delivery systems in Section 2.3). Without such mapping, critical signals may be missed because monitoring may be wrong or too narrow.

Technologies and contextual influences should also be linked to organizational interests. Brown (1980) suggests composing matrices that indicate the likelihood of significant change on one axis and its likely impact on the company on the other (see Figure 8.2). Such a matrix allows quick identification of which changes would constitute milestones for the organization.

8.3.3 Monitoring Strategy

A number of strategic monitoring issues have been mentioned (such as making monitoring choices, meeting different objectives, making do with different resource levels, and performing under accelerated or extended time schedules). In this section, monitoring approaches are differentiated by how familiar you are with the subject. (This discussion assumes technological monitoring, but the implications generally fit contextual monitoring, too.) As a guideline, it is useful to contrast three levels of familiarity:

Level 1 Cold

Level 2 Warm

Level 3 Hot



Figure 8.1 Map of technologies relating to computer systems. (Source: Prepared by S. Cunningham, 1990.)

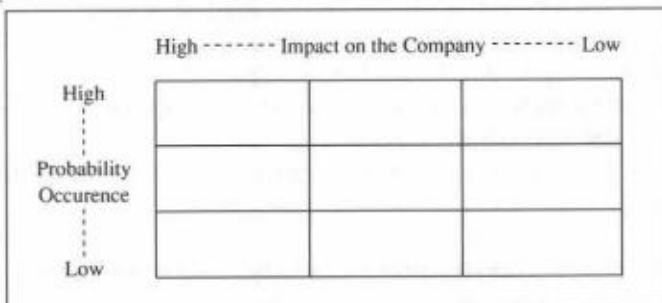


Figure 8.2 Impact and probability of occurrence matrix. (Source: Brown, 1980, p. 31.)

At Level 1, you start “cold”—that is, you are unfamiliar with the subject. At this level, the immediate questions include:

- What is the technology? How is it defined and described? What is the state-of-the-art?
- How do other technologies relate to it? What other contextual factors effect it?
- Who are the key players (individuals, organizations, suppliers, regulators, users)?
- What are plausible future development pathways?

To accomplish Level 1 monitoring:

- Use a “shotgun” approach to gather information—that is, grab anything that is convenient and may be pertinent
- Emphasize recent literature and review state-of-the-art articles or books
- Locate one or two accessible professionals with sufficient expertise in the technology to point out information sources and to help ensure that the monitoring does not drift wide of the mark

Level 2 suggests that you have either completed a Level 1 introductory effort or that you have some familiarity with the subject. Objectives become more focused; pertinent questions include:

- What are the driving forces propelling this technology?
- Can important interdependencies with other technologies or with socioeconomic factors be mapped?
- What are the key uncertainties along the development path?

To accomplish Level 2, the sources of information shift:

- Literature searches become more focused; on-line searches are likely to be more fruitful; and historical searches now may make sense as a way to identify leading indicators of progress and significant influences.
- Locating other forecasts for the focal technology can help answer the pivotal questions and help ground forecasts.
- Networking can be used to identify experts with various perspectives on the technology.

It now makes sense to begin to synthesize the information obtained by formulating an image of what is happening to the technology.

At Level 3, you are very familiar with the subject. Level 3 objectives are even more focused; pertinent questions include:

- Can you specify the key factors to be watched?
- What is the most likely development pattern for the immediate future? For the longer term?
- Can you offer specific projections?
- What recommendations can you make to help your organization manage development?

Actions undertaken to accomplish this level of monitoring include:

- Extending the information search to be as comprehensive as feasible with regard to the key factors
- Developing a conceptual model of what drives and what impedes development of this technology
- Seeking direct confirmation of this model and a review of your projections from experts
- Generating a credible forecast by integrating the monitoring results with other forecasting techniques
- Possibly establishing a structure for an ongoing monitoring system

These three levels suggest different requirements. Subject expertise is not essential at Level 1, but it is necessary at Level 3. Comprehensive information access (such as access to on-line systems) is not critical at Level 1, but it is at Level 3 to ensure thoroughness. Commitments of time and effort also increase at the higher levels. It is sensible to commit only to a Level 1 effort in investigating a technology to ascertain its relevance. The results of such an introductory examination may be sufficient to conclude that the subject is not critical to the organization. On the other hand, if the subject does seem sufficiently important, you may need to revamp the monitoring program to involve experts (from inside or outside the organization), boost resource commitments, and set forth a full-blown monitoring system.

8.3.4 Interpretation and Communication

Data do not speak for themselves; therefore, it is very important that key points be flagged in such a way as to get management's attention. Possible mechanisms to foster communication with potential users include maps, milestones, and multiple channels.

Graphical presentation of technology maps (imagine Figure 8.1 with all of the linkages shown) quickly conveys complex relationships, especially to visually ori-

ented individuals (engineers are great candidates). Technology delivery system sketches fall in this same category, particularly if a few key contextual influences can be highlighted. Any model of what drives the development of a given technology should be interpreted to point out the few, critical factors that should be watched.

Monitoring may generate extensive historical profiles of events and/or trends. These should be offered with further interpretation. The notion of milestones for which to watch can be quite effective. Table 8.3 illustrates innovation milestones in the history of cement manufacturing in the United States. Milestones may reflect events or trends advancing dramatically so as to improve performance significantly.

Multiple communication channels improve the chances of successfully transmitting information. Both parallel channels and serial repetition (same or similar messages repeated at different times) may make sense. For example, the monitoring system discussed in Exhibit 8.2 has multiple channels. Some channels work well for some types of information, but not for others. Paper folders fulfill the configuring objectives quite well, but they do not accomplish the alerting or forecasting functions. Wall maps of relationships (see Figure 8.1) seem excellent for configuring and helpful for alerting. A bulletin board provides a nice alerting mechanism, but little more. Reports appear best for presenting forecasts, but they are too infrequent to alert organizational staff in a timely manner. Finally, meetings can help round out rough spots by providing rich, two-way communication opportunities, but they are costly in personnel time.

In addition to the inherent strengths and weaknesses of the communication channels, consider the individual quirks of the providers and users. Some people favor graphical presentations; some want it written; others thrive on verbal exchange. When feasible, it may be best to use multimedia schemes to reach as many as possible.

8.4 IMPLEMENTATION

Implementation involves three issues: Who should perform the monitoring? How can they put together a monitoring program? How can they integrate monitoring results with other forecasting techniques?

8.4.1 Who Should Conduct Monitoring?

Informal monitoring can be done by anyone who alertly keeps track of information obtained from routine reading of journals, casual browsing, and other such activities. Some sources provide a sort of ready-made monitoring. For instance, *IEEE Spectrum* publishes a "newslog" each month of items pertaining to electronics developments. *Predicasts Forecasts* compiles forecasts from a wide range of government reports and open literature; these are compiled quarterly (and maintained on-line as well) using Standard Industrial Classification (SIC) codes. An example

TABLE 8.3 Significant Technological Discontinuities in Cement Manufacturing

Industry	Year	Event	Importance	Type of Discontinuity	Locus of Innovation			Probability
					New Firms	Existing Firms	1 of 10	
Cement	1872	First production of Portland cement in the United States	Discovery of proper raw materials and importation of knowledge opens new industry	Niche opening	10 of 10	1 of 10		
	1896	Patent for process burning powdered coal as fuel	Permits economical use of efficient rotary kilns	Competence-destroying	4 of 5	1 of 5		.333
	1909	Edison patents long kiln (150 ft)	Higher output with less cost	Competence-enhancing	1 of 6	5 of 6		.001 ^a
	1966	Dundee Cement installs huge kiln far larger than any previous	Use of process control permits operation of very efficient kilns	Competence-enhancing	1 of 6	7 of 8		.000 ^a

Source: Tushman and Anderson, 1986. Reprinted from "Technological Discontinuities and Organizational Environments" by Michael Tushman and Philip Anderson published in *Administrative Science Quarterly* (vol. 31, no. 3) by permission of *Administrative Science Quarterly*.

^a $p < .01$

Note: Fisher's exact test compares the pool of firms that are among the first to enter the niche with the pool of firms that introduce or are among the first to adopt a major technological innovation. The null hypothesis is that the proportion of new firms is the same in each sample; probability is the probability of obtaining the observed proportions if the null hypothesis is correct.

issue contains some 95 items pertaining to industrial robots, reporting projections for robot sales in units or dollars, related employment, prices, or numbers in use (specifying years that the projections address and providing source information on the projections). Any person with a professional level of expertise related to the topic should be able to select pertinent information from these sources.

Formal monitoring varies so greatly in scale that formulating rules would be foolhardy. Instead, consider the range of possible performers. Outside information services may be able to provide the information desired. Options include targeted newsletters, clipping services, trade associations, or contracted monitoring studies done by research firms. Possible disadvantages include less knowledge of the organization's needs, weaker communication bonds to organizational management, and less control over the accessibility of information to others. Advantages include possible strong subject expertise and established access channels.

Within the organization, a single staff person could be assigned to monitor a subject; alternately, that person might be supported by some outside information services. Although the single-person approach is efficient, it may be less effective than a more broadly based approach.

A study by Allen (1977) on where and how scientists and engineers obtain information has pointed to the importance of "gatekeepers." These professionals (perhaps, but not necessarily, managers) serve as critical bridges to the outside world. Most of the flow of ideas in technical development takes place within project groups. According to Allen, relatively few individuals in R&D read much technical literature or, especially in the case of engineers, actively participate in technical meetings. The gatekeeper funnels in a wide sweep of information for use by the project team in their relatively narrow application focus. Gatekeepers tend to emerge, rather than to be created by their managers. They should be well networked within their organizational unit (whether this is a single project team or a large company) and be avid information hounds. Gatekeepers make ideal monitors.

The next logical step is to assign an ad hoc team, from some central staff unit, to conduct monitoring. Alternatively, a team might be set up on a permanent basis to run a monitoring system (see Exhibit 8.2), or monitoring could involve more widely dispersed personnel (see Exhibit 8.4). The dispersed effort can directly involve those with subject expertise and decision-making responsibilities, but it requires incentives to participate and checkpoints to ensure performance and communicate results. Coates et al. (1986) elaborate on possible arrangements.

■ Exhibit 8.4 Dispersed Monitoring in a Food Processing Company

Technology forecasting is a new venture for this company. In the past, innovation has not been a high priority for this marketing-oriented firm. Now, however, management is making new technology a major corporate objective to save on costs through process improvement and to expand markets through new products and packaging.

This company is pursuing a diffused monitoring strategy. The director of engineering has initiated a comprehensive search strategy that categorizes possible application areas for new technology. Within each category, the director has specified technologies to monitor. For instance:

- Pumps, agitators, and mixers
- Ultrafiltration
- Microprocessors
- Wastewater treatment
- Wire-guided vehicles

One engineer is assigned responsibility for each area. That person is charged with monitoring the technology, identifying potential innovations for the company to pursue, and developing specific proposals for action. Such proposals can come forth at any time. (In addition, regular semiannual progress reports are due.) Superior proposals become development projects with specific objectives and budgets.

The great advantage of this approach is that the engineers responsible for development and implementation do the forecasting. No separate "sales pitch" is needed to convince engineers to act on the forecast. ■

8.4.2 How Should Monitoring Be Conducted?

Traditionally, monitoring relied heavily on printed materials. In decades past, technology monitoring relied on composing journals with entries interpreted with respect to prescribed parameters and concerns (see Bright, 1972). That appears too restrictive for ongoing systems that try to track rapidly advancing families of technologies.

Advances in electronic information accessibility represented by the some 5,000 on-line databases (as of 1990), the gateway services (such as BRS and DIALOG), and the wide availability of networked personal computers offer attractive options. Instead of maintaining a monitoring file with journal reports keyed on certain identified parameters, on-line searches on old or new parameters can be generated. As of early 1990, however, searches are still not satisfactory for many purposes. Hard copy systems supplemented by on-line searches to capture additional new items on specific topics appear to be most cost-effective (see Exhibit 8.2). However, as data bases become increasingly "full text" and include more graphic information, electronically based monitoring should emerge as the method of choice. Watch for intelligent information retrieval mechanisms that go beyond Boolean and probabilistic searches (using systems such as SIRE, for instance) to help the user select valuable information (Wiederholt et al., 1989). Indeed, artificial intelligence-supported electronic monitoring could, one day, help the user pose the monitoring questions and come back with suitable answers, not just unprocessed sources.

Today, it may be worthwhile to develop monitoring files on the computer for ease of cross-referencing and/or browsing (that is, using hypermedia systems). Input and maintenance of such a system is more time-consuming than use of hard copy files, but the payoff can be worthwhile gains in effectiveness in some large monitoring efforts with many people involved and many cross-linked interests (such as those suggested by Figure 8.1).

8.4.3 Integrating Monitoring Results with Other Forecasting Techniques

Monitoring provides the basis for most technology or socioeconomic forecasting. Sometimes it provides all that is needed or affordable. Most of the time, however, monitoring provides one component of a forecast. If monitoring is restricted to mean gathering of information from published sources, then it must usually be enriched by expert opinion. For instance, one forecast of intelligent information retrieval (Wiederholt et al., 1989) began with a casual discussion with in-house experts, continued with consolidation of published literature, and then formally requested review and elaboration by outside experts. The process continued with revision and further monitoring over several months before the final report. In such a process, monitoring is tightly integrated with expert opinion methods. Monitoring may initiate forecasting studies that utilize other forecasting techniques as well. For instance, monitoring of developments in personal computers has uncovered salient time series data, suggesting trend extrapolation.

Monitoring also provides a vehicle to keep forecasts alive. Rather than allow a study to be published and become out of date, assign someone to track developments and update the study periodically (this can be done at relatively low cost). Should such monitoring uncover milestone changes, then reexamine the forecast. In the previous paragraph, monitoring was cast as the precursor to a forecast; here it is the follow-on activity.

Because changes in different technologies are interdependent and technological change depends on contextual social influences, it is attractive to consider forecasting such sociotechnical systems rather than focusing on a single technology. Again, monitoring constitutes the main activity of such forecasting efforts. If monitoring turns up a critical change then initiate a more focused study. Such integrated, monitoring-based forecasting systems serve the organization best.

8.5 RECOMMENDED SOURCES

Here are the general categories of published information sources with examples of each:

- Journals: *The Futurist*, *Technological Forecasting and Social Change*, *Futures*, *Science*
- Trade journals: *PC Week*, *Iron Age*
- Newsmagazines: *Time*, *Business Week*

- Newspapers: *New York Times*, *Washington Post*, *The Wall Street Journal* (especially indexes of these, on-line or hard copy)
- Government activity reports:
 - Legislative reports—Congressional Research Service studies, Office of Technology Assessment studies, CIS Index (Congressional Committee hearings and reports), Congressional Record, Congressional Quarterly
 - Executive reports—weekly compilation of presidential documents, monthly catalog of government reports, agency indexes (such as NASA, DoE)
- Indices: Engineering Index (known as Compendex on-line), Science Citation Index (also Social Science Citation Index), Trade and Industry Index, Computer Index (all these are available on-line), Library of Congress Catalog of Books by Subject
- On-Line Database Gateways: services such as BRS and DIALOG provide access to multiple electronic data bases; college library collections increasingly are cataloged on-line and may provide reduced price (or free) access to outside data bases
- Information Services: International Data Corporation (Framingham, Massachusetts), Bacon's Clipping Bureau (Chicago)

EXERCISES

8.1 Exhibit 8.1 chronicles Whirlpool's successful monitoring of advances in chemicals and textiles to affect home appliances. Consider the following industries:

- (A) A major American steel company
 - (B) A Japanese home entertainment company
 - (C) A multinational oil company subsidiary in the pesticide business
- What domains would you monitor for these industries today and why?

8.2 Monitoring can be configured very differently as a function of its purposes and the situation in which it is occurring. Imagine yourself formulating a monitoring effort in three of the following situations:

- Information technology developments—for the government of a small developing country in Latin America
 - Information technology developments—for a major American bank
 - New materials developments—for a "Big Three" American automaker
 - Genetic engineering developments—for a Congressional Committee
 - New drug possibilities—for a multinational pharmaceutical company
 - Socioeconomic contextual changes—for a multinational pharmaceutical company
- (A) Identify the key type of uncertainty for each of these three and your major information acquisition concerns (refer to Table 8.1).

- (B) Following Table 8.2, make each of the pertinent choices and briefly explain your reasoning for these three situations.
- 8.3** For one of the situations in Exercise 8.2 (or another of your choosing), lay out the "big picture" aspects of a monitoring program in the form of a proposal to top management:
- (A) Set out the objectives:
 - Identify the key user(s) and the prime use(s)
 - Pose the key questions to be answered by the monitoring program
 - (B) Will this be a single, ad hoc study or a monitoring system? Why?
 - (C) Describe how the monitoring program will be managed:
 - Who will do the monitoring?
 - What incentives will you use?
 - What resistances do you anticipate?
 - How will you evaluate performance?
 - (D) Propose a schedule for the monitoring effort.
 - (E) Designate the resources required and justify them.
- 8.4** (Project-suitable) Select a topic of interest (a technology or the context for one company's business interests) and conduct Level 1 (or cold start) monitoring. Follow Section 8.3.3, using suggested sources of information to fulfill the four objectives. Your brief report should include a technology description, identification of related technologies (sketch relationships as done in Figure 8.1), identification of key players, and setting out possible development pathways. In addition, note what you found to be promising hard copy sources (such as key serials, books) and on-line data bases.
- 8.5** (Project-suitable) Carry your project forward, conducting a Level 2 (or warm) monitoring program on the topic developed in Exercise 8.4 or on a topic with which you are already familiar. Choose among the following activities:
- (A) Conduct a focused on-line search to round out what you know about a particular technology. Classify how pertinent various sources are.
 - (B) Interview two experts to verify the accuracy of your information. Have 5 to 10 questions ready to help you find out about unpublished sources, forecasts, other experts, who is promoting the technology, and so forth. Summarize what you learn from them.
 - (C) Create a historical chronology of the development of a particular technology or product line. Use this to suggest driving forces and key uncertainties. Consider a precursor analysis, qualitative or probabilistic (see Exhibit 8.3), if your data warrant. Alternatively, consider preparing a product roadmap, as discussed in Chapter 3. If neither approach appears useful in this case, explain why and offer your own analysis of the chronology.
 - (D) Map the competing and supporting technologies that affect this technology or product line. Use an impact matrix (Figure 8.2) to identify those

in which developments could constitute milestones for your technology or product line.

- (E) Map the pertinent contextual factors that affect, or are affected by, your technology or product line. Use an impact matrix to identify potential milestones.

- 8.6** (Project-suitable) Set up a “two-stomach system” for your technology (see Exhibit 8.2). You will need manila folders and an accessible storage site. If you have a project team, write your understanding of the topics to be covered, a deadline for information gathering, and who will be responsible for analyzing each of the topics. Agree on what those analyses will contain, their format, and their initial due date. If sensible, arrange to continue monitoring and have later topic updates scheduled.

--- 9

BASIC TOOLS FOR QUANTITATIVE TREND EXTRAPOLATION

OVERVIEW

This chapter lays the groundwork for the extrapolation of quantitative historical data into the future. It presents basic considerations in performing a trend analysis and then introduces simple models for short-term forecasting. It also discusses least squares estimation for straight lines and associated procedures that support the models in Chapter 10.

9.1 INTRODUCTION

Extrapolation uses the past to anticipate the future. Trend extrapolation uses mathematical and statistical techniques to extend quantitative time series data into the future.

Trend extrapolation is of use to the technology manager only if the future proves to be like the past, at least in some important respects. Observations of natural phenomena have led to the recognition of scientific laws that describe growth processes. These laws also model technological growth, or innovation, processes (see Chapter 2). Unlike many natural growth processes, technological growth relies on complex, socioeconomic factors. Human beings develop patterns of behavior based upon their experiences, and these patterns allow social scientists to anticipate events in a manner analogous to the predictions of physical scientists. While social relationships are more complex and uncertain than physical ones, the past is the richest source of information about the future. As the saying goes, "I'm sorry to be so stupid about the future, but I've had so little experience with it." We do

have vicarious experience with the future, however, to the degree that our past experiences condition our future responses.

A critical assumption underlies trend extrapolation of technological change: *Technical attributes often advance in a relatively orderly and predictable manner.* This is no trivial assumption; exceptions abound. Indeed, other chapters explicitly look for milestones, breakthroughs, or qualitative discontinuities. Figure 9.1 portrays decidedly irregular advances in cement production. Contrast this with the orderly trend shown in other figures in this chapter.

Fortunately, there is a second principle that suggests when one is likely to find technological change orderly and predictable: *A complex mix of influences moderates discontinuities.* This is like the rationale that statisticians use to test observations against a "normal distribution": the cumulation of many small contributions will result in a predictable pattern of deviations about a central value. With respect to technological change, this condition of multiple small influences is best met when progress reflects a series of ongoing engineering improvements that do not require scientific breakthroughs or major inventions. In terms of technological innovation (the commercialization and use of technical inventions), it is best met when adoption depends on market acceptance. That is, extrapolation of sales of a new technology that can be bought by thousands of purchasers will be more predictable than extrapolation that depends on only a few potential buyers.

Sometimes, a single decision determines the fate of a technology. For instance, development of the American supersonic transport airplane (SST) was derailed in the 1970s by the U.S. government's decision to cease major funding (a one-

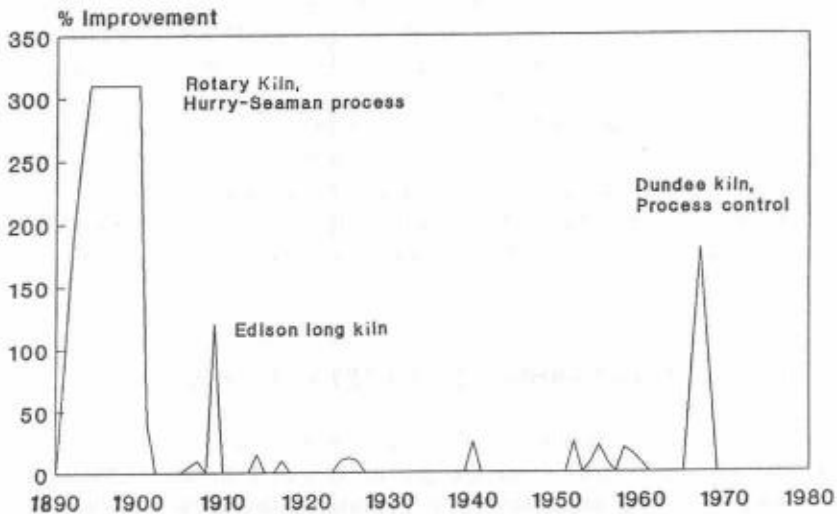


Figure 9.1 Production capacity of largest U.S. cement kiln, 1890–1980. (Reprinted from "Technological Discontinuities and Organizational Environments" by M. L. Tushman and P. Anderson published in *Administrative Science Quarterly*, Vol. 31, no. 3 by permission of *Administrative Science Quarterly*.)

vote margin in the Senate). Such technologies do not lend themselves well to trend extrapolation. (Contrast this example with cable television (CATV) covered later in this chapter. Decisions regarding CATV adoption are made by millions of households.)

Technological breakthroughs can cause fundamental discontinuities in societies. Therefore, extrapolations of technology cannot ignore the impact of technology itself. Nor is technology the only source of discontinuity. Events such as political upheavals, economic recessions, or natural disasters affect the reliability of the past as a guide to forecasting the future. These limitations in extrapolation imply that the forecaster should avoid making single, point value predictions. Any trend should be offered with confidence intervals. Furthermore, explicit sensitivity analysis should consider how a trend would be changed if critical events occurred. Such events could include advances in competing technologies, failure of a supporting technology to develop as expected, alterations in a socioeconomic trend (such as interest rate), or occurrence of a critical event (such as the Senate vote against the SST). Chapter 12 presents cross-impact models that specifically consider such interactions.

Extrapolation methods range from direct to structural techniques (see Table 4.2). The former seek to directly forecast behavior without explicitly considering effects of the broader society; the latter specifically consider cause and effect relationships. Extrapolation methods vary from very simple techniques to highly complicated statistical models that require powerful computers to apply. Sophistication of method, however, does not assure a valid forecast. Validity is seriously limited by the quality of available data.

Changing conditions and data limitations are two reasons why any forecaster should apply multiple forecasting tools. This chapter describes extrapolation tools that can contribute to technology forecasts. Trend extrapolation alone will not suffice. The chapter also describes how to select variables to forecast and presents naive models. Refinement of the basic naive model may consist of more carefully considering the past behavior of the variable, applying knowledge from analogous phenomena, or postulating causal relationships among variables. The key tool used in extrapolating various trends, least squares regression, is also presented. CATV growth provides a continuing subject for examples of trend extrapolation methods throughout the chapter.

9.2 SELECTING THE VARIABLES TO EXTRAPOLATE

In the case of CATV, the fraction of households with cable service provides an obvious variable by which to measure growth. However, in other cases in which the growth of technological capacity is to be forecast, the proper variable(s) is not always so obvious. There are at least three important criteria to apply in selecting an appropriate variable.

First, the variable chosen must measure the level of functionality of the technology. Lenz (1985, p. 14) suggests that "total output of comparable goods or services

by each technology exemplifies this requirement." An understanding of the technology and its application are required for choosing the correct variable. It is not always an obvious choice. For example, in Chapter 4 it was noted that while maximum speed is an appropriate measure of fighter aircraft technology, it does not accurately portray the functionality of cargo or passenger aircraft. A measure such as seat-miles per hour or ton-miles per hour helps determine capacity, which is an important measure as airlines serve high-density markets.

Second, the variable that is chosen must be applicable to both the new technology and to any older technology it replaces. The variable lumens per watt, for example, can be applied to lighting systems ranging from the candle to the latest electrical system, because energy consumption for all lighting systems can be expressed in terms of wattage.

Third, data must be available from which to compute historical values of the variable. This last criterion often is the most limiting. If the ideal indicator is not available or is less complete than alternative measures, a compromise may be necessary. For instance, in devising indicators of high-technology development among nations, Roessner and Porter (1990) substituted technology sales for production measures. Sometimes, available data are so remote that expert opinion or other methods should be used in lieu of extrapolating inappropriate variables.

9.3 NAIVE MODELS

This section describes the basic extrapolation models. These are applicable to short-term forecasting; as one projects further ahead, they lose credibility. These naive models do not explicitly account for the effects of the surrounding environment on technology. A series of models are developed progressively to show how the use of historical data improves the forecast. Note that the simplest models are presented only to show this progression; they are not serious extrapolation models.

In the simplest naive extrapolation, it is assumed that tomorrow, next week, or next year will be exactly like today—that is, no change is expected. Mathematically,

$$X_{t+1} = X_t \quad (9.1)$$

which implies that the value of X at time $(t + 1)$ is no different than the value at time t . Such a model is applied implicitly when current sales are said to be an estimate of future sales.

The forecast represented by Equation 9.1 can be seriously misleading—too conservative in a growing industry or wildly optimistic in an era of decline. For example, consider a firm developing a device for sale to CATV suppliers. This firm might use this model to forecast potential market size as the current number of CATV subscribers. However, data in Table 9.1 show that the number of subscribers has grown each year. Thus the potential for the product is greater than forecast. On the other hand, if a new satellite dish technology is developed that is preferable to cable, the market could evaporate quickly.

TABLE 9.1 U.S. Cable TV Subscribers (In Thousands)

Year	Subscribers
1975	9,800
1976	10,800
1977	11,900
1978	13,000
1979	14,100
1980	16,000
1981	18,300
1982	21,000
1983	25,000
1984	30,000
1985	31,300
1986	37,500
1987	39,700

Source: Statistical Abstract of the United States, 1989, p. 547.

Underestimating or overestimating the potential market can be detrimental. Developing a product for a dying industry obviously courts disaster. However, being too conservative about the potential also can be dangerous. Many fledgling companies, although first to market a successful technology, cannot exploit their position because they have inadequate capital to meet growing demand. In such cases, the firm may lose the market to better-prepared competitors or may be acquired by others with the resources to exploit market position. Thus even those companies seeking a conservative posture should use all available information in the forecast.

More sophisticated extrapolation can be illustrated using data for the growth of CATV subscribers presented in Table 9.1. First, consider the prediction of market size that would be made by a forecaster using Equation 9.1, the simplest naive model. If the forecast was made only once in 1975, market size in 1976 and thereafter would be 9,800,000 subscribers. As the data in Table 9.1 shows, this forecast would be in error by an amount that grows with time.

Figure 9.2 compares actual CATV growth with that predicted by a forecaster who used Equation 9.1 to reforecast the market *at the end of each year*. This is far superior to using Equation 9.1 to project many years ahead. Notice, however, that prediction of the same number of subscribers for 1987 through 1990 looks very strange in light of the historical data available to the forecaster. It also is clear that the model underpredicts market size in every year for which there is data. This analysis suggests that historical information that accumulates during growth can be used to improve the extrapolation provided by the simplest naive model.

Figures 9.3 and 9.4 show the results of somewhat more sophisticated naive predictions that consider historical data. The forecast used to develop the first figure predicts that incremental growth during the coming year will be identical to that in the previous year. In mathematical terms this *constant growth model* is shown as:

$$X_{t+1} = X_t + (X_t - X_{t-1}) \quad (9.2)$$

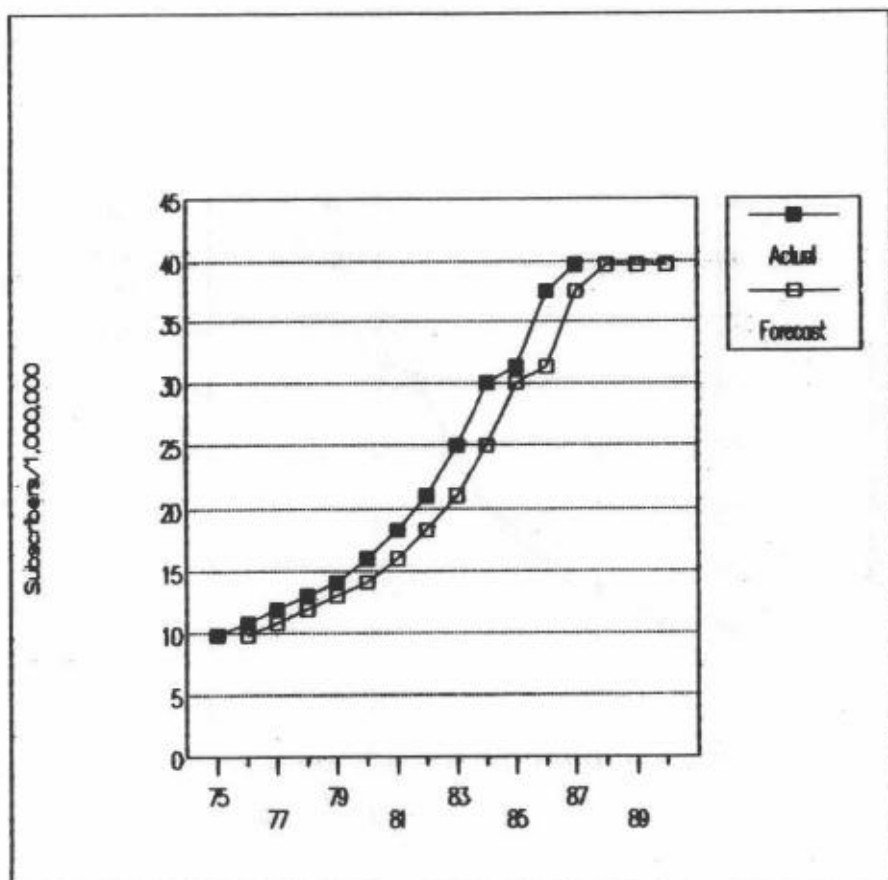


Figure 9.2 CATV subscribers in the United States: actual growth data and “No Growth” forecast.

Applying this approach at the end of 1977, a forecaster would have estimated the number of subscribers in 1978 to be

$$\frac{X_{1978}}{1,000} = 11,900 + (11,900 - 10,800) = 13,000$$

This would have been correct. Although Figure 9.3 shows that forecasts for other years are not quite as accurate, using historical information clearly improves the forecast.

Another variation on the use of historical data is provided by assuming that the fractional growth will be constant from one year to the next. That is

$$X_{t+1} = X_t(X_t/X_{t-1}) \quad (9.3)$$

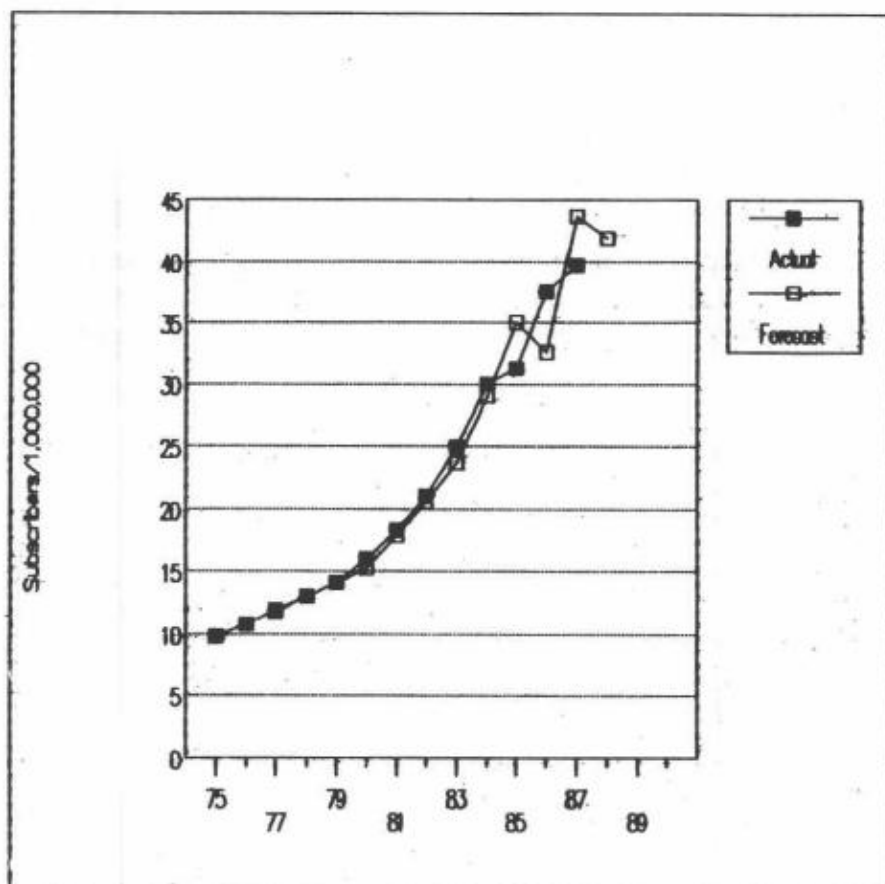


Figure 9.3 CATV subscribers in the United States: actual growth data and constant annual growth forecast.

Using this approach in 1977, a forecaster would have predicted the number of subscribers in 1978 to be 13,112 (in thousands). Figure 9.4 compares the forecast (using Equation 9.3) with the actual CATV growth. Note that the forecasts are very similar to those produced by Equation 9.2. However, as extrapolations are continued further into the future, the two forecasts will diverge due to compound growth effects.

Other models can be created as well (for example, either the increment or the fraction of growth could be averaged over the two preceding years). The important fact to be grasped from this discussion is that forecasts can be improved by using historical information about the process being forecast.

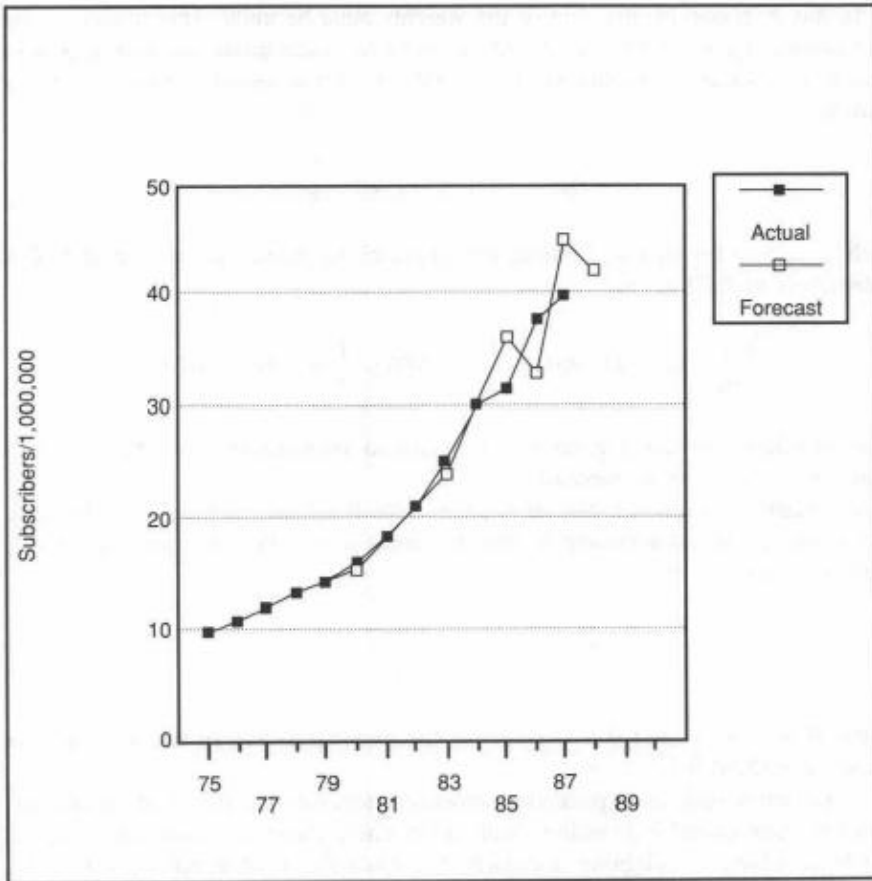


Figure 9.4 CATV subscribers in the United States: actual growth data and constant fractional growth forecast.

9.4 ADAPTIVE WEIGHTING

The techniques described in the previous section rely only on the most recent data points. As a result, valuable information that may be embodied by earlier data is lost or incompletely employed. This limitation is overcome by a class of models that forecasts future values using a weighted sum of past data. The general mathematical expression for such models is

$$X_{t+1} = \sum_{i=0}^n w_{t-i} X_{t-i} \quad (9.4)$$

where X_t is the value of the variable at time t and w_{t-i} is the weight that will be given to the value of the variable at time $(t - i)$.

In this formulation, the sum of the weights must be unity. This model is said to "smooth" the trend because it removes periodic fluctuations that may appear in year-to-year values. For example, one might use a three-period moving average by making

$$w_t = w_{t-1} = w_{t-2} = \frac{1}{3}$$

with $w_{t-i} = 0$ for all $i > 2$. Using this approach to project the number of CATV subscribers in 1978 gives

$$\frac{X_{1978}}{1,000} = \frac{1}{3}(11,900) + \frac{1}{3}(10,800) + \frac{1}{3}(9,800) = 10,833$$

The technique illustrated gives equal weight to information from the last three observations prior to the forecast.

Alternative approaches can be used to give the heaviest weight to the most recent data point and declining weights to earlier points. For example, *exponential smoothing* sets

$$w_{t-i} = \frac{a(1-a)^i}{d}$$

where $0 < a < 1$, and d is the sum of the numerators for all the w_i terms, as shown in Exhibit 9.1.¹

Approaches such as exponential smoothing assume that the most recent data embody more useful information than earlier data. There are many other uses of the basic adaptive weighting approach. For example, if observations were on a quarterly rather than a yearly basis and showed seasonal fluctuations, then taking nonzero values for w only in periods $t-4$, $t-8$, $t-12$, and so on would produce a type of seasonal adjustment.

Adaptive weighting does not make use of the trend information per se. In this respect it is more like the constant model than the constant growth or constant fractional growth models. Such techniques are best used to smooth "lumpy" time series (such as daily stock prices) so that prevailing trends can be identified. They are less useful as a forecasting tool per se.

There are many other methods for improving the accuracy of naive models; the models, however, are still naive in the sense that they only use the past behavior of a variable to determine its future path.² The forecaster can, for example, use a

¹For a more extensive discussion of exponential smoothing, see Chisholm and Whitaker (1971, pp. 22-26) or Granger (1980, pp. 88-90).

²Very good discussions of such methods can be found in textbooks for forecasting such as Chisholm and Whitaker (1971) and Wheelwright and Makridakis (1980).

■ Exhibit 9.1 Exponential Smoothing

Forecast an exponentially smoothed value for the number of CATV subscribers in 1978 using $a = 0.5$ in a three-period model.

$$w_{t-i} = \frac{a(1-a)^i}{d}$$

$$w_{t-1} = \frac{0.5(1-0.5)^1}{d} = \frac{0.25}{d}$$

$$w_{t-2} = \frac{0.5(1-0.5)^2}{d} = \frac{0.125}{d}$$

$$w_{t-3} = \frac{0.5(1-0.5)^3}{d} = \frac{0.0625}{d}$$

$$d = 0.25 + 0.125 + 0.0625 = 0.4375$$

This yields weights of $w_{t-1} = 0.571$, $w_{t-2} = 0.286$, and $w_{t-3} = 0.143$. The estimate of X for 1978 is

$$\frac{X_{1978}}{1,000} = 0.571(11,900) + 0.286(10,800) + 0.143(9,800) = 11,285$$

The value of a can be selected on the basis of the forecaster's judgment or as the value that works best when applied to past observations. Values between 0.2 and 0.4 are generally in order. ■

computer to determine the weights to apply to new observations to minimize forecasting errors (see Wheelwright and Makridakis, 1980, pp. 85–100). However, the general availability of computer facilities also makes it attractive to use statistical forecasting techniques, which may be superior to techniques that rely solely on weighted sums of past observations. A discussion of linear regression analysis and its use in forecasting lays the groundwork for understanding such statistical techniques.

9.5 LINEAR REGRESSION

Linear regression provides the essential tool to determine equations for straight-line relationships. Such equations can be used to extrapolate into the future. They can also be used to fit nonlinear relationships if those relationships can be mathematically transformed into forms that are linear.

9.5.1 Regression Model Validity

The previous sections considered techniques for using past observations to project future values. These techniques were called naive because they did not use information about interactions between the forecast variable and other variables that affect it. At times, however, there is information about such relationships that can add to the forecaster's knowledge of the process—knowledge that can be used to improve forecast quality (Note the discussion in Exhibit 9.2).

For example, to forecast the development of temperature-resistant materials (such as those used in the space shuttle heat shield), it might be expected that the rates of development and diffusion are closely related to federal expenditures on the shuttle program. These expenditures fund research to improve materials characteristics, and to advance production and manufacturing processes. As a result, better materials will be developed, produced, and eventually applied to products manufactured and sold in the private sector. Thus the forecaster might hypothesize that knowledge of projected federal funding could be used to improve predictions about the development of temperature-resistant materials and about their subsequent diffusion in industry.

Knowledge of past causal relationships often can be coupled with independent projections of related variables and used with regression analysis to develop a structural model that will provide an excellent forecast. For example, past data on space expenditures and on the production of heat-resistant materials might be used to formulate a functional relationship linking these variables to material use. Then projections of expenditures on space programs prepared by the Office of Management and Budget could be used with the model to forecast material use.

As a starting point, consider an example of simple regression analysis in which the forecast variable, Y , is thought to be linearly related to one independent variable, X . The model of this causal relationship would be expressed as

$$Y = a + bX + e \quad (9.5)$$

where a and b are the intercept and slope of the function and e is the error between the value of Y predicted by the causal relationship ($a + bX$) and the observed value of Y . This error always exists because the causal relationship does not completely determine the value of Y . For example, Y may be slightly affected by variables W and Z , as well, or the causal relationship may not be quite linear. The goal of regression is to determine values for a and b that minimize the error.

9.5.2 Fitting the Regression Equation

Regression is the name given to a technique for estimating the values of the constants a and b in the linear model given by Equation 9.5. The least squares criterion used to estimate these constants is that they should be chosen to minimize the sum of the squares of the errors between the observed value of Y and the value predicted by the linear model ($a' + b'X$). The primes indicate that these values are estimated

■ Exhibit 9.2 Causality and Regression

"Correlation does not prove causation." This well-worn phrase deserves thoughtful consideration by forecasters. Some correlations may be spurious—the stock market rises in years when the National Football Conference wins the Super Bowl (we cannot conceive a plausible causal mechanism, but who knows!). Other correlations leave one uncertain as to the direction of causal influence. For instance, imagine that government units that spend more on social services have greater unemployment rates. Does the unemployment induce social services expenditures, or do social services decrease people's incentive to find work? In other instances, correlation between two variables may reflect the influence of a third. During the 1980s, U.S. expenditures for the Army increased, as did those for the Navy. Rather than one driving the other, it is far more likely that both reflect a political decision to enhance military preparedness.

Regression among two or more variables can begin from various standpoints. In the strongest case, a solid theoretical basis is reflected in the regression equation. In the weakest, a data base is explored in order to discover correlation among certain variables. In this case, it should not be assumed that causation has been "proven" in any sense.

Four levels of regression modeling should be distinguished (Porter et al., 1981):

1. *Descriptive modeling*—Examination of a data set uncovers correlations; no basis for prediction.
2. *Simple prediction*—Even though the underlying causal processes, are not understood, continuity (stationary nature of behaviors and structures) is assumed to make usable predictions (e.g., stock market models based on previous data patterns rather than structural models).
3. *Causal models*—The underlying causal process generating the present data is known and correctly specified; however, the analysts do not know how the causal system will respond to manipulations (for instance, an ecosystem in the valley may be well described, but construction of a new highway would change it so drastically that no one can predict the results).
4. *Causal predictive models*—The underlying causal process is correctly specified and the results of deliberate interventions can be predicted. This requires knowledge that a change in the manipulable variable will have a sufficiently strong effect in the intended direction (significant regression coefficient), that this effect will not be swamped by other effects, and that the causal system will remain stable while the variable is manipulated. For example, assume production capability for a new technology-intensive product is increased. To predict sales, the system must not be perturbed by competitors increasing their capacity or developing newer products and the market's response to increased supply must be understood.

The forecaster must assess the causal strength of any regression model before using it to forecast. ■

by applying a specific criterion; usually they are not the true values. In practical terms, squaring the deviations between the estimated regression line and the data points places great weight on points far from the line. This makes it important to examine the data carefully before fitting a regression line (probably by using a computer statistical package). *Plot the data first.* Such a plot should be inspected to identify possible "outlier" data points and to decide whether they should be left as is, corrected, or excluded. Such decisions rest on knowledge of the underlying relationships and familiarity with how the data were measured. In addition, it should be determined whether a nonlinear relationship exists. For example, visual inspection of the CATV data (Figure 9.2) should warn that a linear fit will not work well.

The example of CATV growth can be used to illustrate simple regression. The simplest case would be to derive a linear relationship between CATV sales and time. This relationship can be fitted using two-variable linear regression:

$$y_i = a + bX_i + e$$

where i refers to each of the individual data points. To predict Y from X , use the following four-step calculation (in practice one is likely to use a statistical package—such as SPSS™ from SPSS Inc. in Chicago—or a programmable calculator).

First, determine the values of a and b^* that best fit the available data. These values are computed from the following equations:

$$b = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}$$

$$a = \bar{Y} - b\bar{X}$$

where

$$x_i = X_i - \bar{X}$$

*This b is the regression coefficient of Y on X . A different line, with a different slope, typically results if one were to regress X on Y —that is, to predict X values from the observed Y values. The formula for b in that equation would include the sum of the squared Y deviation scores in the denominator instead of the squared X deviation scores.

The correlation coefficient, r (the Pearson product-moment correlation), can be calculated as

$$r = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i^2}}$$

or

$$r = b(s_x)/s_y$$

Conceptually, imagine a scatterplot of dots in two dimensions that represents the X , Y values. The slope of the best-fitting least squares line is given by b , and r indicates how closely the points fall to the line.

The small x (or small y) values, which are called deviation scores, are the difference between the observed X (or Y) values and the mean of the observed sample. Their use simplifies the notation and the calculation.

$$\bar{X} = \sum_{i=1}^n \frac{X_i}{n}$$

where $\bar{X}$ is the mean and n is the number of observations in the data sample.

$$s_y^2 = \frac{\sum_{i=1}^n y_i^2}{(n-1)}$$

where s_y^2 is the variance in Y , or the square of the standard deviation in Y .

$$s_e^2 = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{(n-2)}$$

where s_e^2 is the standard error and $\hat{Y}$ represents values of Y predicted by using the equation

$$\hat{Y} = a + bX$$

In this sample calculation, X = time and Y = CATV subscribers (from Table 9.2). It is assumed that only the first four observations are available to calculate the least squares regression line (to simplify the numerical manipulations, let X = years - 1900):

X	x	x^2	Y	y	xy	$(Y_i - \hat{Y}_i)$	$(Y_i - \hat{Y}_i)^2$
75	-1.5	2.25	9,800	-1,575	2,362.5	30	900
76	-0.5	.25	10,800	-575	287.5	-40	1,600
77	0.5	.25	11,900	525	262.5	-10	100
78	1.5	2.25	13,000	1,625	2,437.5	20	400

$$\bar{X} = \frac{(75 + 76 + 77 + 78)}{4} = 76.5$$

$$\sum_{i=1}^n x_i^2 = 5.0$$

$$\bar{Y} = \frac{(9,800 + 10,800 + 11,900 + 13,000)}{4} = 11,375$$

$$\sum_{i=1}^n x_i y_i = 5,350$$

$$b = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2} = \frac{5,350}{5.0} = 1,070$$

$$a = \bar{Y} - b\bar{X} = 11,375 - [1,070(76.5)] = -70,480$$

The prediction (regression) equation is thus:

$$\hat{Y}_i = -70,480 + [1,070(X_i)]$$

Next examine the residuals (differences between predicted and actual values of Y ; that is, $Y_i - \hat{Y}_i$). Note that the small residuals do not guarantee a good fit for the extrapolation (see Figure 9.5). The sum of the squared residuals $(Y_i - \hat{Y}_i)^2$ from the preceding sample calculation equals 3,000.

Next extrapolate values of Y or X using the best fit equation.

- a. In what year will CATV subscriptions reach a certain level—say, 30,000 (in thousands)? To answer, solve the prediction equation by substituting 30,000 for Y :

$$30,000 = -70,480 + [1,070(X)]$$

$$X = 93.9 \text{ or the year } 1993.9$$

This year, 1993.9, is far later than the actual year, 1984, in which CATV subscribers reached 30,000 (in thousands). This reflects the poor fit between this line and the true nonlinear growth pattern (see Figure 9.5).

- b. What is the predicted number of CATV subscribers in a certain year—say, 1987? To answer, solve the prediction equation by substituting 87 (for the year 1987) for X :

$$\hat{Y} = -70,480 + [1,070(87)]$$

$$\hat{Y} = 22,610$$

Again, note that the predicted value falls far short of the actual. The farther out the extrapolation, the worse the prediction becomes with this misspecified model.

Finally, compute confidence intervals (C.I.) for the predicted Y . To illustrate, use the $\hat{Y}$ value for 1987 calculated in the previous step:

$$\text{C.I.} = \pm t_{\alpha/2, \text{d.f.} = n-2} \sqrt{s_e^2 \left(1 + \frac{1}{n} + \frac{(X_* - \bar{X})^2}{\sum x_i^2} \right)}$$

where t is a statistical value from a t -table and X_* is the value of X at the target extrapolation year.

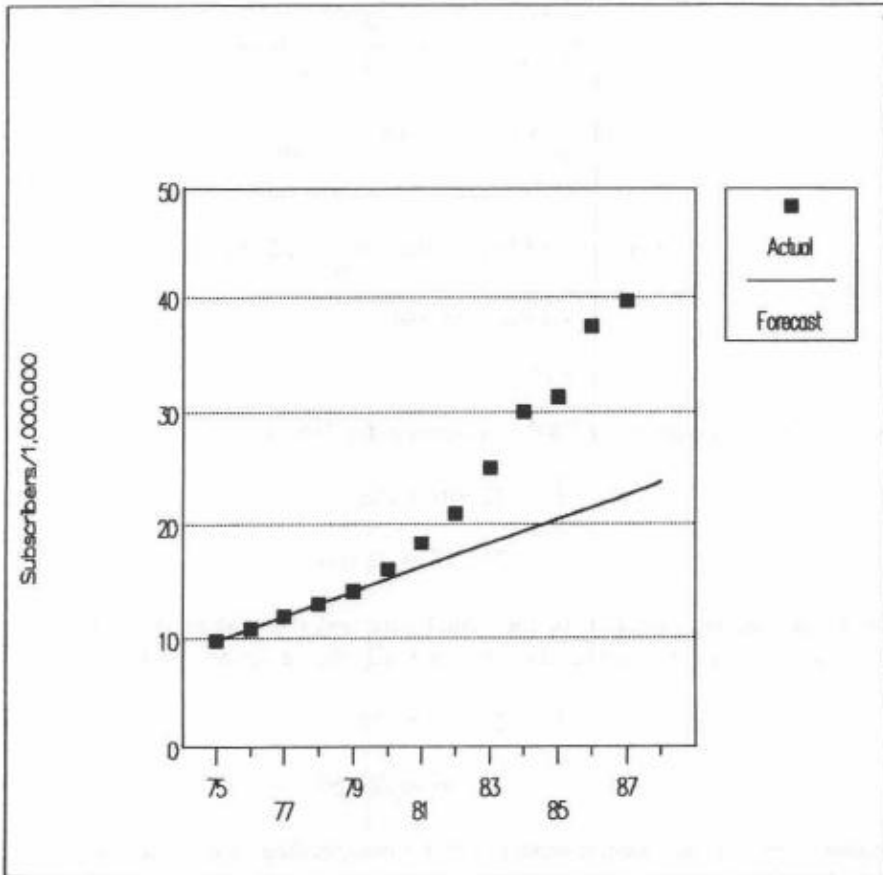


Figure 9.5 CATV subscribers in the United States: actual growth data and linear growth projection based on the first four data points.

Note the three terms, multiplied by s_e^2 , within the square root. These reflect three sources of error in the extrapolated prediction: (1) error about the regression line, (2) error about the mean, and (3) error in the extended slope.

For example, assuming a 50 percent C.I.:

$$\alpha = 1 - \text{C.I.} = 0.50$$

$$\frac{\alpha}{2} = 0.25$$

$$\text{Degrees of freedom (d.f.)} = N - 2 = 2$$

Look up the corresponding t value in Appendix B.2: $t_{0.25,2} = 0.816$ (t values derive from a statistical distribution discussed in detail in most statistics books.)

$$s_e^2 = \sum_{i=1}^n \frac{(Y_i - \hat{Y}_i)^2}{(n-2)} = \frac{3,000}{2} = 1,500$$

$$\frac{(X_* - \hat{X})^2}{\sum x^2} = \frac{(10.5)^2}{5.0} = 22.05$$

$$\begin{aligned} \text{C.I.} &= \pm 0.816 \sqrt{1,500 \left(1 + \frac{1}{4} + 22.05\right)} \\ &= \pm 0.816 \sqrt{34,950} \\ &= \pm 153 \end{aligned}$$

So our predicted number of CATV subscribers for 1987 is

$$\begin{aligned} \hat{Y} &= 22,610 \pm 153 \\ &= 22,457 \text{ to } 22,763 \end{aligned}$$

Given the limited precision of the actual data and the long-term nature of the forecast, this should be rounded off to fewer significant figures, such as

$$\begin{aligned} \hat{Y} &= 22,600 \pm 150 \\ &= 22,450 \text{ to } 22,750 \end{aligned}$$

Again, note that this precise model is badly misspecified; it does not include the actual 1987 Y value (39,700) within this C.I.

The lesson to be learned from this example is that simple linear relationships between the forecast variable and time are unlikely to do a very good job as predictors of technology growth. However, regression can be a powerful tool, as later sections show.

In some cases, it may be necessary to approximate a linear regression equation for a large data set (say, in the absence of a handy computer regression program). Exhibit 9.3 illustrates a method for calculating such approximations.

■ Exhibit 9.3 Three-Point Method for Approximating a Linear Regression

Granger (1980) presents a weighted, three-point method for fitting a linear function to time series data (Y_t). This method may be easier to apply quickly for some users than the standard least squares fit. The technique assumes an odd number of data points, n . If the number is even, the forecaster should discard the earliest data point. Granger refers to the case when $n > 10$ as the "ordinary" series; when $n < 10$ it is referred to as the "short" series. Three averages are used: the weighted averages

of the initial, middle, and final terms. For an ordinary series, these averages are defined as follows:

$$R = \frac{(Y_1 + 2Y_2 + 3Y_3 + 4Y_4 + 5Y_5)}{15}$$

$$S = \frac{(Y_{d-2} + 2Y_{d-1} + 3Y_d + 4Y_{d+1} + 5Y_{d+2})}{15}$$

$$T = \frac{(Y_{n-4} + 2Y_{n-3} + 3Y_{n-2} + 4Y_{n-1} + 5Y_n)}{15}$$

For a short series they are defined as follows:

$$R = \frac{(Y_1 + 2Y_2 + 3Y_3)}{6}$$

$$S = \frac{(2Y_{d-1} + 3Y_d + 4Y_{d+1})}{6}$$

$$T = \frac{(3Y_{n-2} + 4Y_{n-1} + 5Y_n)}{6}$$

For each series, the central point, d , is given by

$$d = \frac{(n+1)}{2}$$

For either series, the curve to be fitted is

$$Y(t) = a + bt$$

For example, for the Fisher-Pry model, $Y(t) = \ln[(1-f)/f]$ where f is the fraction of households with CATV, $a = \ln(c)$ and $b = -b$ (see Equation 10.4). For an ordinary series, the equations for a and b are as follows:

$$a = R - \frac{11b}{3}$$

$$b = \frac{T - R}{n - 5}$$

For a short series, the equations are:

$$a = R - \frac{7b}{3}$$

$$b = \frac{T - R}{n - 3}$$

The first 31 CATV data points presented in Table 10.1 (see p. 180) will be used as an example. For 31 data points, $n = 31$ and $d = 16$.

For the first five points, the following data apply:

YEAR	FRACTION WITH	$Y(t)$
1955	.005	5.293
1956	.009	4.701
1957	.009	4.701
1958	.011	4.499
1959	.013	4.330

For the middle five points, the data are as follows:

YEAR	FRACTION WITH	$Y(t)$
1968	.044	3.079
1969	.061	2.734
1970	.077	2.483
1971	.088	2.338
1972	.097	2.231

For the final five points, the data are as follows:

YEAR	FRACTION WITH	$Y(t)$
1983	.372	0.524
1984	.412	0.356
1985	.446	0.217
1986	.468	0.128
1987	.487	0.052

Thus,

$$R = \frac{[5.293 + 2(4.701) + 3(4.701) + 4(4.499) + 5(4.330)]}{15} = 4.563$$

For this same example, $S = 2.434$ and $T = 0.177$. Therefore,

$$b = \frac{0.177 - 4.563}{31 - 5} = -0.169$$

and

$$a = 4.563 - \frac{11(-0.169)}{3} = 5.179$$

The equation for the fit is

$$Y(t) = \ln[(1 - f)/f] = a + bt = 5.179 - 0.169t$$

where $\ln(c) = 5.179$ ($c = 177.5$) and $b = 0.169$. ■

9.5.3 Using Causal Relationships

A forecaster projecting CATV growth might hypothesize that as real personal incomes increase, consumers would be more willing to pay for cable TV. Therefore, there could be a causal relationship between disposable income and CATV growth. Perhaps, one could use disposable income to predict subsequent CATV subscriber growth. This example, chosen only to exemplify the calculation, assumes that the relationship between the two variables is linear. In practice, this assumption must be carefully evaluated. It is certainly plausible for CATV subscribers to relate linearly with one variable (disposable income) and nonlinearly with another (time). It should also be noted that time can be treated as a causal variable and used in forecasts with the techniques described here. The earlier discussion of regression on time merely sought a linear trend; this section extends the search to more sophisticated relationships among variables.

To check this hypothesis, the forecaster tabulates the number of CATV subscribers versus real disposable income. The results suggest that a linear causal relationship between these two variables, rather than between CATV subscribers and time, might form a workable model for CATV growth.

TABLE 9.2 Comparisons of Actual and Predicted Data Regarding CATV Subscribers

Year	Actual ^a		Predicted ^b	Residuals ^b
	Income	Subscribers	Y_i	$Y_i - \hat{Y}_i$
1975	875	9,800	5,979	3,821
1976	904	10,800	8,357	2,443
1977	940	11,900	11,309	591
1978	983	13,000	14,835	-1,835
1979	1,012	14,100	17,213	-3,113
1980	1,027	16,000	18,443	-2,443
1981	1,042	18,300	19,673	-1,373
1982	1,054	21,000	20,657	343
1983	1,088	25,000	23,445	1,555
1984	1,156	30,000	29,021	979
1985	1,190	31,300	31,800	-500
1986	1,238	37,500	35,749	1,751
1987	1,255	39,700	37,169	2,531

Source: Statistical Abstract of the United States, 1989, p. 547.

^aIncome in billions of 1972 dollars; subscribers in thousands.

^bIn thousands.

Using the linear regression approach described earlier for the full 1975–1987 data set,³ the linear model gives

$$\hat{Y}_i = -65,771 + 82X_i$$

where $\hat{Y}_i$ is the predicted number of CATV subscribers (in thousands) in a given year and X_i is the real disposable income in billions of 1972 dollars in the same year. Comparisons of predicted and observed values are given in Table 9.2 and in Figure 9.6.

Once the forecaster is satisfied that regression has produced a useful model (see Section 9.5.4), the next step is to use that model to forecast. If the forecaster knows or can estimate a specific value of X , call it X_* , then a corresponding value for Y can be produced by the model. Let us estimate real disposable personal income in 1990 as \$1,360 billion (in 1972 dollars). This estimate might be obtained from long-term economic growth forecasts made by the government, or it may be estimated by the forecaster. The forecasted number of cable subscribers (in thousands) in that year would be

$$\hat{Y} = -65,771 + 82(1,360) = 45,749$$

³The precise values of a and b will vary according to how you round off the calculations. For instance, the TOOLKIT calculates a as -66,137 and b as 82.69 for these data. The extrapolated values and statistical significance levels are very much the same from these estimates.

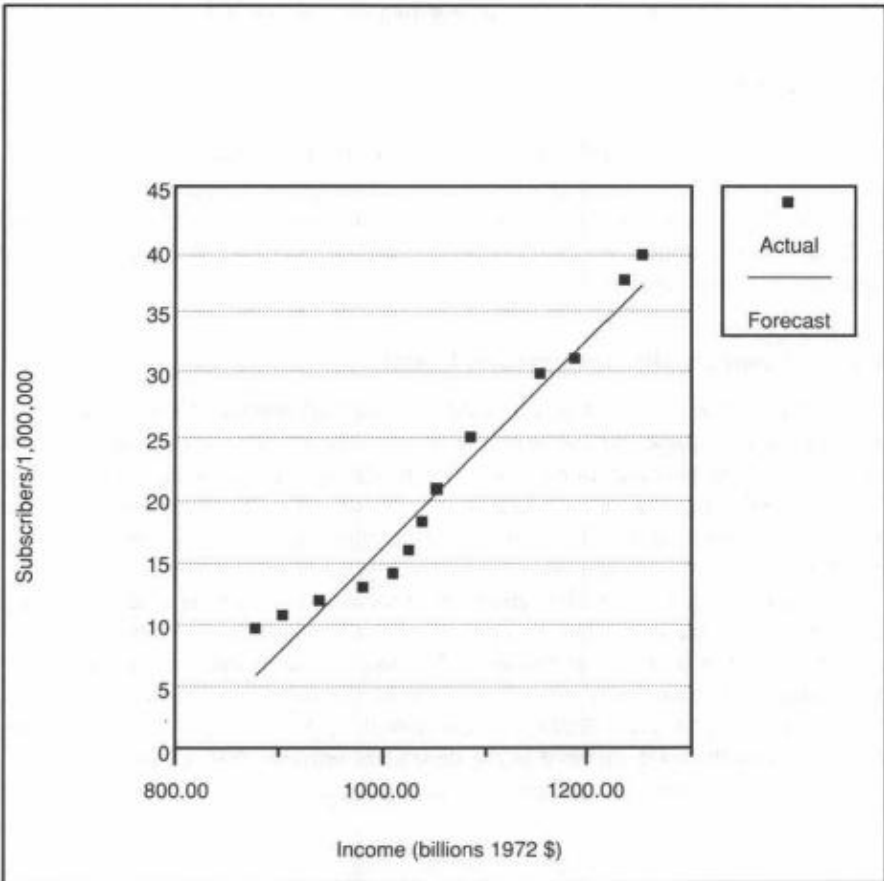


Figure 9.6 CATV subscribers in the United States vs. U.S. disposable income: actual growth data and linear regression of subscribers on income.

The statistical model makes it possible to estimate intervals, rather than simply to make point predictions. Thus it is possible to generate upper and lower bounds for the forecast that have a specific probability of trapping the true value of Y between them—that is, it is possible to determine a *confidence interval* (C.I.). In the cable TV example, a 90 percent C.I. might be generated for the number of subscribers when real disposable personal income is \$1,360 billion. A 90 percent C.I. would be calculated as follows:

$$\begin{aligned}
 \text{C.I.} &= \pm (t_{\alpha/2, d.f. = n-2}) \sqrt{s_e^2 \left[1 + \frac{1}{n} + \frac{(X_* - \bar{X})^2}{\sum_{i=1}^n x_i^2} \right]} \\
 &= \pm (1.796) \sqrt{5.0 \times 10^6 \left[1 + \frac{1}{13} + \frac{(1,360 - 1,058.8)^2}{179,232.2} \right]} \\
 &= \pm (1.796) \sqrt{5.0 \times 10^6 (1.583)}
 \end{aligned}$$

$$= \pm 5,053$$

or, rounding off:

$$\hat{Y}_* = 45,750 \pm 5,050 = 40,700 \text{ to } 50,800$$

(t is taken from Appendix B.2, and all other values were computed from the data in Table 9.2.) Confidence limits can also be used as the basis for best case/worst case scenario construction.

9.5.4 Evaluating the Regression Model

Section 9.5.3 developed a structural model for the dependence of CATV growth on real disposable income. Before applying it, however, investigate the validity of the model by asking two questions: (1) Is the model actually effective in explaining CATV growth? and (2) Can the hypothesis that the true values of a or b are zero be rejected? These questions are answered by applying statistical tests. An F -test is applied to answer the first question, a t -test helps answer the second.⁴

It is important that the model actually be effective in explaining something about the dependent variable, Y . This is a necessary, but not sufficient, condition to forecast future values of Y (recall Exhibit 9.2). One way to evaluate the effectiveness of a model is to examine how well it explains the data being used. This can be done by means of the coefficient of determination, R^2 . This coefficient indicates the fraction of the total variance in the dependent variable that is explained by the model. The coefficient of determination is defined as

$$R^2 = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (9.6)$$

The value of R^2 is provided by the output of most statistical packages. R^2 for the CATV and disposable income example is 0.957. This implies that nearly 96 percent of the variance in the number of cable subscribers is explained by the income variable. Therefore, the model coupled with a good forecast of real disposable income should produce a good forecast of CATV subscribers, as long as the relationship and the pertinent contextual factors remain essentially unchanged.

While R^2 provides an indication of model effectiveness, there are statistical tests that can be used to establish the effectiveness of the model more rigorously. To use these tests, the forecaster must compute the values of the sample variance, s_y^2 and

⁴In this simple linear regression, the model involves a single, independent variable—in this case, disposable income. The two questions essentially are one. For example, if the hypothesis that the true value of b' , the coefficient of the independent variable, is zero cannot be rejected, then one cannot conclude that CATV growth depends on real disposable income. Thus the model will fail both the t -test and the F -test. However, it is worthwhile to illustrate how these tests are applied in this simplest case. The results can then be extended to multiple regression, which is introduced in Section 9.7.

its square root, the sample standard deviation, s_y . The sample variance and sample standard deviation are used in the F -test and t -test described in Exhibit 9.4.

A statistical test used to determine the adequacy of model fit is the F -test. This test is the equivalent of testing whether all coefficients of the regression equation could be zero. Two steps are required to apply the F -test. First, determine a critical value for the F statistic. This value depends on the degrees of freedom of the data set used to develop the model and on the confidence level to be placed on the question of model fit. Next, determine the value of the F statistic for the model. These steps are illustrated for the CATV model in Exhibit 9.4.

■ Exhibit 9.4 F -Test of the Regression Model

Critical F values are determined from standard tables such as those included as Appendix B.1. For this example, a portion of such a table is presented in Table 9.3. To use this table, determine the degrees of freedom (d.f.) of the numerator and denominator of the F statistic, defined as

$$F = \frac{\sum_{i=1}^n \frac{(\hat{y}_i - \bar{y})^2}{k}}{s_e^2}$$

The d.f. of the numerator are the number of independent variables (k)—in this example, the number is one. The d.f. of the denominator equal the number of data points used to construct it (n), less the number of independent variables in the model, minus 1—in this example, the d.f. of the denominator (k) equal $13 - 1 - 1$ or 11. Using 1 as the column entry and 11 as the row entry into Table 9.3, there are several critical values depending on the probability selected. For example, if a

TABLE 9.3 Portion of a Standard F Distribution Table^a

d.f. of Denom- inator	Prob- ability	d.f. of Numerator					
		1	2	3	4	5	...
10	0.25	1.49	1.60	1.60	1.59	1.59	...
	0.10	3.29	2.92	2.73	2.61	2.52	...
	0.05	4.96	4.10	3.71	3.48	3.33	...
	0.01	10.00	7.56	6.55	5.99	5.64	...
11	0.25	1.47	1.58	1.58	1.57	1.56	...
	0.10	3.23	2.86	2.66	2.54	2.45	...
	0.05	4.84	3.98	3.59	3.36	3.20	...
	0.01	9.65	7.21	6.22	5.67	5.32	...
12	0.25	1.46	1.56	1.56	1.55	1.54	...
	0.10	3.18	2.81	2.61	2.48	2.39	...
	0.05	4.75	3.89	3.49	3.26	3.11	...
	0.01	9.33	6.93	5.95	5.41	5.06	...

^aThe full F -table appears in Appendix B.1.

probability of 5 percent (that is, 0.05) is selected, the critical value of F is 4.84. This implies that in constructing many different models with a computed F value of 4.84, the forecaster would err in inferring that a real relationship exists between the dependent and independent variables 5 percent of the time. If the value of F were 9.65, the inference would be incorrect 1 percent of the time.

The next step in applying the F -test is to compute the F statistic for the model. The statistical information necessary to compute F is available from information developed to calculate a' and b' . For the model of disposable income with CATV subscribers, the mean number of calculated subscribers (in thousands) is 21,050 (the sum of column 3 in Table 9.2 divided by 13 years). Computing the numerator using the predicted values of Y in column 4 of Table 9.2, the mean value of Y , and the sample variance gives

$$F = \frac{1.206 \times 10^9 / 1}{5.0 \times 10^6} = 241$$

Since this value is far greater than the critical value for a 1 percent probability, reject the hypothesis that there is no relationship between CATV growth and real disposable income with much less than a 1 percent probability of error. ■

The t -test is performed to assess the probability with which the forecaster can reject the hypothesis that the coefficient b' is different from zero. This test is performed in much the same way as the F -test. That is, a critical value of t is determined from standard tables and then it is compared to the value of t computed for the data used in determining b' . A portion of a standard t -distribution table is shown in Table 9.4 (a more complete table is provided in Appendix B.2.) As before, the row entry is the number of degrees of freedom of the data computed as $(n - 1 - k)$ where n is the number of data points (13) and k is the number of independent variables in the model (1). From Table 9.4, the critical value of t for a 5 percent probability is 1.796, or 2.718 for a 1 percent probability. As before, these probabilities indicate the probability that rejecting the hypothesis that b' is different from zero will be incorrect (see Exhibit 9.5).

■ Exhibit 9.5 t -Test of the Regression Coefficients

As with the F -test, once the appropriate significance level has been set, one is ready to compute the value of the t statistic for the regression coefficient. This statistic is defined as

$$t = \frac{b'}{\sqrt{s_e^2 / \sum_{i=1}^n (x_i)^2}}$$

Recall that x is the deviation from the mean. The mean of X from column 2 of Table 9.2 is \$1,058.8 billion. Squaring the difference between each value in column 2

TABLE 9.4 Portion of a *t*-Distribution Table^a

d.f.	Probability					
	0.25 ^b		0.05 ^b	0.025 ^b	0.01 ^b	
...	0.50 ^c	...	0.10 ^c	0.05 ^c	0.02 ^c	...
...						
10	0.700	...	1.812	2.228	2.764	
11	0.697	...	1.796	2.201	2.718	
12	0.695	...	1.782	2.179	2.681	
...						

^aThe full *t*-distribution table appears in Appendix B.2.^bProbability for one-tailed tests = $\alpha/2$.^cProbability for two-tailed tests = α .

Note: The inferential test alluded to compares a null hypothesis with an alternative hypothesis. When that alternative hypothesis only asserts that the parameter in question is different from the one hypothesized, it is nondirectional, or two-tailed. Sometimes the alternative hypothesis is directional (that is, it might specify that the parameter in question is larger than some value); this sets up a one-tailed test. The tails pertain to the expected distribution of possible parameter values, as in a bell-shaped curve with a bulge in the middle (highly likely values) and narrowing as one moves toward the extremes (the lower and upper tails.)

and the mean, in turn, and summing the squares gives \$179,232.3 billion. Thus:

$$t = \frac{82}{\sqrt{5.0 \times 10^6 / 1.79 \times 10^5}} = 15.52$$

Comparing to the critical values of *t*, reject the hypothesis that *b'* is zero with much less than a 1 percent probability of error. ■

9.6 TRANSFORMATIONS

Many relationships are not linear. Most technological change processes follow nonlinear growth models—in particular, the S-shaped curves. The most effective way to handle these nonlinear models is to find an appropriate way to transform them to a linear relationship. Straight lines are best for graphic extrapolation. For instance, if a technology, *Y*, is growing exponentially, plotting log *Y* versus time will give a straight-line relationship. A number of information technology parameters (such as computer memory per unit volume) have maintained roughly exponential growth for several decades. These are typically plotted with the *Y* axis evenly divided into powers of 10. That is, the interval from 1 to 2 represents growth from 10 to 100

units; the interval from 2 to 3 represents growth from 100 to 1,000 units—decidedly nonlinear!

Transformation to a linear function can at times fit an equation to a nonlinear relationship. Exhibit 9.6 presents the most common transformations and illustrates how to perform the required calculations. These are elaborated further in Chapter 10.

Other functional forms can also be made linear. The power expression

$$Y = aX^b$$

can be transformed into a linear equation simply by taking the log of both sides to produce

$$\log Y = \log a + b \log X$$

which is linear if the dependent variable is taken to equal $\log Y$ and the independent variable as $\log X$. As further examples, the hyperbolic relationship

$$Y = a - b(1/X)$$

can be dealt with by treating $1/X$ as a variable, say X' . And, the parabolic relationship

$$Y = a + bX + cX^2$$

can be handled by treating X^2 as just another variable, say Z , in the linear relationship.

The approach detailed can be easily adapted to calculate a regression equation for one of the models listed in Exhibit 9.6 This is illustrated in Exhibit 9.7.

■ Exhibit 9.6 Common Transformations for Technological Growth Curves

GROWTH MODEL	TRANSFORMATION
Exponential	$Z = \log_{10} Y$ (or $\ln Y$)
Pearl (Fisher-Pry, single variable)	$Z = \ln[(L - Y)/Y]$ where L = upper growth limit
Gompertz	$Z = \ln[\ln(L/Y)]$
Substitution (Fisher-Pry, two variables)	$Z = \ln[f/(1 - f)]$ where L = upper growth limit and f = fraction of the market held by the new technology

■ Exhibit 9.7 Least Squares Calculations with a Transformed Variable

Consider the following imaginary growth statistics for unit sales by a small manufacturing company, Y , (in thousands). Assume that $L = 100$ (in thousands).

YEAR	Y	$Z = \ln[(L - Y)/Y]$
1985	9	2.314
1987	21	1.325
1989	41	0.364
1991	65	-0.619

First, the appropriate growth model must be chosen. (Growth models are discussed in Chapter 10.) This example assumes that a Pearl (or identically, a Fisher-Pry model for one variable changing over time) model is appropriate. Then calculate Z as shown. From this point, follow the same procedures as shown in Section 9.5.2 for Z instead of Y . For instance, calculate the deviation scores, z , and multiply these by the deviation scores, x . Sum these and divide by the sum of x_i^2 to determine b . In this illustration, the linear regression is:

$$\hat{Z} = 43.79 - [0.488(\text{year} - 1900)]$$

Extrapolating to 1997, for example:

$$\hat{Z} = -3.546$$

Next, take the antilog and solve for Y :

$$\frac{(100 - Y)}{Y} = .0288$$

$$Y = 97,200$$

9.7 REGRESSION AND FORECASTING—SOME EXTENSIONS

This chapter has introduced simple regression as a forecasting tool. Such an overview is necessarily brief and omits many important assumptions and qualifications. Additional regression topics can support management decisions that depend on forecasts of technology and its effects. This section introduces some useful extensions to simple regression methods.

Perhaps the most important extension is to expand simple regression analysis to consider simultaneously the causal influence of more than a single explanatory (independent) variable. Multiple regression often can provide a better explanation

of the past behavior of the forecast (dependent) variable and a better basis for predicting its future levels. In the example of CATV growth, the simple regression considered the effects of real disposable personal income on the number of subscribers and the R^2 indicated that much of the variance was captured. However, including the number of households composing the potential market or the hours of nonnetwork programming provided by CATV as variables might improve the fit of the regression and enhance its predictive power.

A multiple regression with two explanatory variables has the form

$$Y = a + bX + cZ + e \quad (9.7)$$

The estimates of a , b , and c once again are determined by the regression procedure so as to minimize the sum of squared errors. In the case of two independent variables, these estimates will specify a two-dimensional plane in a three-dimensional space (Y , X , and Z). However, the procedures can be generalized to provide for any number of independent variables as long as there are adequate numbers of observations available.

There are many computer packages available that will generate results similar to those given earlier for simple regression. Before extensive statistical inference or forecasting is done, the reader should consult expanded discussions on multiple regression because there are many pitfalls underlying the assumptions on which these models rest. Once results are available from the multiple regression, forecasting procedures are similar to those used in simple regression. First, the F -test will determine whether the equation has really explained anything about the dependent variable, and the R^2 will indicate the goodness of the fit of calculated values to actual data. Then, t -tests can be applied to determine whether specific coefficients are significantly different from zero (that is, whether the associated variable is important to explaining and predicting the dependent variable).

The generation of the C.I. for the forecast is somewhat more complicated because of the additional explanatory variables, but the typical statistical package printout provides all the necessary information.⁵

Another way time can be included in regression is by using dummy variables (a variable with a value of either 0 or 1) to seasonally adjust data and forecasts. For example, summer (S), fall (F), and winter (W) variables could be included in the equation.

$$Y = a + b_1X + b_2S + b_3F + b_4W + e$$

For this approach, the data presented in Table 9.1 would be expanded by adding three columns (summer, fall, and winter) following the year. Observations of the

⁵Econometrics textbooks such as those by Gujarati (1978), Johnston (1972), Wonnacott and Wonnacott (1979), and Goldberger (1964) are excellent sources for the techniques and problems of multiple regression. Forecasting methods books, such as those by Wheelwright and Makridakis (1980) or Granger (1980), also discuss the major issues, although they are not as extensive as those on econometrics.

number of subscribers made in the fall quarter, for example, would have a 1 in the fall column and a 0 in each of the winter and summer columns. Using this approach, the estimated coefficients of the regression will change the intercept of the regression line depending on the quarter. Thus, a forecast for a Fall Quarter value would be adjusted up or down depending upon how the number of subscribers changed during that season. In this example, the spring quarter provides the base from which the adjustments for the other quarters are made. This technique is analogous to seasonally adjusting data for projections, and most regression computer packages make it very easy to do.

Analysis of time series data seriously compromises a major assumption of the statistical analysis of regression equations—namely that observations, and errors in those observations, are independent of each other. Econometricians and others who work with time series data have developed sophisticated procedures to cope with this problem. Two-stage least squares estimation procedures and Box-Jenkins time series analyses are two notable approaches. These are beyond the scope of this book, and are not viable for most technological trend analyses. For instance, the Box-Jenkins approach first estimates the autocorrelation process in the series and then makes suitable adjustments so that the resulting series is well behaved. However, for this estimation to be credible, 50 data points or more are required. Rarely would a technology forecaster have the luxury of such extended series. Furthermore, the forecaster usually seeks to predict long-term change (five or more years in the future). The uncertainties of fast-paced technological change and complex socioeconomic influences dwarf the concerns of biased time series statistical tests.

EXERCISES

- 9.1 What type of additional assumptions are made when causal models are used instead of simple extrapolation? If they require more assumptions, how can causal models make the forecaster more confident about a projection?
- 9.2 Adaptive weighting schemes allow for flexibility in extrapolation. Identify three cases in which data may fluctuate in a cyclical fashion (for instance, snowmobile sales are probably seasonal). How might you use adaptive weighting to capture these cyclical characteristics to improve extrapolation in each of the three cases?
- 9.3 The TOOLKIT includes the even-numbered years of the time series of data for U.S. CATV subscribers (extended series compared to Table 9.1) under the file name CATV2.dat. The percent of households with CATV (Y) with time (X) is distinctly nonlinear (examine a plot of the series using TOOLKIT). However, consider the subset of the data for the even years 1966 through 1972. Use exponential smoothing for three periods—1966, 1968, and 1970—to predict a value for 1972, using $a = 0.5$. Repeat the calculation using $a = 0.4$. Repeat it once more using a different a value of your choice. Comment on

what difference the value of a makes. Comment on how well the exponentially smoothed values fit the actual data for 1972.

- 9.4 Using the same four data points as in Exercise 9.3—that is, percent of households with CATV for 1966, 1968, 1970, and 1972—calculate a linear regression equation by hand. Follow the method of Exhibit 9.3. Compare your coefficients a and b with those obtained by a linear model of the full series (use the TOOLKIT). Comment. Calculate R^2 for your four-point linear model.
- 9.5 Use the equation calculated in Exercise 9.4 to extrapolate the percentage of households with CATV through 1978. Compute 50 percent and 90 percent confidence intervals. Do the actual data fall within these confidence intervals? Comment.
- 9.6 Use the first four data points for disposable income versus CATV subscribers (see Table 9.2). Calculate a leading indicator model that predicts subscribers from income. Extrapolate subscribers using \$1,255 (in billions) as the income figure. Compare your prediction accuracy with the fit of the whole series model (available in the text or by manipulating file CATV.dat in the TOOLKIT). Calculate an F -test and a t -test for your four-point regression. Comment.
- 9.7 Repeat Exercise 9.6, but let subscribers be the X variable and income the Y variable. Again, predict subscribers using \$1,255 (in billions) as the income figure. Explain any difference from the previous result.
- 9.8 Give one technology forecasting example in which two or more variables appear to be related in a nonlinear way. Describe how you could still make use of linear regression to derive an equation for this example.

10

EXTRAPOLATING TECHNOLOGICAL TRENDS

OBJECTIVES

This chapter explains and illustrates the use of the most applicable forms of trend extrapolation for technology forecasting and presents trend analysis as a four-step process. The chapter emphasizes two key growth models, Fisher-Pry and Gompertz. Finally, it illustrates how the Lotka-Volterra equations offer a promising general framework for trend modeling.

10.1 TREND ANALYSIS IN TECHNOLOGY FORECASTING

Technology forecasting relies largely on naive (direct) time series analysis. This implies major assumptions about the nature and permanency of both context and structure, as discussed in Chapter 9. Trend analysis methods can yield valid forecasts when supporting and competing mechanisms in the larger environment remain constant over the time horizon of the forecast or when changes in these mechanisms cancel one another. Even under those appropriate conditions, trend extrapolations should be used in conjunction with complementary technology forecasting methods, especially expert opinion and monitoring. This chapter stresses extrapolation over time. It applies regression techniques to fit selected nonlinear relationships that are especially suited for technology forecasting.

10.2 STEPS IN TREND ANALYSIS

Once the variable(s) have been chosen and the necessary data have been obtained (see Chapter 9), trend analysis can begin. Exhibit 10.1 outlines the basic steps.

Step 1, Exhibit 10.1, model identification, draws upon insight into typical patterns of technological innovation (see Chapter 2), solid knowledge about what is driving the change and empirical evidence. *S-shaped growth* should be considered the most likely form (as discussed in Chapter 4). Section 10.3 contrasts the Fisher-Pry and Gompertz models—two approaches that produce S-shaped curves. However, other models sometimes merit serious consideration as well.

■ Exhibit 10.1 Steps in Trend Analysis

1. Identify the proper model. Prominent alternatives include:
 - a. S-shaped growth curves
 1. Fisher-Pry (or equivalently, Pearl)
 2. Gompertz
 3. Technical progress function
 - b. Learning curve
 - c. Exponential growth
 - d. Linear
2. Fit the model to the data
 - a. Graphically
 - b. Solve for the constants in the equation
3. Use the model to project
 - a. Graphically
 - b. Mathematically
4. Perform sensitivity analysis and interpret the projections.
 - a. Compute confidence intervals
 - b. Consider outside factors.

Exponential growth often holds over certain periods, or epochs (Hamblin, Jacobsen, and Miller, *et al.*, 1973). The growth rate then shifts and another epoch emerges. Over multiple epochs, continuing exponential growth becomes apparent, but succeeds at different rates. Exponential growth may hold for the time period of interest, but possible physical or social limits that could slow or stop growth must be anticipated.

An *envelope curve* can be constructed by stacking S-shaped curves, one after and over another. Figure 10.1 shows a classic example of how a series of technological developments, each S-shaped, can combine to drive a parameter forward, possibly along an exponential frontier. As one technology approaches a peak, R&D provides a successor that fulfills the appetite of a still-hungry market. An envelope curve could be plotted through the peaks and the valleys of the successive technologies to depict the general trajectory of the development pattern.

The *technical progress function* is another important model. It measures growth as a function of effort instead of time. As depicted in Figure 10.2A, the notion is

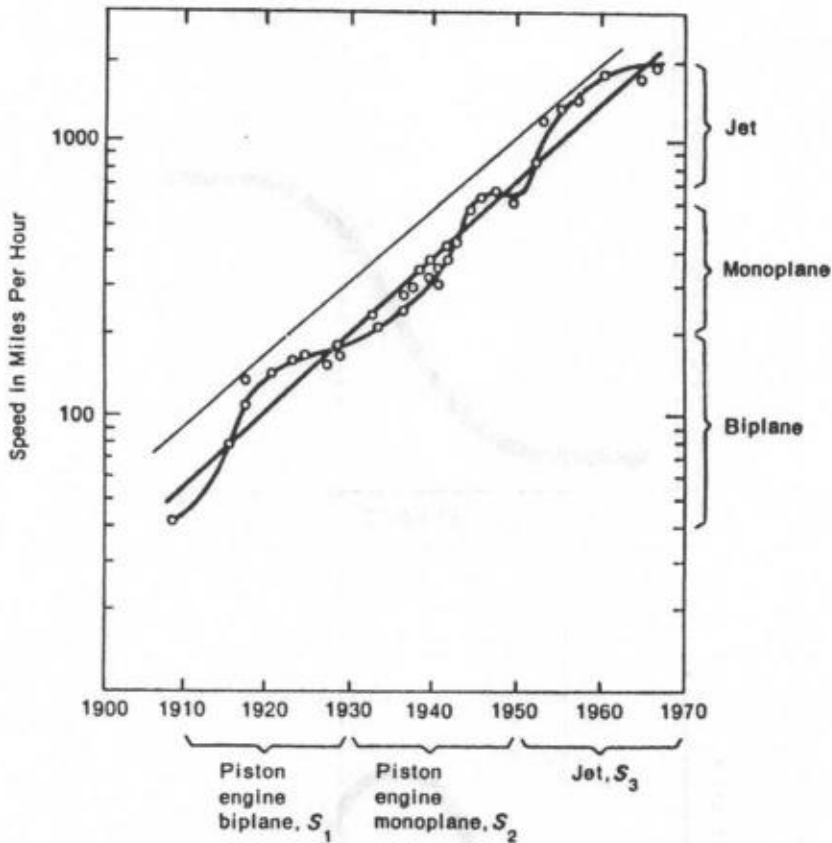


Figure 10.1 Envelope curve. (Reprinted by permission of the publisher from *Technological Forecasting for Decision Making* by J. P. Martino. Copyright 1983 by Elsevier Science Publishing, Co. Inc.)

that progress in developing a technology starts slowly as many impediments must initially be overcome, advances rapidly for a period, and then slows as the easy improvements are “mined out.” This is, of course, the S-shaped growth curve in another guise. The tapering off of technical progress for constant increments of additional effort (Figure 10.2A) implies dwindling productivity (Figure 10.2B). The R&D organization receives the greatest payoff from its effort during the steep portion of the technical progress curve (Figure 10.2A). After that, marginal returns per unit of effort diminish (later portion of Figure 10.2B). The technical progress function provides a vital signal to those who would manage R&D for a technology—that is, at some point continued investment in R&D will deliver less and less. This means that the natural momentum of continuing to do what has delivered good results in the past, must be stopped—that large, successful research group must be reassigned to fresh tasks or R&D productivity will drop.

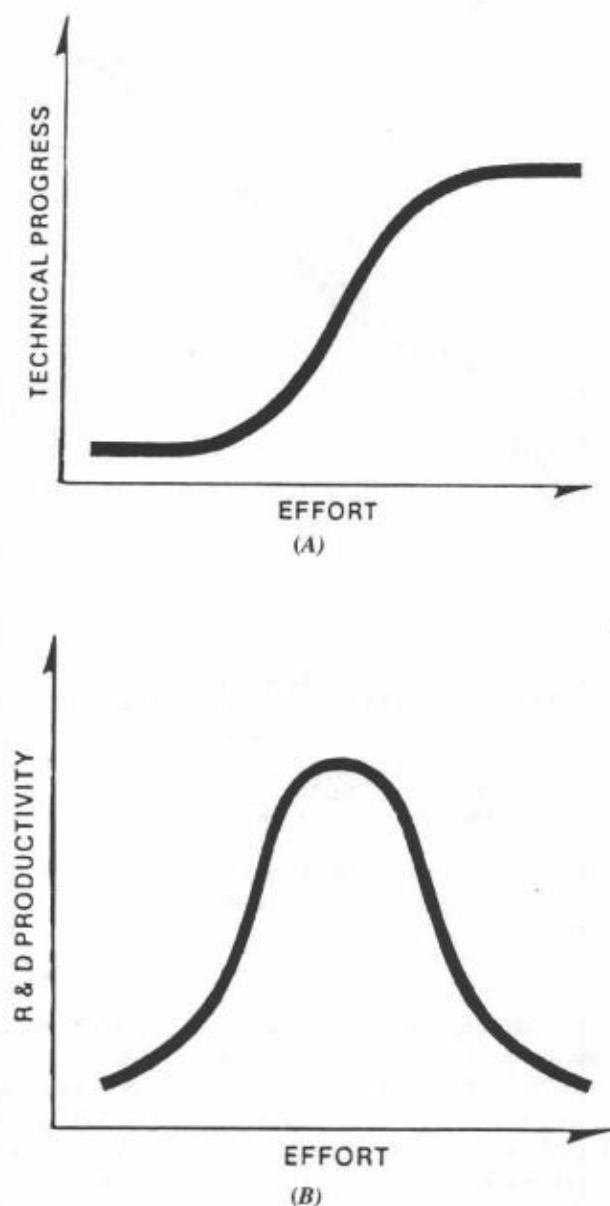


Figure 10.2 (A) The S-curve technical progress versus effort. (B) The R&D productivity curve. (Source: Becker and Speltz, 1986. Reprinted by permission. ©Industrial Research Institute.)

A related notion is the *learning curve*, which addresses the improvement in productivity often seen as a technology production process matures. This is conventionally depicted as a power function:

$$Y = aX^{-b}$$

or equivalently,

$$\log Y = \log a - b \log X \quad (10.1)$$

where Y is the number of direct labor hours required to produce the X th unit; a is the number of direct labor hours required to produce the first unit; X is the cumulative number of units produced (not time); and b is a parameter that measures the rate that labor hours are reduced as cumulative output increases (Argote and Epple, 1990).

Analysis of the learning curve for a technology can help predict declining production costs. The decline is attributable to organizational learning as personnel gain experience in refining the production process—in other words, the learning curve is due to both technology and people factors. Each increase in cumulative output leads to a reduction in unit production cost. While this “progress ratio” varies greatly from technology to technology, the modal value is about 80 percent—that is, each doubling of cumulative output leads to a 20 percent reduction in unit production cost (Argote and Epple, 1990). Figure 10.3 shows the learning curve for production of an advanced military jet built in the 1970s and 1980s. It shows the reduction in the number of hours needed to assemble the aircraft as workers learned the intricacies of construction, and less labor means a lower cost per unit. Projection of a learning curve can help a technology manager monitor his or her own production processes to ensure that learning is progressing reasonably. Projection of a competitor’s learning curve could help gauge whether to enter a market.

Having identified a promising model, the second step is to fit the model to the data. Should that fit prove poor, the forecaster may wish to reconsider the choice of model. The choice of a model should never be made by fishing through a grab bag of models and picking the one that fits the data best. Noisy data can mask true relationships; there is no substitute for solid conceptual underpinning.

Begin by graphing the data. Remember that individual data points may be problematic. This is important as outlier points can exert great influence over mathematical curve fits. Outliers may result from special circumstances, mistaken measurement, transcription errors, or just a divergent value that cannot be ignored. As discussed in Chapter 9, the forecaster may want to examine various transformations. For many forecasting purposes, fitting a line to properly transformed data will provide a satisfactory basis for extrapolation. Even if the equation for the trend is eventually calculated, the graph will provide an excellent check. It is easier to detect a bad extrapolation on a graph than from an equation.

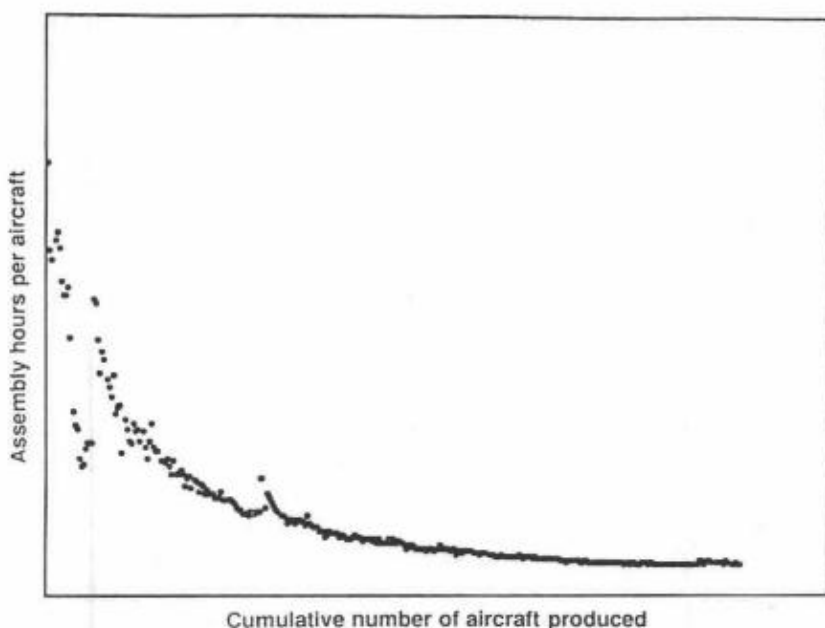


Figure 10.3 Learning curve—assembly hours per aircraft and cumulative number produced. (Source: Argote and Epple, 1990. Reprinted by permission of the author and publisher from *Science*, Vol. 247, pp. 920–924. Copyright 1990 by the AAAS.)

Details about how to fit an equation for a model based on a sample of data are provided in Chapter 9. Once the equation is determined, an extrapolation can be made graphically and/or mathematically (Step 3). Chapter 9 provides examples.

Next, perform a sensitivity analysis for the extrapolation (Step 4). Calculation of confidence intervals (see Chapter 9) provides vital information on the range of future values to be expected. Consider also what outside influences are important—that is, which ones could alter the trend substantially and that are reasonably likely to occur. This sensitivity analysis may be quantitative (see Appendix C) and/or qualitative. Cross-impact methods (Chapter 12) can be of use, and expert opinion (Chapter 11) can be quite helpful at this juncture. A strategy that often proves highly effective includes these steps:

- Show your trend extrapolation to selected experts for their reactions,
- Raise specific questions about the external influences identified to see if the experts agree that they are likely and how they would alter the trend,
- Ask the experts to identify other factors likely to alter the trend.

After the sensitivity analysis has been performed, the projections should be interpreted. Forecasters often consider their job to be done when they provide trend(s),

but this is not so. The implicit knowledge that has been gained in determining that trend should be made explicit. The forecaster needs to indicate

- Why a certain model has been selected
- How strong or weak the data are
- What factors are likely to interact with the trend and how probable they are
- How much confidence he or she has in the trend

Wherever possible, an open dialogue with the intended users will add significant value to the trend analysis.

10.3 THE S-SHAPED CURVES

The familiar S-shape (more formally, sigmoidal) curves describe many natural phenomena and also fit the technological growth processes. This could be stated as the "pumpkin theory" of technological innovation (see Figure 10.4), the yeast theory, or the embryo theory. These natural growth processes share the properties of relatively slow early change, followed by steep growth, then a turnover as size asymptotically approaches the limit. Various mathematical formulations can depict such patterns. Two such models are presented here.

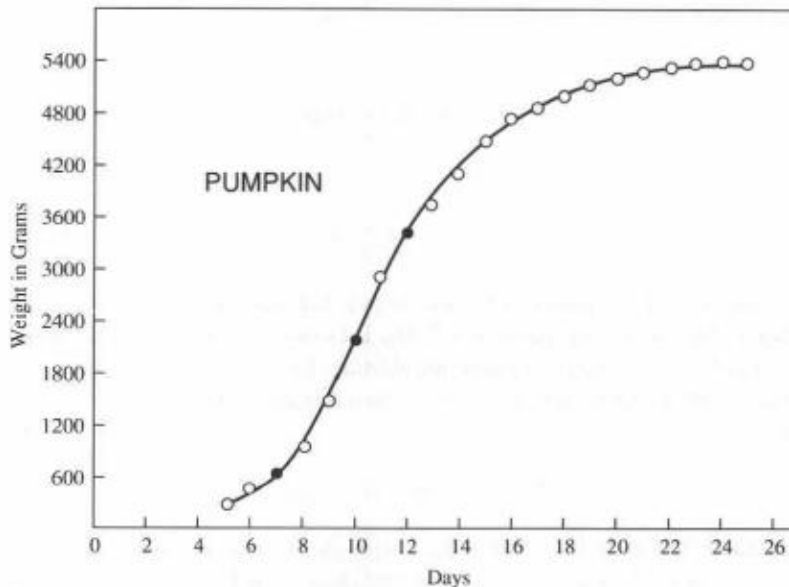


Figure 10.4 The pumpkin curve. (Source: Martino, 1972. Reprinted by permission of the publisher from *Technological Forecasting for Decision Making* by J. P. Martino, p. 184. Copyright 1972 by Elsevier Science Publishing Co., Inc.)

10.3.1 The Fisher-Pry Model

The Fisher-Pry model predicts characteristics loosely analogous to those of biological system growth. It also is referred to as a substitution model because it is often applied to forecast the rate at which one technology will replace (or be substituted for) another.

It is convenient to take the fraction of potential market penetration, f , as the dependent variable where $f = Y/L$ and L is the upper bound for growth of the variable Y (such as the number of units sold). The fundamental assumption of this approach is that the rate of change of f is proportional to both f and to the fraction of the market remaining, $1 - f$.¹ That is,

$$\frac{df}{dt} = b[f/(1 - f)] \quad (10.1)$$

This is a good assumption when initial sales make the sale of subsequent units easier. For example, when sales lead to growth in distribution, service, and/or repair networks that, in turn, encourages additional buyers.

Integrating Equation 10.1 between t_0 , the time at which substitution is 50% complete, and t is displayed as

$$\int_{t_0}^t b[f/(1 - f)] dt$$

Solving for f yields

$$f = \frac{1}{1 + \exp[-b(t - t_0)]}$$

or

$$f = \frac{1}{1 + c \exp(-bt)} \quad (10.2)$$

The expression given by Equation 10.2 was suggested years ago and was utilized in demographic forecasting by Raymond Pearl. It is often referred to as the *Pearl curve*, particularly when used to examine the absolute growth of a technology rather than the fraction of a market taken over by a technology. The form developed by Fisher and Pry,

$$f = 0.5 [1 + \tanh a(t - t_0)]$$

can be reduced to Equation 10.2 with a bit of straightforward manipulation. However, since the name "Fisher-Pry curve" is commonly used for this substitution function, we will refer to all forms of the equation as such here.

¹Such models are also known as logistic curves.

Fisher and Pry actually use a special version of the more general Pearl equation; the linear form given by

$$\ln[f/(1-f)] = 2\alpha(t-t_0) \quad (10.3)$$

where f = the fraction of that year's market held by the new technology; $1-f$ = the fraction held by the old technology; α = half the annual fractional growth in the early years; and t_0 = the time at which $f = 1/2$. Note the different interpretation of f in this case of competition from that in which f is the fraction of the ultimate market limit, L . In sum, this version offers an alternative formulation for an S-shaped curve with the same functional form.

In their original article, John Fisher and Robert Pry (1971) presented 17 cases of substitution that fit their model. Subsequent investigators have expanded the catalog so that nearly 100 cases of technology substitution have been found to follow the Fisher-Pry forecast (Lenz, 1985). Figure 10.5 illustrates the curves for various products and processes. These curves follow Equation 10.3. They appear

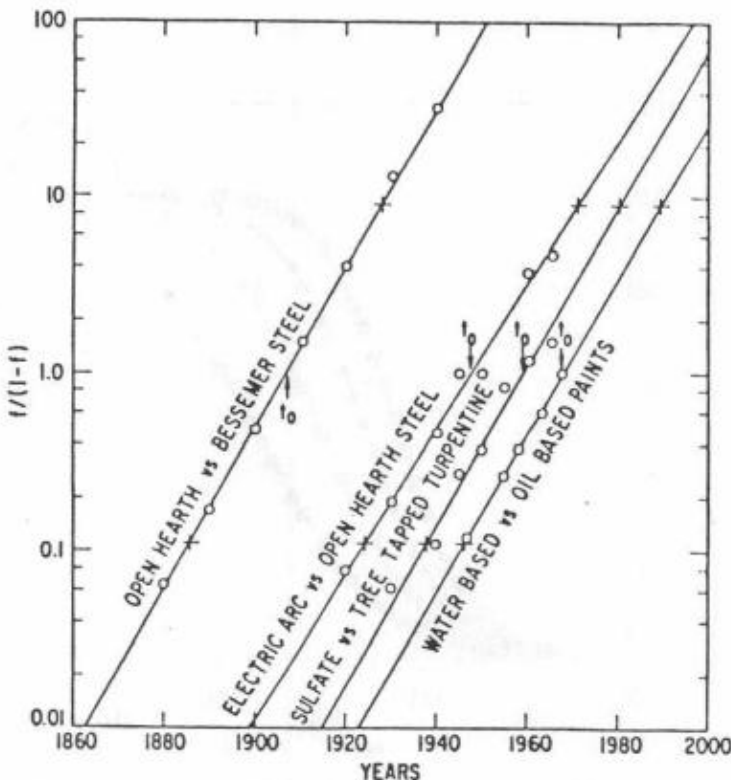


Figure 10.5 Fisher-Pry substitution curve fits for various products and processes. (Source: Fisher and Pry, 1971. Reprinted by permission of the publisher from "A Simple Substitution Model of Technological Change" by Fisher and Pry, *Technological Forecasting and Social Change*, pp. 75-88. Copyright 1971 by Elsevier Science Publishing Co., Inc.)

linear in the plot because the transformation $\ln [f/(1-f)]$ is a linear function of time for the model. The issue of modeling change in a technology over time, in conjunction with competition from other technologies, will be addressed further in Section 10.4.

The Fisher-Pry curves described by Equation 10.2 (equivalently, Equation 10.4) or by Equation 10.3 are both S-shaped, although not identical. The remainder of this section concentrates on the more general form—described by Equation 10.2 (or Equation 10.4).

From Equation 10.2, there are inflection points at $f = 0$, $f = 0.5$, and $f = 1$, which occur at times $t = -\infty$, $\ln(c)/b$ and $+\infty$, respectively. The first corresponds to the minimum value of f , the last to the maximum, while $[\ln(c)/b, 0.5]$ is the point about which the curve is symmetrical. As demonstrated in Figure 10.6A, changes in the value of b alter the shape of the curve; Figure 10.6B illustrates how

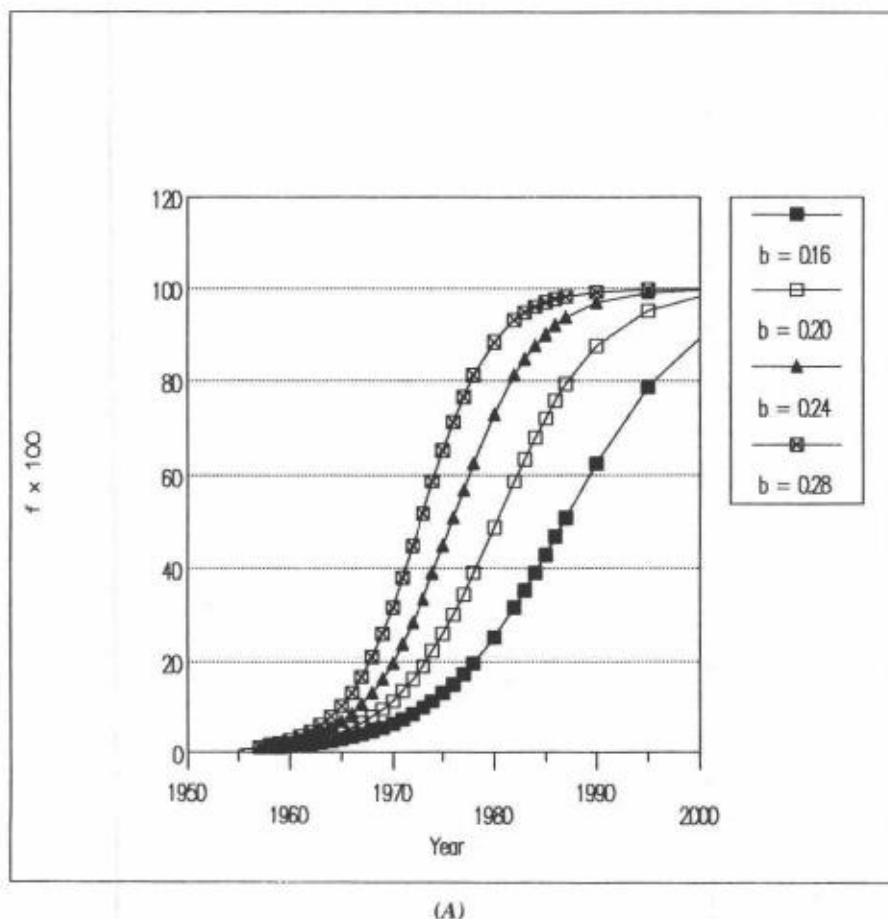
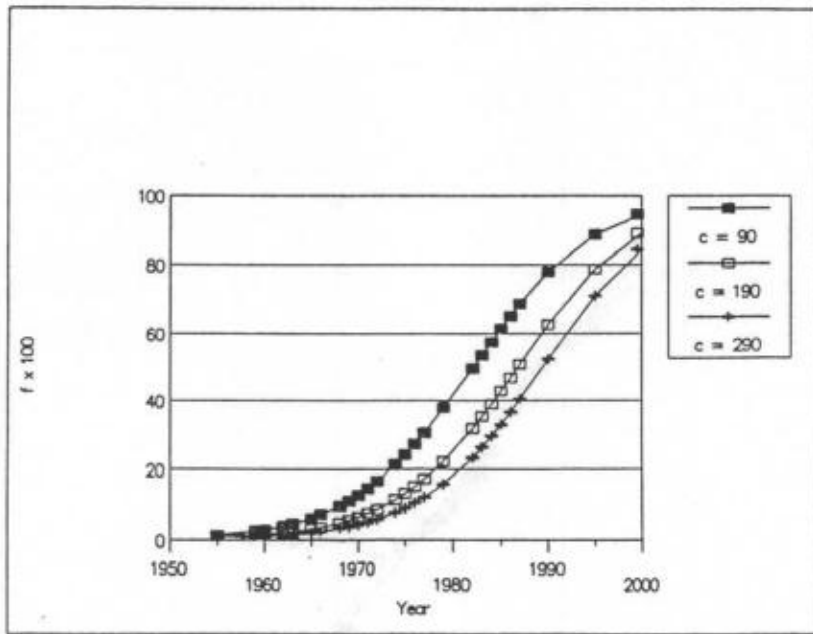


Figure 10.6 Fisher-Pry model. (A) Variation with values of b . (B) Variation with values of c .



(B)

Figure 10.6 (Continued).

changes in the value of c alter the location of the curve. Thus the model offers considerable flexibility to fit time series data.

To create a Fisher-Pry forecast, the analyst must determine the values of b and c that produce the best fit to the time series data. Although this can be done using the form given by Equation 10.2, it is much easier to transform the equation as follows:

$$Z = \ln[(1 - f)/f] = \ln(c) - bt \quad (10.4)$$

In this form, Z is a linear function of time. Existing time series data are used to compute values of Z . These are plotted versus time, and a straight line is fitted to the data to determine the values of b and c . Alternately, values of $(1 - f)/f$ can be plotted versus time using semilog graph paper. Five-cycle paper will allow the forecaster to cover a range from 0.001 to 100, which will accommodate values of f from 1 percent to 99 percent. The forecaster can fit the resulting plot with a straight edge or by using the least squares straight-line expressions presented in Chapter 9. Figure 10.7 shows the results of a least squares fit used to determine b and c for CATV data presented in the first columns of Table 10.1.

The accuracy of the growth portrayed by the resulting expression, of course, depends on the validity of the fundamental assumption about the invariance of the external environment. This assumption is inherent in all direct trend extrapolation

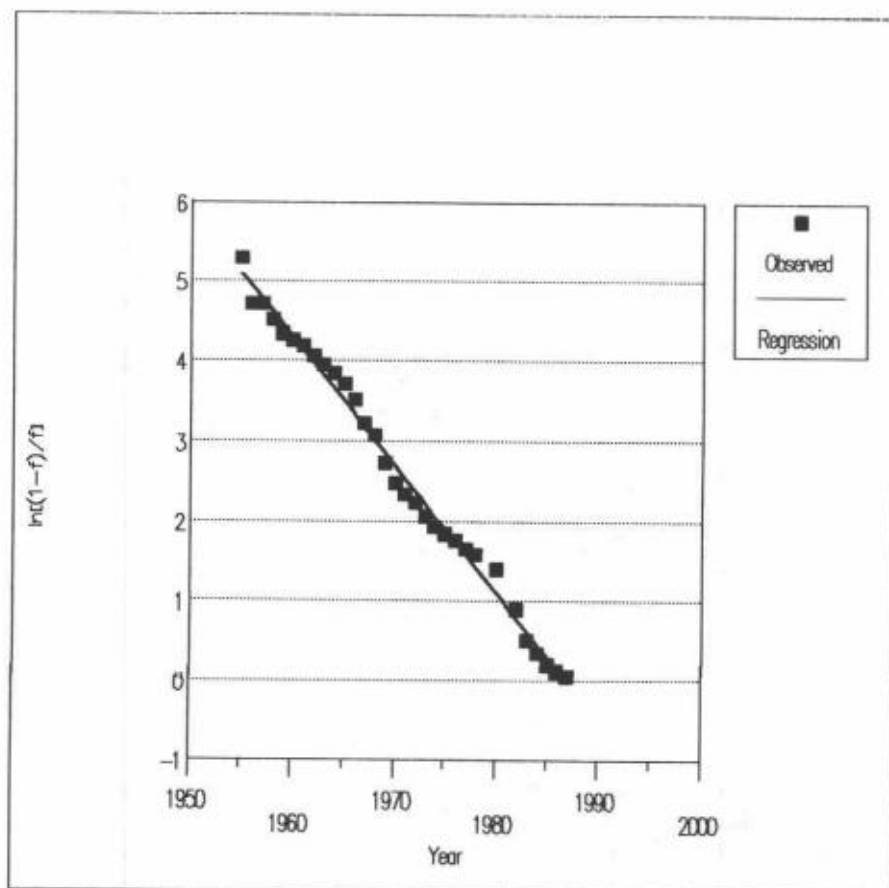


Figure 10.7 Fisher-Pry model: Least squares regression fit.

methods. Validity also depends on the extent and accuracy of the data the forecaster has in hand when b and c are determined. Figure 10.8 illustrates this by showing the different forecasts that would have been prepared for CATV growth depending on the number of years of available data. This variation in forecasts has led some to suggest that the Fisher-Pry model should be employed only after substitution has reached a level of 10 percent (that is, $f = 0.1$). For the CATV example, that point would be reached in 1972 (point number $n = 18$).

Figure 10.8 shows that waiting until substitution is substantial does indeed produce a more accurate forecast of CATV growth. Unfortunately, the technology manager frequently must make critical decisions at a much earlier time to capitalize on technological or market advantages (Lenz, 1985). The only answer to this dilemma is that forecasts should be made early and often—that is, the forecast should be made as soon as data are judged to indicate a trend and revised continually as new information becomes available.

TABLE 10.1 Fisher-Pry Forecast of CATV Penetration Including 90 Percent Confidence Limits

<i>n</i>	Year	Actual Forecast		90% C.I.	
		%	%	Lower	Upper
1	1955	0.5	0.61	0.47	0.79
2	1956	0.9	0.72	0.56	0.93
3	1957	0.9	0.84	0.65	1.09
4	1958	1.1	0.99	0.76	1.27
5	1959	1.3	1.15	0.90	1.48
6	1960	1.4	1.35	1.05	1.74
7	1961	1.5	1.58	1.23	2.03
8	1962	1.7	1.85	1.44	2.37
9	1963	1.9	2.16	1.69	2.76
10	1964	2.1	2.53	1.98	3.23
11	1965	2.4	2.95	2.31	3.76
12	1966	2.9	3.44	2.70	4.38
13	1967	3.8	4.02	3.16	5.10
14	1968	4.4	4.68	3.68	5.92
15	1969	6.1	5.44	4.29	6.88
16	1970	7.7	6.32	5.00	7.97
17	1971	8.8	7.34	5.81	9.22
18	1972	9.7	8.50	6.75	10.65
19	1973	11.3	9.82	7.82	12.27
20	1974	12.7	11.33	9.05	14.09
21	1975	13.8	13.03	10.45	16.14
22	1976	14.8	14.95	12.03	18.42
23	1977	16.1	17.09	13.82	20.95
24	1978	17.1	19.47	15.82	23.72
25	1979	NA	22.09	18.04	26.76
26	1980	19.8	24.96	20.49	30.03
27	1981	NA	28.06	23.19	33.51
28	1982	29.0	31.39	26.13	37.18
29	1983	37.2	34.93	29.30	41.00
30	1984	41.2	38.63	32.69	44.94
31	1985	44.6	42.47	36.26	48.94
32	1986	46.8	46.41	40.05	52.89
33	1987	48.7	50.39	43.90	56.87
36	1990		62.11	55.69	68.12
41	1995		78.44	73.45	82.72
46	2000		88.98	85.87	91.48

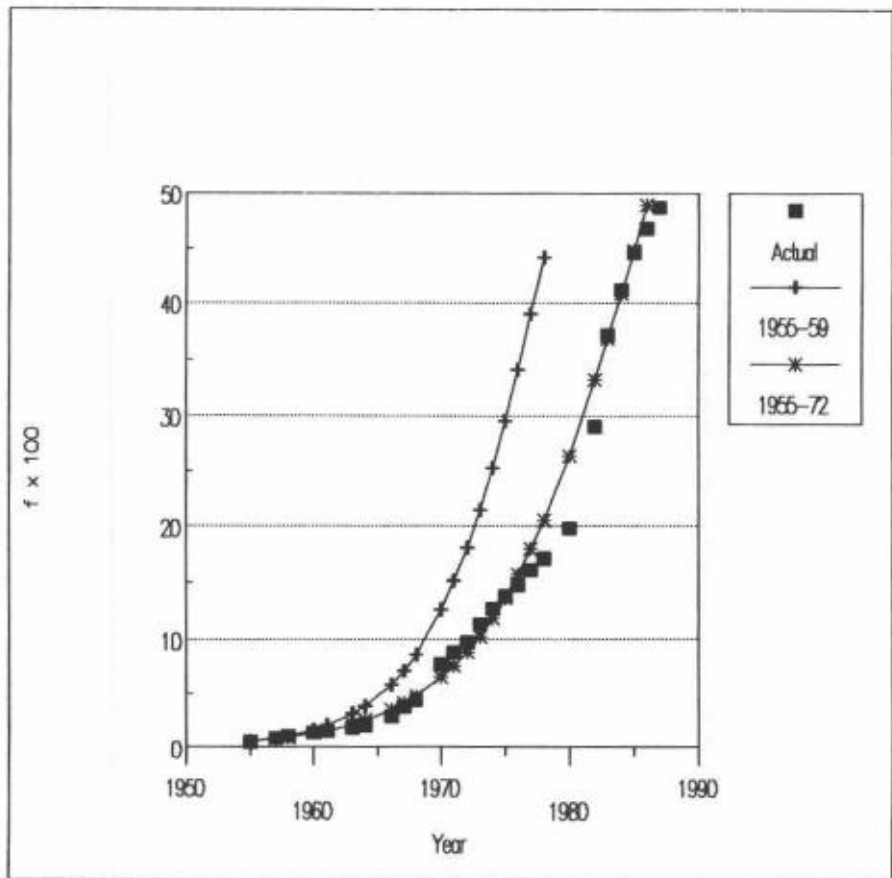


Figure 10.8 Fisher-Pry model: Variation with number of years of available data.

Figure 10.9 compares the Fisher-Pry model using coefficients from a least squares fit of the 31 CATV data points (see Table 10.1) to measured values of f . Also shown in the figure are the 90 percent confidence interval values for the forecast. These confidence limits were computed by the methods presented in Chapter 9 applied to the forecast in the form suggested by Equation 10.4. Values of the 90 percent confidence limits of the forecast transformed into the f versus time domain are given in Table 10.1. Note that the confidence intervals are not perfectly symmetrical about any point (reflecting the nonlinear Fisher-Pry equation). Notice also that, unlike the case for unbounded linear extrapolation, the confidence intervals do not expand in a cone shape as the extrapolation period lengthens. Instead, the confidence intervals, too, are constrained as the curve approaches the limit.

10.3.2 The Gompertz Model

The underlying dynamics of the Gompertz model are quite different from those of the Fisher-Pry model. In fact, Gompertz is sometimes referred to as a mortality

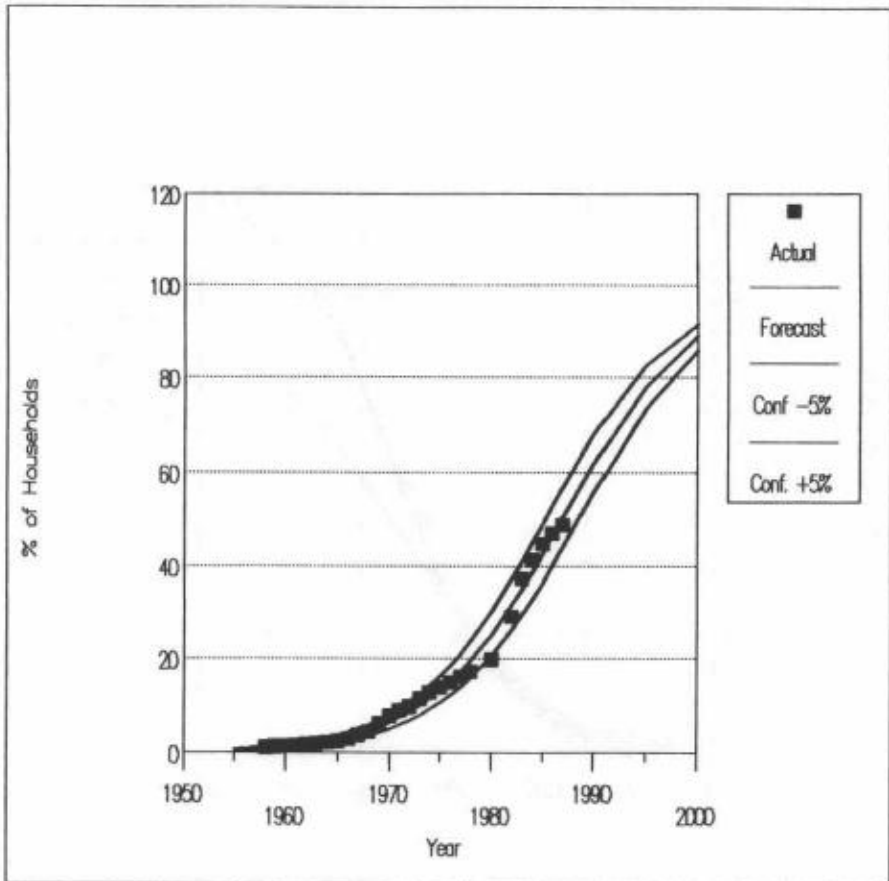


Figure 10.9 CATV subscribers in the United States: Fisher-Pry fit, 90 percent confidence intervals, and actual data.

model, whereas Fisher-Pry is characterized as a growth model. Benjamin Gompertz, an English demographer, found that the mortality rate of a population grew exponentially as it aged. With respect to technology forecasting, the Gompertz model is most appropriate in cases in which equipment replacement is driven by equipment deterioration rather than by technological innovation.

The Gompertz model is defined by this equation:

$$f = \exp[-b \exp(-kt)] \quad (10.5)$$

Once again, $f = Y/L$, and L is the upper limit for the number of units, Y . While equation 10.5 produces an S-shaped variation, as does the Fisher-Pry model, Figure 10.10 shows that the shapes predicted by the two models for a given set of data are quite different. As with the Fisher-Pry expression, Equation 10.5 has three inflection points: $f = 0$; $f = 0.368$, and $f = 1$. These correspond to $t =$

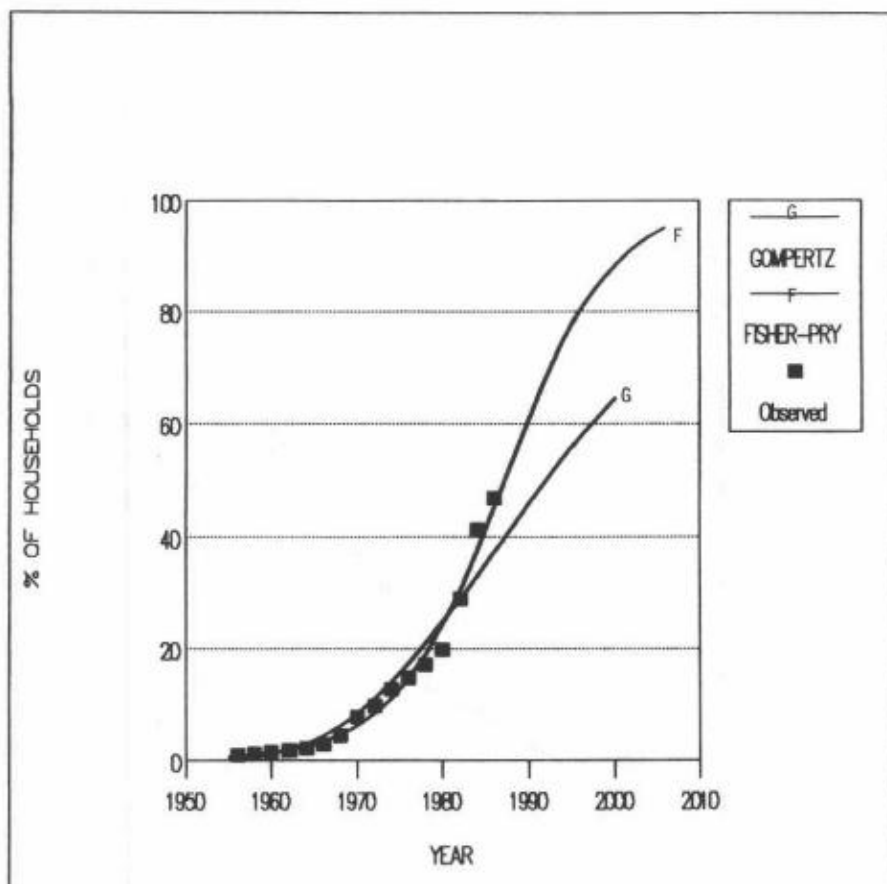


Figure 10.10 CATV subscribers in the United States: Fisher-Pry versus Gompertz models.

$-\infty$, $\pm \ln(b)/k$, and $\pm = +\infty$. However, the Gompertz curve is not symmetrical about its central inflection point.

The position and shape of the Gompertz curve can be adjusted independently—the position by changing b ; the shape by altering k . Thus, as with the Fisher-Pry model, it offers considerable flexibility in obtaining a satisfactory fit of time series data. The process of fitting the Gompertz curve is analogous to fitting the Fisher-Pry curve. However, the log of Equation 10.5 must be taken twice to produce this linear expression:

$$Z = \ln[\ln(L/f)] = \ln(b) - kt \quad (10.6)$$

The forecaster can find values of b and k by fitting a straight line to values of Z computed from the data by using a straight edge on a plot of Z or by least squares regression fit.

10.3.3 Choosing Fisher-Pry or Gompertz

The model chosen by the forecaster must embody underlying characteristics that reflect those of the process to be forecast. This is not always easy: the process often is not well understood. However, the Fisher-Pry and Gompertz models offer an opportunity to illustrate model selection. The following discussion reflects the general concerns of model selection and of differentiation between growth and mortality models.

Equation 10.1 shows that Fisher-Pry, a growth model, assumes that the rate of change of the process is proportional to *both* the fraction of the market penetrated by the technology and the fraction that remains to be penetrated. It depends on the number of uses for which the new technology has been applied and on the number for which it is yet to be applied. This is analogous to the process followed by the diffusion of a new, technically advanced product. In such cases, knowing someone who owns one helps prospective buyers assess the technology's potential for them. Since the availability of the technology and of spares, repair facilities, and advice normally grows with the number of units in the field, diffusion is further enhanced by the sale of additional units. Thus penetration of the market is driven, not only by potential sales, but also by the sales that have been achieved.

In the diffusion process modeled by Fisher-Pry, initial sales of a new technology are difficult despite its promise and the size of the potential market. This is true largely because the technology is unknown and unproven. Moreover, in the early stages, field support is likely to be poor, there may be institutional barriers to overcome (such as, licensing), and survival of the technology and its supplier is uncertain. Thus adoption implies considerable risk. As applications grow, so does general knowledge of the technology and its advantages. The support infrastructure improves, as does confidence that the purchaser will not suffer either from inadequate support or from the financial failure of the manufacturer. From this point, penetration grows rapidly. At some stage (for example, after 50 percent penetration), further penetration becomes increasingly more difficult because it often involves sales to companies that may not benefit as greatly from the technology or that have marginal capability to finance adoption. Thus the rate of penetration slows.

The Gompertz model embodies quite different dynamics. For example, although not readily apparent from Equation 10.5, for penetrations greater than 50 percent, the rate of penetration depends primarily on the fraction of the market remaining. Thus the Gompertz model is appropriate for forecasting market penetration by technologies for which initial sales do not make subsequent sales easier. This dynamic usually is found when a new technology offers no clear-cut advantages over an old. In such instances, an older technology is replaced by a newer technology that performs the same tasks with essentially the same financial and/or functional efficiency. Purchases simply replace equipment that has worn out or has been destroyed. It is this characteristic of the Gompertz model that leads to its classification as a mortality model.

Since the dynamics of the two models are different, it is reasonable to expect significant differences in the forecasts they produce. For example, the Fisher-Pry

model forecasts a more rapid penetration than does the Gompertz. However, either model can be made to fit a handful of initial data points. Thus goodness of fit cannot be used as a criterion to determine which model to use; instead, the dynamics of the model must be matched to the dynamics of the process being forecast.

The process of CATV growth considered in this chapter is an example of new technology diffusion. CATV offers users significant benefits over older telecast delivery technologies. In this case, older technology is not merely replaced when it wears out. In fact, existing TV sets are converted to CATV, and the new and old technologies often function side by side in the same neighborhood. In the initial stages, diffusion of CATV was slowed by uncertainties about the potential of the technology and the future of the cable companies themselves. These uncertainties were mitigated, however, because conversion to CATV did not prevent TV sets from functioning in their original mode. Initial diffusion was more seriously inhibited by the need to build a cable infrastructure and by licensing requirements that usually must be satisfied location by location. In short, the dynamics of CATV growth provide a compelling case for selecting the Fisher-Pry model. Data show that diffusion has now proceeded to nearly the 50 percent level. The forecast (see Figure 10.9 and Table 10.1) suggests that it will proceed to 90 percent completion in the next decade.

Not all substitution processes are as clear-cut as the example of CATV. Some, for example, appear to be driven by a mix of new technology diffusion and old technology mortality. For example, some owners of prestige automobiles may replace them with newer models that offer significant technological improvements (such as anti-lock braking systems or fuel injection). Others may replace to "keep up with the Joneses" even when no clear technological superiority is offered. Still others may be driven (no pun intended) to replace worn-out units by nearly identical ones because of the prestige and tradition of an established design ("only the very best people own a . . ."). In cases in which the underlying dynamics of the penetration are unclear, the best guide is to consult those whose job it is to understand the dynamics of the market. The overly optimistic forecasts of demand for innovative technology products, such as home computers and expert systems, for example, can be partly attributed to the failure of forecasters to consult such experts (Wheeler and Shelley, 1987). However, Schnaars (1989) notes that such experts often get caught up in the "zeitgeist" (prevailing wisdom) of the times and tend to be overly optimistic. Historical analogy to similar situations also can be helpful.

The forecaster must match the dynamics of the model to those of the process. To reiterate, the dynamics of the Fisher-Pry model are appropriate to cases of technology diffusion; the dynamics of the Gompertz model are appropriate to cases of replacement driven by equipment deterioration rather than technological advantages. Occasionally, as Lenz (1985) notes, industries have been able to match equipment deterioration to new technology innovation; however, the difference in the rates associated with each mitigate against achieving that balance often. When the dynamics of the process are unclear, consult the experts, compare alternative models, and pray.

10.3.4 Selecting an Upper Bound for the Forecast

The preceding discussions, have emphasized forecasting the market fraction, $f = Y/L$, where L is the upper bound for Y . This approach simplifies the presentation; however, it also disguises the fact that L is a third parameter that must be estimated to employ forecasts made with the Fisher-Pry and Gompertz models. An accurate estimate of L is important for a number of reasons. First, L must be known to formulate the time series data for forecasting. Second, an incorrect upper bound can seriously distort values of the fitting coefficients (b and c in the Fisher-Pry model) and hence the subsequent forecast. Further, the technology manager generally needs a forecast, not only of the fraction of the market that will be penetrated, but also of the number of units that will be sold.

In the case of CATV, the upper limit was set as the number of households with TV sets—a number that grows with time and, therefore, is itself a subject for forecast (as considered in Chapter 9). The range of f for CATV could be established as zero to one with some confidence. In other instances, this cannot be done so easily (for example, in forecasting the upper limit for the sales of subcompact automobiles, the total number of automobiles sold would not be an accurate upper bound). Nor is the limit for the functional capacity of a technology (for example, the precision of a manufacturing method or the level of concentration of a chemical compound that can be detected) so easily set.

In some instances, forecasters have used the data to establish the upper limit (for example, in the Fisher-Pry method, data fits would be performed for b , c , and L). As Martino (1983) notes, this is bad practice because initial data are relatively insensitive to the upper bound and thus do not provide reliable guidance. Rather, L should be set by natural or fundamental limits to the process. This, of course, requires knowledge of the technology and the market place that the forecaster may not possess. Therefore, in cases where the upper bound is unclear, the forecaster should work closely with experts in the field to determine a reliable estimate.

10.4 TOWARD A GENERAL MODEL OF TECHNOLOGICAL CHANGE: THE LOTKA-VOLTERRA EQUATIONS²

10.4.1 Theory

The Italian mathematician Vito Volterra proposed the Lotka-Volterra equations at the turn of the century to model population changes of sharks and food fish in the Adriatic Sea. Since then, the model has been expanded and applied in the fields of demography and ecology. It successfully represents many biological populations based upon data from field observations in the wild. Although simple, the model is one of the theoretical cornerstones of the discipline of mathematical ecology.

As per earlier discussion, technology forecasting has frequently and successfully borrowed from demographic models for the purpose of trend extrapolation.

²This section was contributed primarily by Scott Cunningham at Georgia Tech. The authors greatly appreciate his contribution.

Both the Pearl (Fisher-Pry for a single technology) and the Gompertz curves were originally developed in demography and later applied to technology. In using these models, the technology forecaster draws upon an analogy between populations of technologies and populations of organisms. Demographic models often produce highly accurate forecasts for technological growth.

The Lotka-Volterra equations have helped to frame a number of long-range technology forecasts. The accuracy of such extrapolations and the strong biological bases for the Lotka-Volterra model suggest that these mathematical formulations may relate to the underlying sociotechnical processes that lead to the growth and diffusion of technologies.

A system of Lotka-Volterra equations (describing interactions between two technologies) is shown below:³

$$\frac{dX}{dt} = X(\alpha_1 - \beta_1 X - \gamma_1 Y) \quad (10.7)$$

$$\frac{dY}{dt} = Y(\alpha_2 - \beta_2 Y - \gamma_2 X)$$

More elaborate systems may be created by adding additional variables (technologies) and by introducing new competition terms.

Consider the basic dynamic assumptions made by this model. First and foremost, the Lotka-Volterra model assumes that technology develops within a larger market system in which technologies may compete; the model expressly includes variables to represent these features. Second, such systems tend to be open to many forces not included in the model (such as governmental regulations), but the Lotka-Volterra model assumes that Equation 10.7 captures the essential dynamics.

Specifically, the Lotka-Volterra model assumes that technological growth depends on two factors: existing technology populations and market potential. The resultant growth depends on multiplication of these two factors, as shown in Table 10.2.

Consider next the purpose and meaning of the coefficients. Note in the table that α , β , and γ all occur within the market potential portion of the equations. Let X and Y constitute two competing technologies (Equation 10.7), measured either by units sold or by profits earned. α pertains to the production or sales capacity of the market and has units of 1/time. α_1 and α_2 will be equal when both technologies have equal market capacity. β is the inhibition coefficient. It gauges the loss of potential market caused by increases in the technology. When forecasting sales it has units of 1/[money $\times$ time], and when forecasting units of technology it has units of 1/[unit $\times$ time]. γ is the competition coefficient. It reflects the share of market

³This book refers to Equation 10.7 as the Lotka-Volterra equations—plural because more than a single technology can be modeled. The original Lotka-Volterra equation lacked the second term, $-bX^2$ (when multiplied out), so that in the absence of competition, growth would be exponential. Equation 10.7 is known as the predator-prey version of the equations; it provides a quite general depiction of technological growth.

TABLE 10.2 Dynamic Origins of the Lotka-Volterra Equation

Growth in technology is the result of current levels of technology times current market potential:

$$\frac{dX}{dt} = X \cdot L(X) \quad \text{where}$$

Symbol	Description	Used in system as:
$\frac{dX}{dt}$	Growth in technology	$\frac{dX}{dt}$, $\frac{dY}{dt}$, etc.
X	Existing levels of technology	X , Y , etc.
$L(X)$	Market potential	$(\alpha_1 - \beta_1 X - \gamma_1 Y)$, $(\alpha_2 - \beta_2 Y - \gamma_2 X)$, etc.

capacity taken by the competing technologies and has the same units as β . β_1 and β_2 need not be equal, nor must γ_1 equal γ_2 . These three sets of coefficients are necessary for a complete solution to the Lotka-Volterra equations and will vary according to the characteristics of the technologies in question. The value of the coefficients will also vary depending upon the units chosen for study (such as years and thousands of items shipped per year.)

The form of Equation 10.7 may be unfamiliar to forecasters who usually address changes in amounts of technology over time (e.g., annual sales) rather than changes in degree of technological growth over time (e.g., rate of sales gains). Most trend extrapolation models employ a single mathematical equation that describes amount of technology over all time periods. These equations are known as *closed form solutions* by mathematicians. Equation 10.7, a system of differential equations, describes changes in growth. Although it will give a definite value of growth for any instant at which all the conditions of the equation are known, it is impossible to solve these equations in a single closed form. The growth formulation of the Lotka-Volterra equations is sufficiently general to embrace a very large family of growth curves. The Lotka-Volterra equations can approximate each of the closed form forecasting models discussed earlier in this chapter.

Given the dynamic form of the Lotka-Volterra equations and the general description of its variables, it is now appropriate to discuss several relevant concepts borrowed from biology. The first is the definition of an ecology, defined here in a technological context as a "system." The properties of a system, for the purposes of the Lotka-Volterra equations, are as follows:

- A system is limited to a few closely interrelated technologies and markets.
- Each technology and market within the system is at the same level of generality.
- Systems are hierarchial and may be refined at progressively higher or lower levels of detail.

- The system is abstracted so as to be relatively closed (that is, relatively immune to external influences).
- Systems are simplified models of more complex sociotechnical environments—models that may be manipulated mathematically.

Another important concept in population biology is that of the niche. For technology growth modeling, the niche can be considered to be the prospective market for which the technology is suited. Considering this definition, technological competition is the process by which two technologies vie for a niche, or more broadly, the ways in which two technologies interact with each other within a sociotechnical system.

Ecologists speak of the concept of *biopotential*. Species vie for biopotential; technologies compete for market potential. The actual potential of the market might consist of human, economic, and/or technological resources. However, for the purposes of the Lotka-Volterra equations, this market potential is given in terms of the potential production or sales capacity for the technology. Potential is shown in Table 10.2 as $L(X)$.

There are two mutually exclusive types of biological systems that are of particular interest for technology forecasting: competitive and inhibitive systems. Inhibitive systems are those in which market size primarily limits growth. They are equivalent to the growth models discussed earlier in the chapter. In contrast, competitive systems are primarily limited in growth by the nature and extent of competitive behaviors with other technologies. In competitive systems, one technology steadily drives the other into extinction or disuse. Competitive systems relate closely to the mortality models discussed earlier in this chapter. Competitive systems occur when $\beta_1 \times \beta_2 < \gamma_1 \times \gamma_2$. Inhibitive systems predominate when $\beta_1 \times \beta_2 > \gamma_1 \times \gamma_2$.

10.4.2 Functional Equivalences of the Model

Under what conditions do the Lotka-Volterra equations produce the special case trend extrapolation models discussed earlier in the chapter? Although the equations may not be reduced to a closed form solution for the vast majority of potential systems, they can produce solutions that duplicate any of the standard forecasting curves to any arbitrarily high level of precision. The Lotka-Volterra equations are of interest because they can depict such a wide range of dynamic behaviors. They are sufficiently general to encompass many different kinds of trend extrapolation models. Linear, exponential, Pearl, Gompertz, substitution models, and oscillatory behaviors can all be matched by the Lotka-Volterra model. Table 10.3 lists specific conditions under which these solutions will be evidenced by the Lotka-Volterra equations.

Linear behaviors will be evidenced by slow-growing technologies or for any technology examined over a sufficiently short period of time. If examined for greater lengths of time, exponential or S-curve characteristics will likely appear.

TABLE 10.3 Summary of Relationships between Lotka-Volterra Equation and Other Forecasting Models

Trend Model	Closed-form Solution	Growth Description	The Lotka-Volterra Equations	
			$\begin{bmatrix} \bar{X} & X(\alpha_1 - \beta_1 X - \gamma_1 Y) \\ \bar{Y} & Y(\alpha_2 - \beta_2 X - \gamma_2 Y) \end{bmatrix}$	Approximate this trend Model when ...
Linear	$X = At + B$	$\frac{dx}{dt} = C$		Slow growth or short periods of time, competition, and/or inhibition are significant ($\gamma, B \rightarrow \infty$)
Exponential	$X = \exp(At)$	$\frac{dx}{dt} = AX$		Occurs for low saturation ($\alpha > X$)
Decaying exponential	$X = A - A \exp(-Bt)$	$\frac{dx}{dt} = (A - CX)$		Occurs for high saturation ($X \rightarrow \alpha$)
Pearl	$X = \frac{A}{1 + B \exp(-Ct)}$	$\frac{dx}{dt} = X(-A - CX)$		Usually inhibitive systems, but can model competition for technologies that have similar growth rates ($\gamma \rightarrow 0$ or $\alpha_1 \sim \alpha_2$)
Gompertz	$X = A \exp(-C \exp(-Dt))$	$\frac{dx}{dt} = CX e^{-At}$		Competitive systems; models older technology growing obsolescent ($\beta \rightarrow 0$)
Substitution models	No general closed-form solution	Various		Significant competition between growth and equilibrium technologies or between two technologies with similar growth rates

The Lotka-Volterra equations approximate such behaviors when the coefficients β and/or γ are very large.

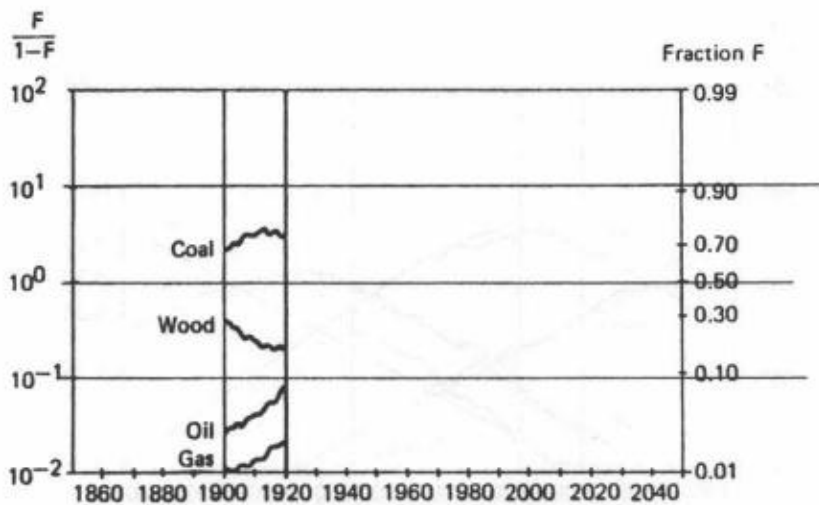
Exponential behaviors often occur at the beginning of growth regimes. Many exponential growth patterns will in fact begin to level out and produce curves more characteristic of the Pearl or Gompertz models, when observed for a sufficient length of time. The Lotka-Volterra equations will produce solutions that approximate the exponential models when the market is not saturated. It should be noted again that the Lotka-Volterra equations will not produce a closed-form exponential curve. They will, however, produce a curve that is equivalent to the exponential model for any arbitrarily defined level of accuracy and any desired length of time.

The Lotka-Volterra equations reduce in closed form to the Pearl (or one-variable Fisher-Pry) model when there is no competing technology. The Pearl curve can otherwise be approximated by the Lotka-Volterra equations whenever competition is insignificant, such as when the competing technology Y is not near saturation, or whenever the competition coefficient γ is relatively small. Pearl curves are also produced in competitive situations between technologies that have similar growth rates. The Pearl curve is basically an inhibitive model of technology growth.

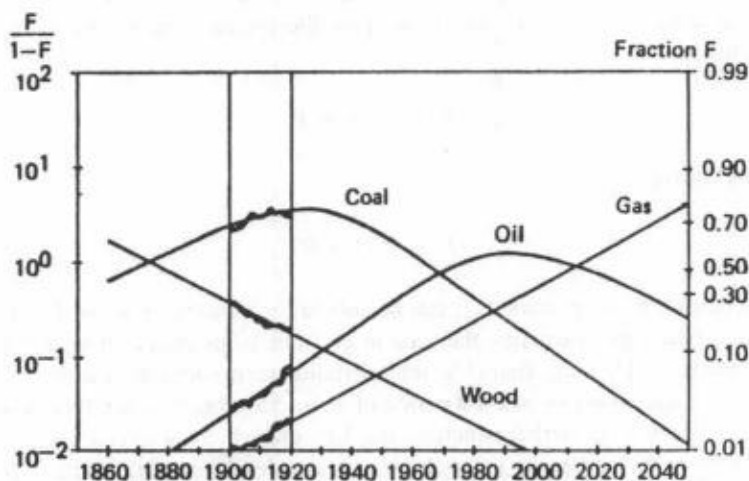
In contrast, the Gompertz curve occurs in situations where there is intense competition between technologies. The competition equation will produce behaviors that are functionally equivalent to the Gompertz model when there are competing technologies that take prospective markets away from the technology X of interest. The Gompertz equation, then, is chiefly a model useful for competitive systems.

Substitution is an important class of behavior often evidenced by growing technologies. The Fisher-Pry model represents substitution of a single, saturated market by a growth technology. Competition between more than two technologies, or competition between two growing technologies can be forecasted using Marchetti's multiple substitution analysis. Details of this analysis are available (Marchetti, 1983; 1988). Substitution analyses that rely on logistic growth and logistic decay as descriptors of technological takeover constitute a specific set of solutions to the Lotka-Volterra equations. The Lotka-Volterra equations will produce this kind of growth pattern whenever competitive systems occur that include technologies of similar growth and takeover rates.

Figure 10.11b shows the forecasting potential of multiple substitution analyses. Marchetti and Nakicenovic (1979) show that equations fit from 20 years of data for four competing energy technologies provide good fits with historical developments over the next 60 years. Marchetti (1988) attributes the relatively poorer fit for natural gas to the fact that its market penetration was still small (about 2 percent) at the end of the base 20-year period. Note that this model anticipates a window of opportunity for a new energy source beginning in the 1970s. Nuclear power seemed to be growing to fill this niche, but external influences may well limit it. Extension of these energy substitution models at the International Institute for Applied Systems Analysis (IIASA) suggests another substitution due to begin about 2025 (possibly solar or fusion energy?).



(A)



(B)

Figure 10.11 Lotka-Volterra energy projections. (A) 20-year data record for four comparing energy technologies. (B) Resulting projections. (C) Fit with actual data. (Source: Marchetti and Nakicenovic, 1979. Reprinted by permission.)

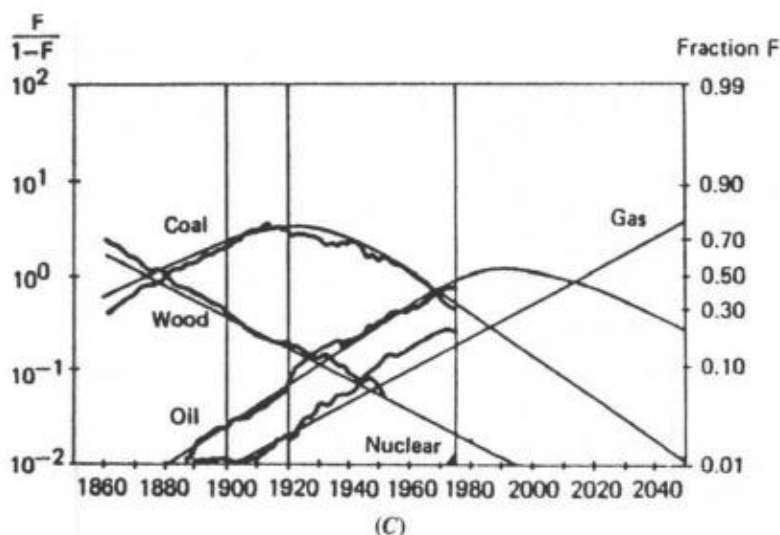


Figure 10.11 (Continued)

Note the left side of Figure 10.11 is scaled as $f/(1-f)$ in powers of 10 (a log scale), as used previously in the book. The alternative scale on the right results from setting

$$f/(1-f) = P$$

then solving for f

$$f = P/(1+P)$$

The fraction of the total market, f , can be solved for various values of P ; Marchetti emphasized the time constant—the time to go from 10 percent to 90 percent of the potential market. He notes that U.S. transportation technologies—canals, railroads, paved roads, and airways—show a period of about 55 years between their respective $f = 0.5$ (50 percent) market penetrations. This underlies his forecasts of windows of opportunity for successor technologies. It also lends support to the hypothesis of bursts of technological innovation in long waves (Volland, 1987).

Of his substitution analysis and its relationship to the Lotka-Volterra equations, Marchetti (1983, p. 3) comments:

I am fairly convinced that the equations Volterra developed for ecological systems are very good descriptors of human affairs. In a nutshell, I suppose that the social system can be reduced to structures that compete in a Darwinian way, their flow and ebb being described by the Volterra equations, the simplest solution of which is a logistic.

Marchetti's multiple substitution analysis is strongly suggestive of the importance and power of successive families of logistic curves. While Marchetti normally examines single families of competing technologies, he notes that such families often grow within a larger market whose growth also can be described as logistic. This conception of successive layers that describe hierarchies of logistic functions has interesting implications for the technology forecaster.

It can be illustrated that logistic growth that occurs within a larger logistic market will indeed produce another Pearl or Fisher-Pry curve. It is clear, however, that the forecaster makes choices about the market and level of detail he or she wishes to study. Figure 10.12 illustrates a hierarchy of transportation technologies. At each successive level logistic growth and competition functions may be successfully used to describe growth. In technology, as in nature, the exact taxonomy used for classification is not as important as representing the correct functional form. Substitution analysis, such as Marchetti's or Fisher-Pry, cannot be successfully used to forecast growth when addressing two technologies of radically different growth rates or from different market niches.

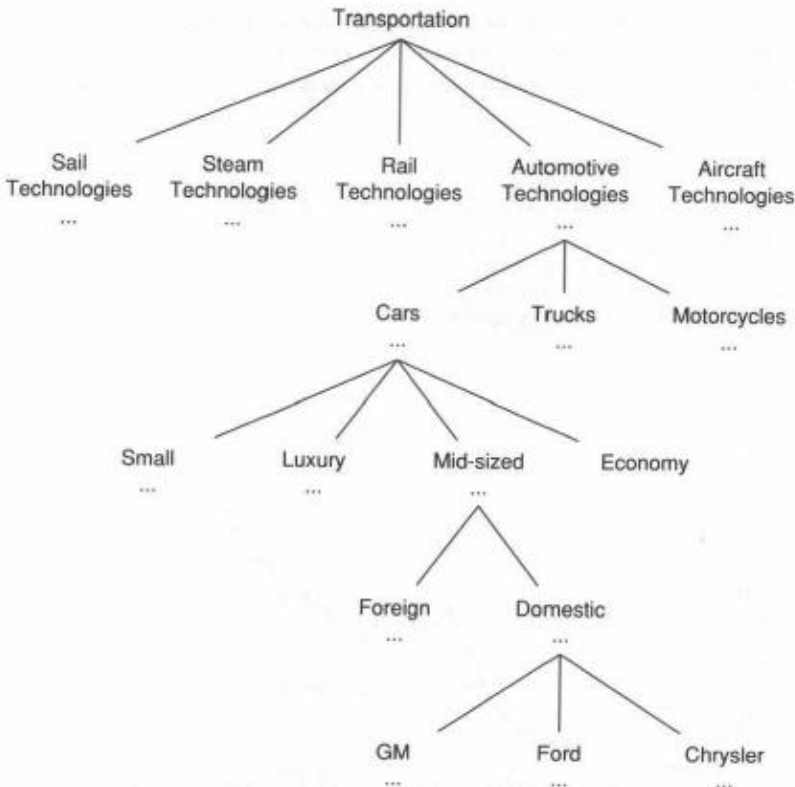


Figure 10.12 A hierarchy of transportation technologies.

Oscillatory models are a final class of models accommodated by the Lotka-Volterra equations. Periodic behaviors are commonly found in natural populations, and they can be successfully modeled using the Lotka-Volterra equations. Oscillatory behaviors have been observed in consumption and mining patterns in the United States and in car and transportation systems in Europe. It is suspected that the growth of certain technologies (such as robotics in the United States) follow an oscillatory rather than S-shaped growth pattern. Oscillatory behaviors present a special challenge to the forecaster using trend extrapolation methods. These growths often show a logistic start followed by an overshoot and then oscillation around a supposed limit. The unwary forecaster who naively applies a Pearl or Gompertz curve to start-up data, to data which are in fact oscillatory, will produce forecasts that are in serious error. The more complex population models, such as Lotka-Volterra, can represent such behaviors if the forecaster has correctly surmised their form.

Mathematical modeling using the Lotka-Volterra equations shows that oscillatory behaviors occur in certain specially defined systems that are known to biologists as predator-prey systems. In the equivalent technological systems, the predator technology benefits from interaction in a marketplace, while the prey technology loses from interaction. A gain from interaction is modeled with a negative competition coefficient (which subtracts from the available market). Figure 10.13 shows such oscillatory behavior seen in copper mining in the United States. Note the logisitic start-up in the early 1900s and the subsequent behavior not accounted for by Pearl or Gompertz models.



Figure 10.13 Oscillatory behavior in U.S. copper production.

10.4.3 Issues

This section has discussed the theoretical derivation of the Lotka-Volterra equations, their potential as an effective model of competing technologies, and the ways in which the model provides an overarching rationale for many conventional curve-fitting routines. Growth data, applied in the basic formulation of the Lotka-Volterra equations, are noisier than population data (for instance, it is easier to specify precisely the location of an automobile than its rate of change of location—velocity). Fitting change data to systems of equations can be done using nonlinear multiple regression, a procedure that grows more complicated as the number of competing technologies increases. Extensive data sets, often not available to the forecaster, are needed.

Until a technique is developed that allows a general solution to the Lotka-Volterra equations based upon time-series data, they cannot be used as a completely general trend extrapolation tool. Marchetti (1988), however, refers to excellent results in fitting some 100 empirical cases of technological growth using a numerical approximation program that constitutes a significant subset of potential solutions to the Lotka-Volterra equations (Marchetti and Nakicenovic, 1979).

The Lotka-Volterra equations offer a number of possible aids to the manager of technology:

1. An overarching model that is sufficiently general to describe most technological growth situations
2. A convincing rationale for trend extrapolation that provides clearly defined assumptions about the nature of technological growth (this rationale may be applied no matter what specific curve-fitting model is being used)
3. A rigorous and analytical treatment of competition within technological systems
4. A host of useful biological and population dynamics concepts and analogies that may be used when considering the interrelationships of technology and society
5. A solution set of curves that are richer than all current trend extrapolation techniques; such curves, once refined, might well prove as useful as current models have in forecasting technology growth (the Lotka-Volterra equations can model oscillatory as well as inhibitory systems not represented by current forecasting tools)
6. Links between trend extrapolation and simulation, modeling, and systems dynamics (the Lotka-Volterra equations offer an analytical technique by which alternative cases and scenarios about new and emerging technologies can be explored and refined).

10.5 CONCLUSION

Chapters 9 and 10 provide the essential tools for trend extrapolation. Chapter 9 presents the basics, including computational procedures to fit linear equations. By

using appropriate transformations of nonlinear functions, these procedures can be readily adapted to all the commonly used technology trend models.

Chapter 10 provides a four-step approach to analyze technological trends. It emphasizes the Fisher-Pry (Pearl) and Gompertz curves—the S-shaped curves—and introduces the Lotka-Volterra equations as a promising general model of technological growth and competition.

The validity of trend analysis depends critically on a number of assumptions. Extrapolation assumes that the underlying structures and interacting forces remain the same throughout the period of extrapolation. This assumption always needs to be carefully considered. Sometimes, it is determined that trend extrapolation would not be justified. Even when it is justified, trend analyses should include sensitivity analyses, seeking to identify factors likely to disrupt the trend and estimating how the trend would be altered. Furthermore, trend analyses need to be combined with the use of complementary methods, especially monitoring and expert opinion. Taken together, these can provide the bases for a sound forecast.

EXERCISES

- 10.1** Examine the data for 1978 to 1988 in Table 10.4 for tapes and CDs. Try to establish a leading indicator relationship between the two series (recall the presentation on leading indicators and regression in Chapter 9). Use the relationship to project CD shipments for 1990. Comment on your results.

TABLE 10.4 Shipments of Audio Media (millions of units)

Year	Long Play/ Extended Play Records (LPs)	Cassette Tapes (Tapes)	Compact Discs (CDs)
1978	340	65	0
1979	315	80	0
1980	320	110	0
1981	295	140	0
1982	245	185	0
1983	210	240	2
1984	205	335	10
1985	170	342	22
1986	130	345	53
1987	110	412	110
1988	70	452	145

Source: Recording Industry Association

- 10.2** Develop a Fisher-Pry substitution model for CDs. Figure f as the share of audio shipments held by CDs in a given year (lump records and tapes as the old technology). Project f for 1990. Calculate a 90 percent confidence interval for 1990. Comment on the strengths and weaknesses of your Fisher-Pry substitution model.

- 10.3** Try the alternative procedure given in Equation 10.3 to calculate Fisher-Pry substitution for CDs (Table 10.4). Which year-to-year increase would you choose to figure α ? Note that α is defined in terms of annual fractional growth. You cannot simply compute this on the basis of growth over, say, a five-year period, divided by 5. Instead, much like compound interest, you must compute it as

$$(1 + \alpha)^n$$

where n is the number of years spanned.

Using this alternative procedure, you should:

1. Calculate a value for α ;
2. Then, using the f values for one known year, calculate t_0 ;
3. Use this equation to project f in a certain year (or, conversely, to determine the year f will reach a certain level).

Project f for 1990. Compare with the results of Exercise 10.2.

- 10.4** Enter data for tapes in the TOOLKIT. Compute both Fisher-Pry and Gompertz models. Comment on their strengths and weaknesses.
- 10.5** Show that the Fisher-Pry equation given by Equation 10.4 and the linear form of the Pearl equation (Equation 10.3) are the same equation when b and $\ln(c)$ are expressed in terms of α and t_0 .

--- 11

EXPERT OPINION

OVERVIEW

Gathering and analyzing the opinions of experts are very important activities for the technology manager. This chapter describes approaches for identifying experts and selecting a technique. Common methods for gathering opinion are introduced and contrasted: interviews, meetings, surveys, the Nominal Group Process, the Delphi technique (and variations on it), and the POSTURE method. The example of building a computing environment for a small engineering college (Commission for Computer Integration, 1989) is used as a unifying theme throughout the chapter.

11.1 INTRODUCTION

Throughout his or her career, the technology manager will be faced with many decisions for which he or she lacks specific expertise. In some instances, the need for decisions based on this expertise will be immediate. In others, the manager will need to make preliminary decisions to position the firm advantageously vis-à-vis longer-term projections of technology growth. Regardless of which type of decision is faced, expert input is often crucial. However, it is important that the manager not forfeit responsibility for any decision to experts. Experts seldom command the whole range of conditions that the manager must consider. Nor do they have the ultimate responsibility to sell the decision to management, to implement programs, and to reap the reward or shoulder the blame for failure. The manager must synthesize expert input with a personal assessment of risks and potential benefits and

add an understanding of the corporate environment that only he or she can provide. The manager must devise methods to secure expert input that can be used to reach a decision.

Two fundamentally different kinds of expert input may be required: *fact* and *opinion*. The former deals with existing facts, the latter with future facts. The search for facts generally is the more tractable situation. Experts may be biased or may have differing interpretations of the facts, but their inputs can be cross-checked against each other and against those of other sources. The search for future facts is more uncertain. While individuals in the field or related fields can be expected to give informed opinions, bias often is difficult to isolate, and even a group of experts can be caught up in prevailing fashions (Schnaars, 1989). While expert input about facts *may* differ, expert opinions *will* differ. This chapter emphasizes special problems inherent in obtaining opinion.

To gather expert input, a manager must accomplish two fundamental tasks: (1) the expertise required and the individuals who possess it must be identified, and (2) method(s) to obtain their input must be chosen. The choice of method often will influence the selection of experts and vice versa. Different methods will be dictated by the location, availability, and schedules of experts, and by their individual and group characteristics. For example, the experts identified may not be employed by the firm or may be geographically scattered. They may hold vastly different world views, speak different technical languages, be mutually antagonistic, or simply have no history of effective communication and cooperation. Under such circumstances, a traditional, interactive, face-to-face meeting may be inadvisable or impossible.

Before selecting a method, the manager also must assess the type of information needed, the detail and accuracy required, the timing, and its potential importance to achieving corporate goals. Each method has different strengths and weaknesses, produces different forms of input, and implies a wide range of costs. Often, more than one method will be required. The tool must be adapted to the task—a sledgehammer is of little use to attack a fly on a window.

11.2 SELECTING EXPERTS

Before selecting experts, the technology manager must determine the range of expertise he or she requires to provide input for the decision. This is not a simple or an error-free task, and it must be carried out iteratively. Perhaps the best place to start is by probing the dimensions of the decision. Only when these are understood can one determine the input required. Decision dimensions often are controlled by factors internal to the manager's organization. Therefore, he or she should turn first to individuals within the firm; they compose the first group of experts to be consulted. While they may not be experts in the technology being considered, they are expert in those areas in which the technology will be used, maintained, administered, sold, and/or manufactured. Thus their input is vital as is their commitment to and support of the decision.

To identify those within the company that the decision will affect, the manager might begin with a series of questions.

- Who will implement the new technology?
- Who will maintain it?
- For what tasks will it be used?
- What departments or activities will be influenced by the results of the application?
- What resources are available?
- What potential costs and benefits might be expected?

Answers to these questions, no matter how tentative, help the manager map information needs.

For example, suppose a small engineering college (SmEC) with a limited computer base wishes to expand and modernize its computer system to provide a platform for modern educational and engineering applications. The data systems manager is to determine the computing system required and to make decisions about the way in which its use will be integrated into the engineering curricula at SmEC. The data systems manager might begin by soliciting input from management (the president, business manager, development director) about resources that will be available. Next he or she might turn to the faculty to determine academic uses that might be made of the system, to students who will use it, and to the computing center director who will implement, maintain, and operate the system. After consulting with these individuals, the data systems manager might form a committee to probe more deeply.

Next the data systems manager should determine if others within the organization have faced similar decisions. Who are they? How did the decision differ? What information did they seek? From whom? What information do they wish they had possessed?

The manager should then determine if there are individuals in other organizations who recently have made similar decisions. Articles and ads in trade literature can be useful sources for such contacts, as can trade and professional organizations. In this step, the manager broadens the range of experts. The individuals identified may have experienced the changes produced by decisions similar to that which the data systems manager must make.

Having completed these steps, the data systems manager can make a preliminary estimate of the expertise that will be required to provide an information base. For instance, the data systems manager at SmEC might decide that he or she needs to know what hardware and software systems are available and what the requirements and costs (factual information) of such systems are. Other factual information needs could include how the systems presently are used in other engineering colleges, faculty resource requirements, and measures of effectiveness and acceptance by faculty and students. Opinions might also be needed to acquire other necessary

information: projections of improvements in computer workstations; software and networking technologies; probable changes in industry and computer standards; assessments of potential educational applications and changes in faculty and student productivity; and estimates of probable accrediting agency positions on computer-based learning.

Vendors, technology developers, and established users are good sources for factual information. After confidentiality agreements are executed, they are knowledgeable about the directions of their company's product development and research. They also are good sources for opinions about industry directions and new product introduction and timing. While vendors are biased, their bias is clearly indicated by the logos on their business cards; the bias of experts from other areas is not so easily determined.

For opinions about issues farther removed in time and/or content from production and sales or about issues that span multiple technical, social, or economic fields, the problem of identifying experts may require trial and error. Lipinski and Loveridge (1982) note that in nearly any field there is a peer group who knows the current pecking order of the group. These various peer group networks should be searched until individuals emerge who are recommended repeatedly. Individuals within the firm, government, or university, and "think tank" personnel often can supply starting points for network searches, as can professional organizations, journals, and periodicals.

Merely identifying individuals is not the end of the process. Although various classifications of the type and degree of expertise are possible, Lipinski and Loveridge (1982) suggest considering the following types of experts:

- Generalists with a spread of interests and perceptions that give them a high level of awareness of the broad context
- Persons of thought who have particular and deep knowledge of specific fields
- Persons of present or future action whose present or likely future positions make it possible for them to affect the technology (p. 214)

They conclude that experts possess three important characteristics: substantive knowledge in a particular field, ability to cope when faced with uncertain extensions of that knowledge, and imagination. In most cases, the technology manager will not have the time or resources to determine all three characteristics of each expert. However, the scale for self-ranking of expertise presented in Table 11.1 may be useful in assessing substantive knowledge.

Finally, the technology manager must ascertain that the range of expertise represented by the experts selected covers that previously determined to be necessary to support the decision. It should be realized that most experts will have subsidiary knowledge in related fields as well as deeper expertise in their primary field. This subsidiary knowledge may be sufficient to cover less-critical categories of information. Galanc and Mikus (1986) describe means to determine the minimum number of persons from each area necessary to cover the field of interest.

TABLE 11.1 Self-evaluation Criteria: Guidance to Self-ranking of Expertise

1. You are unfamiliar with the subject matter if the mention of it encounters a veritable blank in your memory or if you have heard of the subject yet are unable to say anything meaningful about it.
2. You are casually acquainted with the subject matter if you know at least what the issue is about, have read something on the subject, and/or have heard a debate about it on either a major television or radio network or an educational channel.
3. You are familiar with the subject matter if you know most of the arguments advanced for and against some of the controversial issues surrounding this subject, have read a substantial amount about it, and have formed some opinion about it. However, if someone tried to pin you down . . . you would soon have to admit that your knowledge is inadequate.
4. You were an expert [with the subject matter] some time ago but feel somewhat rusty now because other assignments have intervened, [However,] you have kept reasonably abreast of current developments in the area, thus trading breadth of understanding for depth of specialization.
5. You should consider yourself an expert if you belong to that small community of people who currently study, work on, and dedicate themselves to the subject. Typically, you know who else works in this area; you know the literature of your country and probably the foreign literature; you attend conferences and seminars on the subject, sometimes reading a paper and sometimes chairing the sessions; you are most likely to have written and/or published the results of your work. . . . Other experts in this field may disagree with your views but invariably respect your judgment; comments such as "this is an excellent person on this subject" would be typical when inquiring about you.

Source: Lipinski and Loveridge, 1982. Reprinted by permission from *Futures*, vol. 14, no. 2, 1982.

11.3 SELECTING THE TECHNIQUE

Nelms and Porter (1985) provide a useful scheme for classifying the various expert opinion techniques and, more importantly, for choosing among them. They note that Gustafson and colleagues (1973) identify three distinct processes in expert opinion capture: *talk* (interaction), *feedback* (one-way communication), and *estimate* (decision). All techniques described in this chapter embody combinations of these three processes. Table 11.2 shows the techniques discussed in this chapter (plus brainstorming, which is covered in Chapter 7) with their classification relative to the three processes.

There are six factors (Nelms and Porter, 1985, pp. 57–58) to consider in choosing a technique or designing one especially for a given situation:

1. *Logistics*—Resources largely determine the method. If severely limited, a one-round estimate (that is, no feedback with additional input from participants) provided by a survey is the most appropriate. However, feedback

TABLE 11.2 Classification of Expert Opinion Techniques

Technique	Talk (T)	Feedback (F)	Estimate (E)	Process
Committees	X		X	TE
Brainstorming	X			T
Nominal Group Process	X	X	X	EFTE
Surveys			X	E
Delphi		X	X	EFE
Shang Inquiry		X	X	EFE
EFTE	X	X	X	EFTE
POSTURE	X		X	ETE

is desirable, so estimation plus feedback is usually the minimum desirable format.

2. *Feedback*—The format used for feedback does not seem important as long as extreme positions are not overemphasized. Minimizing delay between successive feedbacks seems desirable.
3. *Communication medium*—The medium of exchange appears to be a function of access to experts, resources, and time.
4. *Sample size*—While it usually is desirable to maximize sample size, studies indicate that groups of 12 or fewer converge on valid results even when individual judgments are only moderately valid.
5. *Stopping rule (multiple communications)*—The number of rounds used is a tradeoff of time and resources against the stability of group and individual responses. Consensus can be used as a criterion, but it occurs at the cost of increased group pressures.
6. *Interaction*—It appears that interaction is most usefully seen as a function of task complexity and of group social structure (for example, social complexity increases if there are serious disagreements that individual members care about deeply). The greater the complexities, the more interaction is needed.

The techniques discussed in the following section differ in their strengths and weaknesses, as well as in their time and resource demands. In many instances, a combination of techniques will be the most satisfactory.

11.4 TECHNIQUES FOR GATHERING EXPERT OPINION

Common methods for gathering expert input are considered in this section. Those used in the individual mode, such as a face-to-face (in person) or telephone interview, are detailed first. Those used in the immediate group mode—that is, face to face but with several experts at one time—are considered next, followed by a discussion of group methods usually applied at arms length. Many technology managers favor face-to-face methods because such techniques allow the manager

to assess the candor, judgment, and creativity of the expert. However, it often is difficult to meet individually with numerous experts in person or collect all of them at a single place and time. Arm's-length techniques, such as surveys and the Delphi technique, can provide structure and depth that is often impossible in the give-and-take of a face-to-face meeting. Thus each approach has value in the right circumstances. (For an excellent review of various input techniques see Wolfe, 1987.)

11.4.1 Individual Input

Individual input can be obtained in person, by telephone, or by electronic mail. In-person and telephone interviews are quite common. These interviews can be structured, focused, or nonstructured.

A *structured interview* has an explicit schedule of fixed questions that usually are in closed form (the responses to open-ended questions are sometimes difficult to decipher and categorize). An example of a recent structured interview asked schoolchildren specific questions concerning the level of technology that they expect to experience at the beginning of the twenty-first century: "Do you think that you will be living and working in space in the year 2000?"

A *focused interview* is directed to respondents known to have pertinent knowledge. Prior to the interview, the interviewer must carefully study the topic. An interview guide should be prepared that specifies a set of topics to be discussed. The emphasis of this type of interview is to obtain subjective information regarding the situation under study. For example, the following question might be directed toward a computer chip manufacturer: "What advances do you see in the development of VLSI (very large scale integrated) chips in the next 10-year period?" This is a question that does not have one answer. It would be asked only after the interviewer has carefully considered the possibilities for future chip development and has studied the progression of integrated circuitry. The interviewer should have enough background knowledge to ask for clarification if necessary or to ask for expansion if the response is not sufficiently detailed or if the response goes off in a tangential direction.

The *nonstructured, or nondirected, interview* has no schedule and no prespecified set of questions. Respondents are encouraged to reveal their opinions and attitudes as they see fit. This type of interview is difficult to control, and the responses obtained are difficult to compare. A sample opening question for such an interview might be: "What advances can we expect to see in office automation in the next 10 years?" From this point, the interviewer must wing it, following up responses with additional leading questions. After talking with several respondents, the interviewer draws inferences from the collected responses.

When using either focused or nonstructured interviews to gather information, the interviewer must be clear on the objectives to be accomplished since interviewees may ask you to explain the purpose of a particular question. It also is important in all three types of interviews that only critical questions be asked; questions that are merely interesting should be saved for a cozy discussion by the fire. The

interviewee's attention can be captured for a limited time, and any question that is not directly pertinent to the technology forecast could sacrifice more important information. Interviews should last no longer than 45 minutes, which limits the number of questions that can be asked.

It is highly advisable to conduct a *pilot interview* prior to the real thing. In the pilot, several respondents, or stand-ins, are asked the questions that will be raised in the actual interview. If they have problems answering these questions, or if they give responses in a different manner than that anticipated, revision of the interview is needed.

These revisions should remove ambiguity, maintain pertinence, and revise leading questions. Leading questions can bias responses severely; asking the same question different ways can provide strikingly different responses. "Office automation seems to have reached a plateau in 1990, do you agree or disagree?" is an example of a leading question. The respondent is inclined to agree, thinking: "Yes, by 1990 a lot did happen in office automation. Perhaps the interviewer knows a lot about this subject; I don't want to disagree with someone who is knowledgeable about this."

The sequencing of questions in a personal interview also affects the answers that are given. Questions can be funneled from the general to the specific, or, less frequently, in the opposite direction. Funneled questions lead from one to another: (1) "What do you think are going to be some of the most important technological developments in computer operating systems in the next 25 years?" (2) "Of all those developments mentioned, which do you think will be the most important?" (3) "On what basis did you select these as the most important?" In the inverted funnel sequence, narrower questions are followed by broader ones.

11.4.2 Committees, Seminars, and Conferences

These group techniques require experts to be in the same place at the same time. Proceeding from committee to seminar to conference, the formality of meeting format and the number of participants generally increase, while the opportunities for interaction decrease. Committees usually meet more frequently than seminars, and seminars meet more often than conferences. Generally, committees are charged with a specific task, whereas conferences and seminars exist to exchange information. Finally, committees most often have fixed membership; conference and seminar attendance is less rigid, although there usually is a core of frequent participants. Committees, seminars, and conferences are all potentially beneficial platforms for expert input; however, the committee is the most familiar, and it is the only one that will be discussed in detail.

The committee is ubiquitous in modern organizations. It is the most common method used when more than one expert's input is required. It is the easiest face-to-face meeting to convene (especially if experts are employees of the organization) and a comfortable format for participants. Martino (1983) notes two major advantages to the committee method: (1) the total knowledge of the group is at least equal to that of any one member and (2) the number of factors that it can consider

at least equal those of any one member. The former is important because the meeting is intended to increase the knowledge base for a decision; the latter is significant because bad decisions often can be traced to neglect of a critical factor. While the committee also possesses at least as much incorrect information as any member, it is assumed that the group's pooled knowledge will cancel out this misinformation.

Although the committee meeting is common, it often is poorly conducted and frequently is not seen as productive relative to time invested. A successful meeting depends heavily on the skill and preparation of the chairperson and of the committee members. Although the meeting usually is conducted in an unstructured give-and-take format, the chairperson should prepare an agenda and ensure that members understand the topic and goals in advance. Members also should, but seldom do, prepare. Moreover, there are limitations inherent in the give-and-take process that can limit the success of the unstructured committee meeting.

In a face-to-face meeting no one is anonymous. Experts are identified with positions that they feel must be defended to preserve their credibility. Other members suppress input in deference to experts or to those of status, authority, or power. The pressure to conform often produces a false or premature consensus through bandwagon effects. Group conclusions sometimes can be influenced more by the repetition of an input than by its validity. Thus the premise that information will cancel misinformation often goes unrealized. Although some of these deficiencies can be countered by effective leadership, the complexity of the topic often defeats efforts to stay on track, and the meeting wanders from the issue.

There are benefits of the committee meeting that are overlooked when productivity is considered too narrowly. For example, if the pressures for consensus inherent in the meeting are handled effectively, a group identity and a commitment to the subsequent decision can be forged. Moreover, the meeting can educate committee members about dimensions of the decision and about developments and concerns in areas of expertise other than their own. Thus benefits from committee activities can sometimes surface in areas quite different from immediate meeting goals.

Returning to our example of the small engineering college, based on preliminary discussions with internal experts, the data systems manager forms a committee to explore the dimensions of the computer system decision. This committee is made up of representatives of the administration, students, and one faculty member from each academic department at SmEC. Each is expert in some dimension of the problem, and some also are knowledgeable about computer hardware, software, and applications. The data systems manager sends the first meeting agenda to each member and states that the committee's goal is to establish the characteristics required of the new computing system. Before the meeting, the manager discusses the agenda and goal with each member and suggests types of input that seem most useful. The data systems manager plans the first few meetings as opportunities for participants to educate each other about their needs and expectations. Each meeting is conducted so that everyone participates, and overly vocal or dominant individuals are neutralized by a statement such as "Thanks for that input, does anyone else have a comment?" Additional meetings (described in later

sections) are planned to develop a substantive definition of the new computing system.

Brainstorming is a technique used to produce creative ideas without evaluating them (this technique is discussed in Chapter 7). Imagine that the data systems manager at SmEC might devote an early committee meeting to a brainstorming session to elicit ideas about the kinds of applications the new computer system should support. Later, the manager and/or the committee might evaluate the suggestions to determine those with promise. These might require additional expert input for full evaluation.

11.4.3 The Nominal Group Process

The Nominal Group Process (NGP) is a powerful group technique developed by Delbecq and Van de Ven (1971). It has been used effectively in the early stages of problem definition (Roper, 1986, 1988; Sachs and Clark, 1980); it also can be used to define the dimensions of a decision and to identify options and questions for conventional or Delphi surveys. Like brainstorming, the NGP is designed to overcome the unproductive aspects of unstructured, face-to-face meetings and to stimulate creative thinking by a group of experts. When carefully applied, undue pressures for consensus, bandwagon effects, and the negative effects of hostility between participants can be avoided. The process also is structured to build strong group identity and support for subsequent decisions.

The NGP is based on indications that small groups that *do not interact* are the most creative idea generators, while groups that *do interact* perform the best evaluation (Delbecq, 1974). Thus, unlike brainstorming, the NGP incorporates evaluation as well as idea generation. At the close of an NGP meeting, the manager not only will have a rich list of ideas, he or she will also have a preliminary rating of those the group feels are most important. The six steps of the NGP are presented in Exhibit 11.1. Also given there are brief statements of the objectives of each step. As in any group meeting, the manager should carefully explain the agenda and the meeting goals before starting.

■ Exhibit 11.1 Steps in the Nominal Group Process (Based on Roper, 1988; and Delbecq, 1974).

Step 1: Silent (Nominal) Idea Generation. Each participant is urged to list all of the factors that he or she believes are central. Participants work silently and individually from 10 to 25 minutes.

Objectives: "To give the participants time to think, to force them by tension and social competition to produce, to avoid polarization and premature evaluation and closure, and to avoid pressure for conformity and deference to status."

Step 2: Group Round-Robin Listing of Factors. Sequentially, each participant reads one factor from his or her list. This is done without explanation or advocacy and without question or comment by the group. Each factor is written on a flip chart

and posted for all to see. This process continues until all factors have been listed; generally it takes 45 minutes to an hour. The round-robin sequence gives each participant control of the floor in turn and draws input from reticent individuals unlikely to contribute in the give-and-take of a normal meeting format. The process often triggers new factors not listed during Step 1.

Objectives: "Equal sharing of ideas, depersonalization of ideas, production of a written record, focusing of the problem, visual and auditory concentration, polarization of conflicting viewpoints."

Step 3: Discussion and Clarification of Listed Factors. Each factor listed on the flip charts in Step 2 is discussed informally by the group. Factors are considered in sequence. The primary thrust is clarification. The group leader probably will find it best to discourage advocacy at this point, as the cooperative spirit of the group may still be fragile—especially if antagonism initially existed among participants. During the discussion, additional factors may be added to the list. These factors may be new or may combine factors already listed to reduce repetition and overlap. New factors may be formulated by individuals but are generally group products.

Objectives: "Clarification, reduction of repetition, overlap, and ambiguity."

Step 4: Silent (Nominal) Individual Written Voting on Priorities. Each participant is asked to silently and independently select the most important factors from the list compiled through Step 3. This is not the final ranking of factors; rather it is intended to focus attention on the quest for the most important factors, thus reorienting the thrust of the first three steps during which breadth, not focus, was sought. The specific number of factors participants are directed to select is flexible. However, Shillito (1973) suggests having each participant vote on 20 percent of the number of factors on the list. Occasionally, this may prove an unwieldy number. If so, a smaller percentage can be selected.

Objectives: "Focus on most important issues, rank ordering of items, consideration of alternatives."

Step 5: Discussion of Voting Results. Before the start of this step, the group leader tabulates the votes from Step 4 and displays the factors with the highest totals on the flip chart. The group is then asked to evaluate the result. Typically, they are not satisfied. This dissatisfaction sharpens the focus. A general discussion of why the preliminary result is unsatisfactory leads to a more general discussion of evaluation criteria and finally to the formulation of new factors that encompass the concerns of the group. New formulations often are produced by categorizing factors from the list under major areas of concern. By the time this step has been reached, a common group identity generally has been formed and a genuine concern to determine the most important factors is manifested. Antagonism is seldom displayed among participants. While this step often can take as long as the group leader is prepared to spend, a useful limit is 30 minutes.

Objectives: "To compare perceptions of the approaching consensus and to clarify misinformation."

Step 6: Final, Silent (Nominal), Individual Written Voting. In this step, participants are asked to select the most important factors from the list as expanded through Step 5. Selection is done silently and individually. Often a rating scale that ranges from 0 (no importance) to 100 (absolutely necessary) is used, with each participant displaying the relative importance of the factors he or she selects. The number of factors that participants are directed to choose should follow the guidelines suggested in Step 4. It should be emphasized that the preliminary ranking of factors (Step 4) has served its purpose—that is, it has sharpened the focus of the group. Therefore, factors should be selected from the entire list without regard to the preliminary ranking. In fact, it is best to remove the flip chart reporting the preliminary results before voting.

Objectives: "Final re-ranking of ideas and closure."

The NGP provides an effective platform for communication among experts from different fields. Comparisons (Bakus et al., 1982; Delbecq, Van de Ven, and Gustafson, 1975; Van de Ven, 1974) have shown that NGP meetings typically generate nearly twice as many ideas as unstructured, face-to-face meetings and somewhat more ideas than Delphi. The chairperson should run one or more trial NGP meetings before the real thing. He or she also may wish to review the detailed instructions given by Delbecq, Van de Ven, and Gustafson (1975).

Suppose the data systems manager at SmEC decides to use the NGP for a meeting of the committee of internal experts. Before the meeting the manager explains the process and that it will be used to examine academic objectives that should be sought through classroom/laboratory computer applications. The manager also explains that this information will form part of the input necessary to define the computing environment that the new system must foster. The goals of the meeting are to develop a broad list of objectives and an initial indication of which are most important. At the NGP meeting, the seven participants compile a list of 60 objectives. The final ranking of objectives is shown in Figure 11.1.

The NGP format can be modified to fit special circumstances. The most common is that the number of experts is greater than effectively can be accommodated in a single session. Delbecq (1974) suggests, for example, that the optimum number of participants is six. Others (see Roper, 1988) have found groups larger than ten unsatisfactory because the number of ideas generated is too large to be handled effectively in Steps 4 through 6 (see Exhibit 11.1). In this situation, Delbecq (1974) suggests dividing participants into subgroups that individually pursue Steps 1 through 3. These subgroups are then combined to report the results of their efforts. The combined group completes the remaining steps using a master list developed from individual subgroup results. However, this process is difficult to complete at a single session. Also, unless duplications are eliminated, the master list may be too ambiguous, overlapping, and repetitive to produce a satisfactory result. Therefore, Roper (1988) suggests planning two sessions and consolidating subgroup lists before beginning Steps 4 through 6 with the combined group.

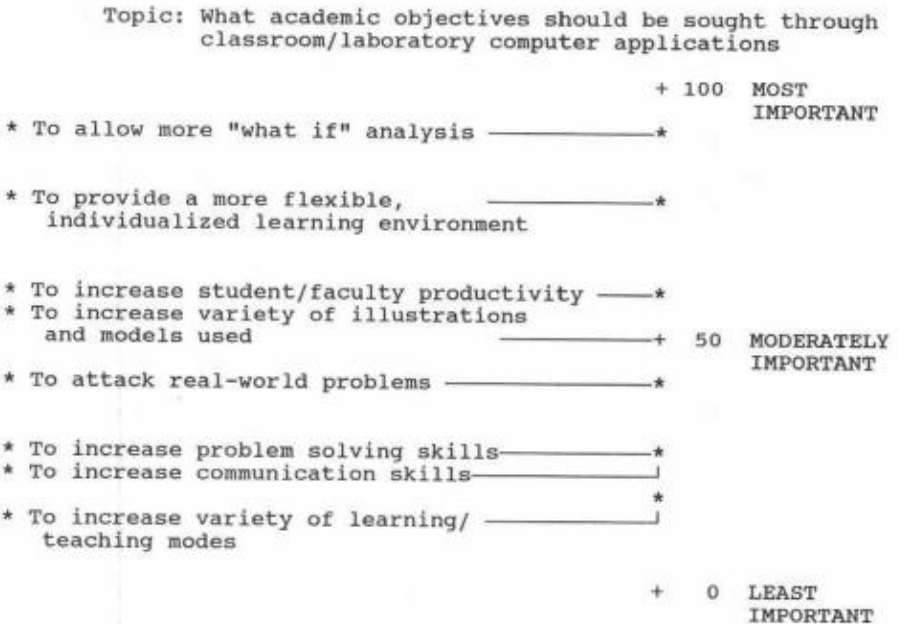


Figure 11.1 Final ranking of academic objectives from the Computer Committee at Sm EC. From materials developed during deliberations of the Commission for Computer Integration (1989).

When the primary purpose of the meeting is to develop breadth of perspective or to compile a rich list of ideas, Steps 1 through 3 can be very effective by themselves. In this format, the NGP resembles a structured brainstorming session. Limiting the process to these steps also can be used to gather input while delaying the perception or advent of closure. However, this weakens group identity and hence support for the final decision (Roper, 1988).

11.4.4 Surveys

Surveys are the most common method for soliciting input from groups of experts when face-to-face meetings are impractical. The method is popular because it is relatively quick, reasonably easy, and inexpensive; it can be used to collect input from a large number of experts in a variety of formats; and it avoids the negative dynamics of face-to-face meetings. Disadvantages (such as failure to produce consensus and problems with bad samples and bias) are shared to various degrees by all expert opinion techniques. Traditional surveys are one-time affairs and do not provide a formal feedback mechanism such as that built into the Delphi technique.

The survey embodies a number of basic assumptions. For example, group input is more likely to be correct than individual input. As in the face-to-face meeting, it is assumed that information will cancel misinformation. Since traditional polls do

not incorporate feedback, however, canceling must be achieved by some scheme for averaging responses. This technique also assumes that questions have been posed clearly, concisely, unambiguously, and in vocabulary shared by the respondents. It is also assumed that surveys are unbiased and do not include unstated or misleading assumptions. Such qualities are not always easy to ensure. The reader may wish to pursue examples given by Emory (1976) for a better understanding of the problems. Generally, at least some questions should be open-ended. This allows respondents to "use their own words" and thus gives a clearer picture of meaning and rationale.

Surveys can be carried out in the personal mode (face-to-face), the impersonal mode (printed questionnaire sent by mail), or in a combination of these modes (Emory, 1976). They also can be executed by computer via an electronic network (such as Bitnet) if the means are available to most respondents. This electronic mode offers some opportunities for unstructured feedback even in the traditional survey format. It also provides a means to begin processing some of the responses quickly. The personal mode is likely to be most expensive and time-consuming, but offers opportunities for increased flexibility and first-hand evaluation of expert credibility.

Once information needs have been identified and experts chosen, the next step is to construct the survey. Before beginning, the manager should review a basic text in the field such as Emory (1976), Jolliffe (1986), or Raj (1972). The manager then should *begin at the end* by considering the expected results: What response format will be most useful? How will input be organized, processed, and analyzed? Will input analyzed in this way fulfill information needs? Next, a draft survey should be developed and tested on a sample group (perhaps colleagues). Finally, the survey must be revised to remove deficiencies that have been found.

To be effective, the survey must be carried out carefully, with due consideration given to sampling, bias, construction, and analysis. Before executing a survey, the manager must carefully consider the form and substance of the input to be gathered in the context of the decision it is to inform.

Martino (1983) gives hints for good survey questions. Compound statements, for instance, should be avoided because it is hard to decide what responses to them mean. Suppose a question is posed to determine how important it is to provide symbolic manipulation support for teaching calculus on SmEC's mainframe computer. A low rating may mean that the expert does not think symbolic manipulation should be used in calculus; on the other hand, it may mean that the expert thinks personal workstations are better solutions than the mainframe. Ambiguous questions should be avoided for similar reasons. Ambiguity often stems from terms that are not well defined or from terms that mean different things to experts from different fields.

Too much descriptive material in a survey can be as harmful as too little. Additional information appears to help with concepts unfamiliar to participants and confuse them when concepts are familiar. Individuals also seem to find it more difficult to deal with negatively posed questions and questions that use negative comparative adjectives (the term "less", for instance) than with those framed positively (Cassels and Johnstone, 1979). Finally, Lipinski and Loveridge (1982) note

that many persons find it difficult to estimate probability. For example, "When will there be a 25 percent probability that something will occur?") is a difficult question to answer. Estimates of "nearly certain" (greater than a 90 percent probability), "an even bet" (50 percent probability), and "almost no chance" (less than 10 percent chance) appear to be easier to handle.

11.4.5 The Delphi Technique

The Delphi technique is a special form of survey. Delphi, however, is designed to ensure a participant's *anonymity*, *controlled feedback and iteration*, and a *statistical group response*. The process was developed in the early 1950s at the Rand Corporation for military applications. After declassification, it was recognized as a useful method for technology forecasting and planning by American business and industry. Delphi since has spread around the world. Unfortunately, its rapid diffusion often has led to careless use. This carelessness stems, in part, from uncritical acceptance of the technique without adequate attention to underlying assumptions and requirements (Rieger, 1986; Sackman, 1974).

A Delphi survey is composed of a series of questionnaires known as "rounds." Input from each round is gathered, analyzed, and fed back to participants in the next. Typical feedback includes the median and the interquartile range of group response and the rationale for agreement or disagreement with group input. The first round normally should be open-ended to ensure that important aspects of the topic have not been overlooked by the manager in preparing the survey. The process continues until response stabilizes—usually within four rounds. A sample application of the Delphi technique is presented in Exhibit 11.2.

Delphi participants generally act anonymously. This reduces tendencies for an individual to defend untenable positions in order to preserve credibility and for a vocal or powerful individual to dominate. Feedback is intended to inform participants of the current position of the group and arguments for and against different points of view inherent in that position. Martino (1983) notes that an important aspect of the feedback is to allow the group to concentrate on the goals of the *process* rather than on social goals "such as winning the argument or reaching agreement for its own sake" (p. 17). Statistical treatment of responses is intended to portray differences of opinion within the group; the arguments are intended to portray the rationale supporting those differences.

The growth of Delphi has not been without controversy. In the 1970s, many issues of contention were raised (such as the value of anonymity, superiority of pooled group judgments, the validity of qualitative versus quantitative research within the scientific paradigm). However, at this stage of Delphi development, questions of sound application appear more important to the technology manager. Rieger (1986) suggested that Delphi is most appropriate for normative forecasts and noted that areas that are value-laden and that demand participation by multiple publics are prime targets for Delphi.

Many users pay insufficient attention to basics such as population sampling, piloting procedures, questionnaire validity, and reliability. Population sampling and piloting procedures are common concerns of all survey techniques. Questions about

validity and reliability, however, are more critical to Delphi because of its multiple-round character. These primarily deal with whether or not stability has been reached in the responses to succeeding rounds. Mouly (1963) suggested retesting to determine if responses change from the original survey. However, the time and work involved in Delphi surveys (for both manager and participants) make this impractical in most cases. A more practical approach is application of the chi-squared test for group and individual response stability suggested by Chaffin and Talley (1980) and Dajani, Sincoff, and Talley (1979).

Martino (1983) gives useful suggestions for Delphi surveying. First, the manager should obtain the agreement of experts to participate, which may require contracting and paying them. Careful selection of experts and attention to sampling are wasted if the survey has a low response rate. Second, the Delphi method, especially its iterative nature, should be explained carefully to all participants. Next, the manager should make the survey as easy to complete as possible. The number of questions also should be limited. Martino suggests a maximum of 25 in most cases. Sometimes mutually exclusive events or conditions may be included as a result of input from the unstructured round. The reason for these inclusions should be explained so that participants do not feel their consistency is being checked. Finally, managers must resist temptations to include their own input to avoid distorting survey results.

A lot of work is required to conduct a Delphi survey (Linstone and Turoff, 1975). To maintain an acceptable turnaround between rounds, computer analysis is desirable, even for modest-sized polls. Even then, each round of a mail survey will take about one month. (This time can be reduced considerably by using electronic mail or next-day delivery.) Internal surveys, however, can be completed in less than one-half this time.

■ Exhibit 11.2 Application of Delphi to Determine Computer System Functionality (Based on Commission for Computer Integration, 1989).

The data systems manager from our earlier examples has combined Nominal Group Process input with other input to define the ideal computing environment for SmEC. The next step is to determine the system characteristics necessary to deliver that environment. The manager would like to pose these characteristics independent of hardware and software considerations so as not to limit the range of possible vendor equipment or to cloud group perspective by what specific equipment will or will not do. Based on analysis of various inputs, the manager poses system characteristics in terms of *functionality*—that is, the capability of the system to deliver the services required to provide the ideal computing environment. The group decides functionality, F , can be described in terms of three components—ease of use, E ; versatility, V ; and capacity, C :

$$F = w_E E + w_V V + w_C C$$

where each w represents the relative importance (or weight) of each component.

The manager needs three types of information to flesh out the descriptor F :

1. The factors upon which the components depend
2. The appropriate weight for each component
3. The appropriate weights for each factor composing each component

Once this information is obtained, the manager can evaluate the existing system and determine specific areas in which changes are needed and the relative importance of those changes. F also will provide a metric with which to measure products offered by different vendors. Because of the complexity of the deliberation, the manager decides to use Delphi.

The manager begins by explaining Delphi, the type of input that is desired, and how it will be used. These descriptions are given in writing for future reference and are discussed. Each component of F is defined:

Ease of use—Measures the system's physical and operational characteristics that affect the ease with which it can be used to accomplish educational tasks.

Versatility—Measures the ability of the system to accommodate both a broad range of uses and changes in those uses—in other words, measures the system's robustness and flexibility.

Capacity—Measures system capability to perform the tasks necessary to support the academic goals of students and faculty.

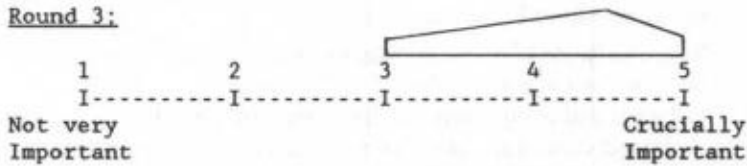
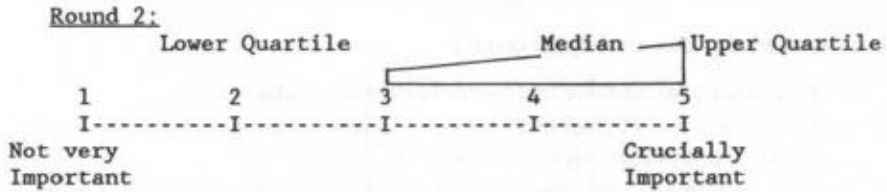
Round 1 of the Delphi poll identifies 9 factors upon which participants feel E depends, 12 factors that determine V , and 7 that define C . Results for Rounds 2 through 4 are shown in Figure 11.2 for one of the factors defining E . (Figure 11.2 is shown in the same form used to provide feedback to the group.) The weights for each component determined by normalizing group responses are

$$F = 0.33E + 0.30V + 0.37C$$

Weights for each factor of each component are found in a similar manner. The values for E , V , and C would be unity for a perfect system. Such a system would produce an F of 1.

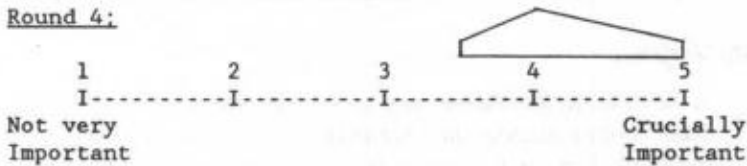
A second Delphi survey is conducted to evaluate the functionality of the system presently operated by SmEC. Tight control is possible since all participants work at SmEC and a 100 percent response is achieved for each round of each survey. Chi-squared tests for group stability (Dajani, Sincoff, and Talley, 1979) and individual stability (Chaffin and Talley, 1980) both are satisfied for the question shown in Figure 11.2. Differences in administrative, staff, faculty, and student responses are not probed. ■

The Delphi survey is a powerful method to obtain the subjective input of experts. However, it takes longer, requires more work, and is more expensive than conventional surveys. When conducting a Delphi survey, the manager must pay close attention to fundamental questions of technique and application as these are areas in which many surveys fail.



More Important: - The variety of operating systems now available that have a common user interface would greatly enhance ease of use and willingness to use the system for different applications.

Less Important: - All command structures are somewhat arbitrary. Uniformity and intuitiveness would be nice, but not as important as other factors.



More Important: - The variety of operating systems now available that have a common user interface would greatly enhance ease of use and willingness to use the system for different applications.

- Intuitiveness of commands would result in greater acceptance and less fear.

Less Important: - All command structures are somewhat arbitrary.

Uniformity and intuitiveness would be nice, but not as important as other factors.

Figure 11.2 Delphi responses by round regarding uniformity and intuitiveness of command structure.

■ Exhibit 11.3 The EFTE Procedure

1. Participants are provided pertinent background information.
 2. Questions about background information are resolved in a face-to-face meeting by the manager who discourages discussion.
 3. Each participant completes a Delphi survey and returns it to the manager.
 4. Group input is summarized and displayed in the usual format. In addition, responses are ranked from highest to lowest.
 5. Group response is discussed freely among participants, but the anonymity of individual responses is preserved.
 6. A second Delphi round is completed. Index cards are provided for anonymous questions and comments. Both survey and cards are returned to the manager.
 7. Delphi results again are summarized and displayed. Anonymous comments and questions also are read and posted.
 8. Participants once more discuss group input.
 9. Delphi responses are examined for individual stability. If stability is not found, Steps 6 through 9 are repeated.
 10. Final results are summarized and given to all participants. Additional statistical analysis may be performed. ■
-

11.4.6 Delphi Variants

Many variations on the basic Delphi theme have been proposed over the years. The simplest involve detail (not necessarily unimportant) modifications to the process. For example, Martino (1983) notes variations such as (1) beginning with a list of events or concerns rather than an unstructured round designed to elicit them and (2) executing a technology forecasting Delphi within a predefined social, economic, and/or political context. Other variants limit the number of rounds, although this perhaps is best done by measuring response stability (see previous section). Even the medium has been changed by computerizing the entire process so that it is conducted in "real time." The most radical changes, however, involve breaching the anonymity condition by face-to-face discussion of intermediate or final results and/or feedback.

An interesting variant, *Shang Inquiry* (Ford, 1975), combines Delphi with a method used by the ancient Chinese to divine information from their gods (or, since few of us are students of Far Eastern social customs and all of us were kids, the game we called "yer gettin' hotter"). In this method, experts estimate the highest and lowest values of a variable in the opening round. The manager uses these to find the lower and upper bounds of the variable and the midpoint, R . In subsequent rounds, group modal response to the query "Is the correct value higher or lower than R ?" is used to zero in on the correct value. For example, if the response is that the value should be higher, the midpoint for the next round is computed by averaging R with the upper bound; if the response is that the value

should be lower, R is averaged with the lower bound. Ties are broken by a coin toss. Thus each round elicits a progressively more precise estimate.

Nelms and Porter (1985) suggest a technique they call Estimate-Feedback-Talk-Estimate or *EFTE* (see Table 11.2). This technique's most significant variation from classical Delphi is the free discussion of results and the relatively fast feedback embodied by the process. Ten steps are involved, excluding iteration (see Exhibit 11.3).

There are other modifications to Delphi, and new ones undoubtedly will continue to be proposed.

11.4.7 POSTURE

The Policy Specification Technique Using Realistic Environments (POSTURE) is based on concepts of policy capture (see Chapter 18) and social judgment theory. This technique is intended not only to determine the judgment of a group of experts, but also to model the judgment process itself. The rationale is

If one understands the judgment process, there is a good chance of improving it. If one can improve the judgment process, one should be able to improve the resulting judgments and the decisions they imply. (Brockhaus, 1975, p. 127)

POSTURE is an attempt to formulate a process to accomplish these lofty goals. It combines some of the input methods described in earlier sections with the technique of multiple linear regression, described in Chapter 9, to capture and model the judgment process of individual experts.

To use the process, the manager must first identify, perhaps using an earlier method such as brainstorming, the factors used to formulate the judgment. Next, weights indicating relative importance must be assigned to each factor. This usually is done by formulating a diverse set of short scenarios that incorporate the range of possible combinations of factors. Each expert evaluates every scenario using a numerical scale to indicate its desirability. Then the manager calculates a multiple linear regression equation for each participant. These equations presumably make explicit how each individual weighted the various factors to reach a judgment of desirability. Next the manager clusters the regression equations to produce an expression for subgroups of participants. Finally, participants may be brought together to try to formulate a consensus judgment about the scenarios. If a consensus is achieved, an equation is developed to represent the group.

POSTURE results heavily depend on the quality of the scenarios, successful identification of all factors pertinent to the judgment, and the selection and quality of participants. Its advantages primarily accrue from (1) capturing the judgment process and (2) realizing the value of the process as a means for individuals to learn more precisely the factors on which different groups disagree. The former allows the manager to simulate responses to similar decisions. The latter emphasizes the importance of this and most other input methods as learning processes for participants. The technique is fairly resource-dependent and may appear unrealistic or contrived to participants and, perhaps, to upper-level management. It seems

most appropriate to support situations in which a decision or similar decisions are made repeatedly (for instance, in the hiring process).

11.5 CONCLUSION

The basic concepts of obtaining expert opinion have been presented in this chapter. While it is impossible to cover all aspects of this complex and important task, concepts for selecting experts and methods have been presented as have the most common techniques. The manager should pay special attention to selecting experts who have the information required and to choosing a method that will provide the information needed for an informed decision. Care should be taken to review the six factors required to choose a technique (see Section 11.3) to ensure that the method chosen is compatible with information needs, resources, and time. The manager is cautioned that in the final analysis he or she, and not a group of experts, has the responsibility for a decision.

EXERCISES

- 11.1 You are the chief strategic planner for a major American aerospace company that seeks new commercial initiatives. You want to ascertain (1) if new trains (such as bullet trains and Mag-Lev trains) are likely to offer commercial promise in the coming decades and (2) whether your firm should enter this business. For each issue, discuss promising sources of expertise; whether individual or group expert opinion processes appear more compatible with your needs; and ways you could go about obtaining the cooperation of the appropriate experts in the appropriate format.
- 11.2 Prepare three survey questions that address the commercial promise of the new trains discussed in Exercise 11.1. Write whether these questions appear suitable for in-person interview, mail survey, and electronic mail administration. Set your written comments on suitability aside. Pilot test your three questions with three people (colleagues, friends) using the mail survey or electronic mail. Pilot the same three questions with three different people using an in-person interview. Consider the pilot test in terms of how well people understood your questions, surprising responses, and needs for revision. Compare these results with your written comments.
- 11.3 Form into groups of at least five persons and designate a manager. Your assignment is to conduct an NGP inquiry to identify and prioritize the factors necessary to make your campus (or company) more accessible to visually impaired persons. Follow the procedures outlined in Exhibit 11.1.
- 11.4 Conduct a three-round Delphi exercise on the following two questions. Identify an appropriate group of colleagues or fellow students to take the role of experts on "nanotechnology." Consider nanotechnology as the deliberate

manipulation of materials at the atomic level to build purposive structures (such as nanomachines).

Question 1—In what year do you anticipate the first laboratory assembly of a functional, inorganic nanomachine from atoms using physical manipulation (not chemical processes)? (“Never” is a legitimate answer; think in terms of molecules being assembled to have mechanical properties such as wheels or levers.)

Question 2—What is the probability (to the nearest 5 percent) that such nanomachines will have been implemented for initial, limited applications (commercial or military) within 25 years? (Think in terms of the introduction to market of a nanomachine product that a user could buy off the shelf.)

Have each group member anonymously answer each question. Compile responses and provide statistical feedback (in the form shown in Figure 11.1) to everyone. For the second round, request that those who choose to respond outside the interquartile range provide a one-sentence rationale. Again, provide statistical feedback on the second-round responses along with comments (anonymously). Complete a third round; no comments needed. Report the final-round results statistically. Did the process move toward consensus? Had results stabilized adequately?

- 11.5 Describe how you could adapt Exercise 11.4 to perform a Shang Inquiry. Would this be a better process than ordinary Delphi for the nanotechnology issue? Explain. (Optional: Perform a Shang Inquiry on these same two questions.)
- 11.6 Describe how you could adapt Exercise 11.4 to perform an EFTE process. Would this be a preferable process to Delphi for this issue? Explain. (Optional: Perform the EFTE process on these same two questions.)
- 11.7 Describe a technology forecasting or assessment issue for which POSTURE would be a good candidate.

--- 12

SIMULATION

OVERVIEW

Simulations portray some aspect of a real-world system that can be used to study its behavior. This chapter introduces simulation and several simulation methods useful in forecasting and technology management. The first method discussed is cross-impact analysis, a relatively new technique for which some even newer perspectives are offered. Cross-impact in turn provides the basic information for the second technique, KSIM modeling. This method can be easily employed to provide information about complex real-world situations. System dynamics methods are also described as a philosophy for analyzing and understanding complex real-world systems. Gaming is introduced as a means for studying the behavior of decision makers in search of goals or in competition with each other.

12.1 INTRODUCTION

Simulation means different things to different people. To an airline pilot, for instance, it means physical emulation of the cockpit and analog or digital emulation of aircraft flight behavior. An operations researcher, however, generally thinks of simulation in terms of discrete-event computer models that emulate the system he or she is studying. Once such models have been verified and validated, they can be used to study the effect of changes in the real system. A military analyst may see simulations as war games with stochastic behavior and may use them to study strategy and tactics, the effects of new weapon systems, or other battlefield possi-

bilities. War games are especially useful since the real system, fortunately, is rarely employed. The common notion among these different perspectives is that simulation simplifies a real-world system yet captures and portrays its major characteristics in ways that are useful for study and training.

While employed for the same purposes, different types of simulation are used in forecasting and managing technological systems. Several methods are considered in this chapter: cross-impact analysis, KSIM modeling, system dynamics, and gaming.

12.2 CROSS-IMPACT ANALYSIS

A basic limitation of many forecasting techniques is that they project events and/or trends independently (Stover and Gordon, 1978); thus they fail to account for the impact of events or trends on each other. For example, a successful nuclear fusion process could have a major effect on petroleum exploration. Likewise, the scarcity of petroleum resources holds great economic implications for the development of nuclear fusion. These two technologies do not exist in isolation. Each has a history; each is affected by developments in the other.

One approach to capturing interactions between events is to construct a model, that is, a formal representation of interactions among significant variables. There are several types that can be employed. A *mathematical model* uses equations to represent the system in which the events occur. Such models often require a major time investment to construct. Even with this investment, model coverage usually is limited (for example, mathematical models of inventory systems, of the economy, or of resource allocation systems). There are, however, special models that cut across disciplines and account for the effect of one event upon another. In the technology forecasting arena, one such model is *cross-impact analysis* (CI). Basic CI concepts are widely used and have applications in many areas, including natural resource depletion, strategy and tactics for warfare, institutional change, organizational goals, communication capability, and computer capabilities.

Since CI deals with the future, it involves uncertainty. Therefore, it is a stochastic rather than a deterministic model. Traditional CI is focused on the effects that interactions between events have on their probabilities of occurrence. Thus it deals with discrete events and incorporates no dynamic (time) dimension. While still discrete, the dynamic dimension can be added to CI using the concepts of Markov chains. This modification of traditional CI can be employed to study the chain reaction of events/trends on other events/trends over time. These CI approaches are described in the following subsections.

12.2.1 Traditional Cross-Impact

The concept of CI arose from a game called "Future" which Gordon and Helmer devised for Kaiser Aluminum in 1966 (Helmer, 1983). In the game, a future world was

constructed in which some or all of 60 events might have taken place (technological breakthroughs, passage of legislative measures, natural occurrences, international treaties, etc.). Each event was assigned an initial probability of occurring. As play progressed; these probabilities changed. Part of the change was due to actions of the players, the remainder was determined by the occurrence or nonoccurrence of other events. Changes of the latter type gave rise to the concept of CI.

A specific example is useful to understand how traditional CI works. Suppose we are planning for a particular communication technology, say facsimile transmission (fax). Fax technology allows text and images to be transmitted over normal telephone lines. Hard copy input is provided by the sender; it is then transmitted, providing hard copy at the receiver's end. Transmission is fast, 15 seconds or so per page; thus it is much cheaper than conveying the same information verbally (even when that is possible) and much faster than using the mail service. We wish to know what the future of fax technology will be over the next five years. First, we must determine the ways in which the operating environment is likely to change over that time horizon. Then we must identify events that could have noticeable impacts on the use and/or on the planned uses of fax and the probabilities that these will occur.

Suppose the events are identified as $E_1, E_2, E_3, \dots, E_m$. These represent entirely *external determinants*—that is, natural or man-made events over which we have no control (such as a global economic depression or legislation imposing a large tax on each fax received). *Events completely under our control are not included.* These must be treated differently. If the number of events grows too large, it may be necessary to rank them and retain only those that are most important. This could be accomplished by Delphi polling (see Chapter 11) or by having interested parties assign points to each event on a scale of 0 to 100. For the example of fax transmission, suppose we have identified the four events shown in Table 12.1. For

TABLE 12.1 Occurrence Matrix for a FAX Example

The Probability of This Event Becomes:	If This Event Occurs:			
	Increased Taxes/Cost	Negative Legislation	Replacement Technology	Market Saturation
Increased taxes or costs on transmission (0.70)*	1.00 $P(1 1)$	0.40 $P(1 2)$	0.70 $P(1 3)$	0.51 $P(1 4)$
Negative legislation (0.40)*	0.20 $P(2 1)$	1.00 $P(2 2)$	0.38 $P(2 3)$	0.31 $P(2 4)$
Development of a replacement technology (0.60)*	0.90 $P(3 1)$	0.72 $P(3 2)$	1.00 $P(3 3)$	0.33 $P(3 4)$
Market saturation (0.45)*	0.33 $P(4 1)$	0.35 $P(4 2)$	0.05 $P(4 3)$	1.00 $P(4 4)$

* Initial marginal (ceteris paribus) probability

convenience, we have arranged our work space in an occurrence matrix with the events E_1 through E_m ordered both across the top and down the lefthand side of the array. The next step is to estimate the probability that each will occur. These estimates are called the *marginal probabilities*. (They also are sometimes referred to as *ceteris paribus*—all else equal—probabilities to indicate that they are estimated without considering any of the other events.) These probabilities are subjective and might be estimated through some Delphi-like procedure. For our example, Table 12.1 shows that we have estimated the probability that there will be increased taxes or costs on fax transmissions to be 0.70.

We have completed two components of the CI matrix: the events critical to the forecast have been identified and their initial (marginal) probabilities of occurrence have been estimated. The cells of the matrix will be used to record the *conditional probabilities* (that is, the probability that event i occurs given that event j occurs). These probabilities are the heart of CI: they portray the impact that the occurrence of any event has on the probability that any other event will occur.

The conditional probabilities must be estimated next. First note, however, that the matrix diagonal entries all will be 1.00, for it is certain that event i will occur given that it has occurred. The first step is to compute the statistically acceptable range of conditional probability for each cell (pair of interactions) *above the diagonal*. These ranges will provide guidelines if we have no other basis from which to estimate the conditionals. This can be done using the marginal probabilities established previously for each event. To explain how to compute this statistical range, we must first introduce some statistical notation.

$P(i)$	= probability that event i will occur (the marginal probability of i)
$P(i j)$	= probability that event i will occur given that event j has occurred (the conditional probability of i given j)
$P(\bar{i})$	= probability that event i does not occur
$P(i \bar{j})$	= conditional probability that event i will occur given that event j does not occur
$P(i \cap j)$	= probability that both events i and j will occur (the intersection of events i and j)
$P(i \cup j)$	= probability that event i or j or both will occur (the union of events i and j)

By using the laws of conditional probability and the probability of compound events, Sage (1977) showed that limits exist to the range of statistically acceptable conditional probabilities. If the occurrence of event j enhances (increases) the probability that i will occur, then

$$P(i) \leq P(i|j) \leq [P(i)/P(j)] \quad (12.1)$$

On the other hand, if the occurrence of j inhibits (decreases) the probability that i will occur, then

$$1 + \{[P(i) - 1]/P(j)\} \leq P(i|j) \leq P(i) \quad (12.2)$$

Note that only the initial marginal probabilities $P(i)$ and $P(j)$ are necessary to compute these ranges, and they already have been estimated.

Next we must estimate a conditional probability for each of the cells above the diagonal and compare the estimates to the ranges computed from Equation 12.1 or 12.2. *Estimates that violate the computed ranges should be retained if a solid rationale for them can be given.* For example, in Table 12.1, the conditional probability $P(1|2)$ has been estimated as 0.40, which is within the statistically acceptable range, 0.25 to 0.70, computed from Equation 12.2. However, if we had estimated that it should be 0.15 and had evidence to support our estimate, we would enter 0.15 instead. Alternately, we could elect to assign one of the extreme values of the range to such a probability. Thus, lacking strong evidence to support our estimate of 0.15, we might choose $P(1|2)$ to be 0.25 instead.

Now that conditional probabilities above the diagonal have been estimated (the $P(i|j)$ s), we can turn to those below the diagonal (the $P(j|i)$ s). Here, we can use Bayes' rule to help. If the $P(i|j)$ was in the range established by Equation 12.1 or 12.2, Bayes' rule says that the corresponding probability below the diagonal should be

$$P(j|i) = [P(i|j)/P(i)]P(j) \quad (12.3)$$

If $P(i|j)$ was not in the range or if we do not agree with the value produced by equation 12.3, we should subjectively estimate the value of $P(j|i)$. In other words, if the values computed using Bayes' rule are reasonable, keep them. Otherwise, estimate values believed to be more appropriate. For example, in Table 12.1, the conditional probability $P(3|4)$ was estimated as 0.33, within the range of 0.11 to 0.60 computed from Equation 12.2. Therefore, Bayes's rule can be applied to give a value of $P(4|3) = [P(3|4)/P(3)]P(4) = 0.25$. Table 12.1 indicates, however, that we apparently had a strong rationale to support a lower estimate, 0.05.

Just as the occurrence of an event can affect the probability that another will occur, its nonoccurrence can have an impact as well. In our fax example, for instance, if increased taxes or costs of transmission fail to materialize, then the impetus for and probability of replacement technologies will decrease. Thus we need to construct a nonoccurrence matrix (see Table 12.2).

Our last step is to estimate the entries for the nonoccurrence matrix, using the same philosophy as we did for the occurrence matrix. First we will compute the entries statistically from the following equation:

$$P(i|\bar{j}) = [P(i) - P(j)P(i|j)]/[1 - P(j)] \quad (12.4)$$

Lacking evidence to the contrary, these values will be entered. However,

TABLE 12.2 Nonoccurrence Matrix for a FAX Example

The Probability of This Event Becomes:	If This Event Does Not Occur:			
	Increased Taxes/Cost	Negative Legislation	Replacement Technology	Market Saturation
Increased taxes or costs on transmission (0.30)*	0.00 $P(1 \bar{1})$	0.85 $P(1 \bar{2})$	0.60 $P(1 \bar{3})$	0.90 $P(1 \bar{4})$
Negative legislation (0.60)*	0.87 $P(2 \bar{1})$	0.00 $P(2 \bar{2})$	0.35 $P(2 \bar{3})$	0.40 $P(2 \bar{4})$
Development of a replacement technology (0.50)*	0.48 $P(3 \bar{1})$	0.52 $P(3 \bar{2})$	0.00 $P(3 \bar{3})$	0.78 $P(3 \bar{4})$
Market saturation (0.55)*	0.73 $P(4 \bar{1})$	0.56 $P(4 \bar{2})$	0.75 $P(4 \bar{3})$	0.00 $P(4 \bar{4})$

* Initial marginal (ceteris paribus) probability of nonoccurrence: $P(\bar{i}) = 1 - P(i)$

if evidence supports a different estimate, that estimate will be entered instead. Returning to the example

$$P(2|\bar{1}) = [P(2) - P(1)P(2|1)]/[1 - P(1)] = 0.87$$

If we have no reason to estimate some other probability, then 0.87 should be entered into the nonoccurrence matrix.

Note that the diagonal entries in the nonoccurrence matrix will all be 0.00 since the probability of an event given that it has not occurred is 0. Negative probabilities predicted by Equation 12.4 should be set at 0, while predicted probabilities greater than 1 (certainty) should be set to 1.

The next stage in CI analysis is to simulate the effects of these conditional relationships. We must determine whether the initial estimates of event marginal probabilities are mutually consistent given these perceptions of how events impact each other.

If all the entries in the two matrices agree with results computed from Equations 12.1 through 12.4, then the initial marginal and conditional probabilities are mutually consistent. However, if one or more of the conditional probabilities differ from computed results, we will have to "play" the CI matrices to determine a consistent set of marginal probabilities. A computer-based Monte Carlo simulation can be used to do this:

1. An event is selected randomly (say Event 2 in Table 12.1).
2. A random number between 0 and 1 is generated and compared to the marginal probability of the event to determine if it occurs. Suppose the random number is 0.26, since $0.26 \leq 0.40$, Event 2 is assumed to occur. If the random

number were greater than 0.40, it would be assumed that Event 2 did not occur.

3. The marginal probability of each remaining event is replaced by its conditional probability given that the event in Step 2 occurs or does not occur. That is, in our example $P(i)$ is replaced by $P(i|2)$ if Event 2 occurs, or by $P(i|\bar{2})$ if it does not ($i \neq 2$). Thus, since Event 2 occurred in Step 2, the replacement values will be $P(1) = 0.40$, $P(3) = 0.72$, $P(4) = 0.35$.
4. A second event is selected randomly from those remaining (Events 1, 3, and 4), and Steps 1 through 3 are repeated. In this play, the probability used in Step 2 is the value produced in Step 3 of the previous play. Thus, if Event 2 occurred in the first play and Event 4 is selected in the second, the probability of Event 4 used in Step 2 of the second play is $P(4|2) = 0.35$.
5. The process described in Steps 1 through 4 is repeated until all four events have been selected. All marginal probabilities are then returned to their initial values and the game is "replayed," typically 1,000 or more times.
6. Each time the game is "played" the events that occur are noted. The total number of occurrences divided by the number of games is taken as the final (marginal) probability for each event. The initial marginal probabilities are then replaced by the final marginal probabilities, which account for event interaction.

The conditional probability of an event occurring given two or more other events is required in Step 3 after the first two events have been determined. For instance, the examples woven into these six steps suggested that Event 2 occurred in the first round of play. If in the second round Event 4 is found to occur, then we need conditional probabilities such as $P(1|2 \cap 4)$ to proceed. The occurrence and nonoccurrence matrices only specify pairwise interactions—that is, joint probabilities of one event conditioned on one other event. Joint conditionals are called *second-order conditionals*. There also are third-order and higher-order conditionals, for example, $P(i|j \cap k)$ and $P(i|j \cap k \cap l)$. These probabilities are difficult to determine. Instead, they usually are approximated by averaging second-order probabilities. For instance,

$$P(1|2 \cap 3) = [P(1|2) + P(1|3)]/2$$

A similar averaging procedure is used for higher-order, nonoccurrence probabilities, and probabilities such as

$$P(1|2 \cap \bar{3}) = [P(1|2) + P(1|\bar{3})]/2$$

These approximations are acceptable when the conditional probabilities being averaged are close in value.

Software in the Technology Forecasting (TF) TOOLKIT, which supplements this book, can be used to play this CI game. Of course, the game is probabilistic and so results vary somewhat from game to game. However, after 1,000 plays this program

produces final marginal probabilities of $P(1)' = 0.60$, $P(2)' = 0.50$, $P(3)' = 0.51$, and $P(4)' = 0.51$ for our fax example, where the prime symbol has been added to distinguish the final marginal probabilities from the initial values. Note that there have been changes in the marginal probabilities of all events. After 10,000 plays, the resulting marginal probabilities are 0.61, 0.50, 0.51 and 0.50 for Events 1 through 4, respectively. If Equations 12.1 through 12.4 were used to generate all the conditional probabilities in the two matrices, we likely would have seen little difference between initial and final marginal probabilities. This is because using those equations tacitly assumes that the conditional probabilities are consistent with the initial marginal probability estimates.

The CI game is attractive since it can be generated from relatively little data. It can be used to examine the interaction of events and to ensure, insofar as possible, that the probabilities to be used in, say a scenario, account for those interactions. The technology manager also could use the CI matrices to plan strategies to enhance or inhibit the probability that key events occur. Moreover, constructing the CI matrices can provide a useful format within which to frame discussions of interactions between events. Equally important, merely constructing the matrices forces the manager or forecaster to consider interactions between events. However, it is important to note that, even though the probabilities produced by CI are consistent with our perceptions of how they interact, they still may be incorrect.

Halverson, Swain, and Porter (1989) point out that the effect of an event on the probabilities of other events is not necessarily determined by the size of its marginal probability. This stems from the somewhat trivial observation that an event with a low probability of occurrence has a high probability of nonoccurrence. Since the nonoccurrence of an event can have a major effect on other events, the effect of an unlikely event thus may be quite high.

Halverson, Swain, and Porter (1989) also observe that the conditional probability estimates may be more important than the marginal probability estimates. The initial marginal probability estimates likely will be changed by the conditional estimates during the game. Therefore, we should concentrate on accurately estimating the impact of events on each other (that is, estimating the conditional probability) and expend less effort on estimating the marginal probabilities.

Despite its advantages, classic CI has a number of deficiencies. These primarily stem from the fact that it is a discrete model with no time dimension; thus it cannot deal with cases in which the time dimension is important. For instance, in the fax example, suppose that one event cannot occur unless another has occurred and that the second event cannot occur earlier than three years from today. CI, as presented, is powerless to portray this situation. Nor can it portray trends since they, by definition, are time dependent. Modifications of classic CI designed to address these limitations are described in the following subsections.

12.2.2 Time-Dependent CI

As noted previously, the traditional CI approach includes no dynamic (time) dimension. The probabilities of occurrence and nonoccurrence are assumed to be

valid for all time. However, it is possible, even likely, that the probability of most events will change as time passes. For instance, in the fax example, Event 3 (the development of a replacement technology) likely will become more probable as time passes. Further, while Event 1 (increased taxes or costs of transmission) could occur at any time, its effect on Event 3 may be cumulative over time—that is, as the costs grow, other means of communication may become more and more desirable. Rather than state the CI game in terms of $P(i : j)$, we could use time-dependent probabilities such as $P(E_i^{t+1} : E_j^t)$. This notation indicates both time-dependence and ordering by specifying the probability of Event i at time $t + 1$, given that Event j has already occurred by time t .

The fax example model can be reformulated as a Markov process, which will allow us to answer questions about the probabilities of events over time, the sequences of events over time, and the expected times of occurrences of events. This also will allow us to model how the effects on other events vary with time.

A Markov process is a system of states governed by a matrix of *transition probabilities*.¹ These are the probabilities that the system will go from any state $N_t = i$ to some other state $N_{t+1} = j$ during the next step in time. It is assumed that these transition probabilities only depend on the current state—that is, the Markov process has no memory. For n events there are 2^n states which define a $2^n \times 2^n$ Markov transition matrix. Each of the numbered states represents a distinct scenario defining the occurrence or nonoccurrence of the n events. Suppose that there are three events. We will adopt a shorthand notation to indicate any scenario, call it N . If the scenario, N , is $(1\bar{2}\bar{3})$, then Event 1 occurred, but Events 2 and 3 did not. We will write this as (100) . The order of the three digits represents Events 1, 2, and 3; a 1 indicates occurrence, a 0 indicates nonoccurrence.

Because there are three events, ($2^3 = 8$), the 8×8 transition matrix shown by Table 12.3 will be needed to order this work. In that matrix, the probability of moving from one state at time t to another state at time $t + 1$ (that is, the transition probabilities) are displayed in the cells. For instance, the probability of going from state 010 (state number $N = 3$ in the matrix) at time t , to state 110 ($N = 7$) at time $t + 1$ is given by $p(3:7)$, which is recorded in the cell formed by row 3 and column 7. In general notation, the probability of transition from state i to state j in a one-unit passage of time is $p(i : j)$, recorded in cell (i, j) .

Since each row in the array contains the probabilities of going from state i to each of the eight possible states in the next step, the sum of the probabilities in each row must be unity (that is, one or the other of the states must occur at time $t + 1$). Note that diagonal entries are the probabilities that the state will not change. If these probabilities are large (close to 1.0) the system will change slowly. The probabilities below the diagonal may be set to 0 if events cannot be undone (cannot be reversed).

A Markov matrix requires a great deal of data. A matrix for n events will require that we estimate 2^{2n} transition probabilities. Even if entries below the diagonal can be set to 0, we will need $(2^n - 1)2^n/2$ probability estimates. For example, a six-event model would require a minimum of $(2^6 - 1)2^6/2$ or 2,016 entries. Such

¹This discussion draws heavily on Swain et al. (1989).

TABLE 12.3 Markov Transition Matrix for Three Events

Transition From This State at Time t	To This State at Time $t + 1$							
	(000) $N = 1$	(001) $N = 2$	(010) $N = 3$	(011) $N = 4$	(100) $N = 5$	(101) $N = 6$	(110) $N = 7$	(111) $N = 8$
(000) $N = 1$	$P(1:1)$	$P(1:2)$	$P(1:3)$	$P(1:4)$	$P(1:5)$	$P(1:6)$	$P(1:7)$	$P(1:8)$
(001) $N = 2$	$P(2:1)$	$P(2:2)$	$P(2:3)$	$P(2:4)$	$P(2:5)$	$P(2:6)$	$P(2:7)$	$P(2:8)$
(010) $N = 3$	$P(3:1)$	$P(3:2)$	$P(3:3)$	$P(3:4)$	$P(3:5)$	$P(3:6)$	$P(3:7)$	$P(3:8)$
(011) $N = 4$	$P(4:1)$	$P(4:2)$	$P(4:3)$	$P(4:4)$	$P(4:5)$	$P(4:6)$	$P(4:7)$	$P(4:8)$
(100) $N = 5$	$P(5:1)$	$P(5:2)$	$P(5:3)$	$P(5:4)$	$P(5:5)$	$P(5:6)$	$P(5:7)$	$P(5:8)$
(101) $N = 6$	$P(6:1)$	$P(6:2)$	$P(6:3)$	$P(6:4)$	$P(6:5)$	$P(6:6)$	$P(6:7)$	$P(6:8)$
(110) $N = 7$	$P(7:1)$	$P(7:2)$	$P(7:3)$	$P(7:4)$	$P(7:5)$	$P(7:6)$	$P(7:7)$	$P(7:8)$
(111) $N = 8$	$P(8:1)$	$P(8:2)$	$P(8:3)$	$P(8:4)$	$P(8:5)$	$P(8:6)$	$P(8:7)$	$P(8:8)$

data requirements greatly exceed those of the traditional CI game and greatly diminish the number of practical applications for this approach. However, Swain et al. (1989) present a scheme for estimating these probabilities with data requirements more nearly equal to those of the traditional CI approach. For example, consider a Markov matrix for a two-event model with the following entries:

$$P^{(t,1)} = \begin{bmatrix} 0.550 & 0.300 & 0.098 & 0.052 \\ 0.000 & 0.850 & 0.000 & 0.150 \\ 0.000 & 0.000 & 0.650 & 0.350 \\ 0.000 & 0.000 & 0.000 & 1.000 \end{bmatrix}$$

The same scheme used to order the larger matrix in Table 12.3 has been used here so that the first row and column represent the state 00 (or $N = 1$), the second row and column represent state 01 ($N = 2$) and so forth.

For convenience we will use $P^{(t,1)}$ to represent the entire matrix. The first superscript indicates that the matrix gives the transition probabilities for changes starting at time t ; the second superscript indicates that it applies for one time step (from t to $t + 1$). In this example, entries below the diagonal have been set to 0

and the transition probabilities in each row sum to 1. Note also that $P(4 : 4) = 1.0$. If the system reaches this state, it will not change in subsequent steps. This is called the *trapping state*. Depending on the situation being modeled, there may be one trapping state, more than one, or none at all.

One problem noted with the classic CI is that state probabilities are assigned a single (marginal) probability that does not change with time, which is an unrealistic formulation. However, the transition matrices allow us to represent such variations. Suppose the probabilities of the four states in our sample matrix are estimated to be 0.7, 0.3, 0.4, and 0.1 at time t . For convenience, we will represent this as the row matrix

$$S^t = [0.7, 0.3, 0.4, 0.1]$$

The transition matrix allows us to compute the state probabilities at $t + 1$ as

$$S^{t+1} = S^t P^{(t,1)} \quad (12.5)$$

or

$$S^{t+1} = [0.7, 0.3, 0.4, 0.1] \begin{bmatrix} 0.550 & 0.300 & 0.098 & 0.052 \\ 0.000 & 0.850 & 0.000 & 0.150 \\ 0.000 & 0.000 & 0.650 & 0.350 \\ 0.000 & 0.000 & 0.000 & 1.000 \end{bmatrix}$$

so that

$$S^{t+1} = [0.385, 0.465, 0.329, 0.321]$$

By using the transition matrices and the initial probabilities, we can compute the changes in state probabilities sequentially over the time period of interest.

To time market entry and/or resource allocations, we often are interested in knowing when a state might be reached. The time it takes to first reach a particular state is called the *first passage time*; the associated probability is called the *first passage probability*. We also may be interested in knowing how the cumulative probability of occurrence of a particular state may change over time. Both can be investigated using the time-dependent CI approach as described in the following paragraphs.

The probability that the system will make a transition from any state i at time t to any state j sometime during n steps is given by

$$P^{(t,n)} = P^{(t,1)} P^{(t+1,1)} P^{(t+2,1)} \dots P^{(t+n-1,1)} \quad (12.6)$$

For simplicity, we will examine the special case in which the transition matrices do not change from step to step (that is, the entries in the matrix do not depend on t).

$$P^{(t,n)} = [P^{(t,1)}]^n$$

The n -step transition matrix is obtained by multiplying $P^{(t,1)}$ by itself n times. For example, to compute $P^{(t,2)}$, the matrix $P^{(t,1)}$ would be multiplied by itself to produce

$$P^{(t,2)} = \begin{bmatrix} 0.302 & 0.420 & 0.118 & 0.160 \\ 0.000 & 0.722 & 0.000 & 0.278 \\ 0.000 & 0.000 & 0.422 & 0.578 \\ 0.000 & 0.000 & 0.000 & 1.000 \end{bmatrix}$$

Thus the probability of remaining in state 1 for the period from t to $t+2$ is 0.302, and the probability of going from State 1 to State 2 sometime in that period is 0.420. Note that the probability of staying in state 4 (the trapping state) is 1.0. Therefore, it is impossible to retreat to a lower numbered step from this state.

Following the process outlined above, $P^{(t,n)}$ for this example is given by

$$P^{(t,n)} = \begin{bmatrix} [0.55^n] & [0.85^n - 0.55^n] & [0.98(0.65^n - 0.55^n)] & [1 - 0.85^n - 0.98(0.65^n - 0.55^n)] \\ 0 & [0.85^n] & 0 & [1 - 0.85^n] \\ 0 & 0 & [0.65^n] & [1 - 0.65^n] \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

We can use this matrix to study various transitions of interest. For instance, we may wish to know how long it might take to go from a base state ($N = 1$) where neither event has occurred (state 00) to a state ($N = 4$) where both events have

TABLE 12.4 $P(1:4)$ and $f(1:4)$ for Selected n

n	$P(1:4)$	$f(1:4)$
0	0.000	-
1	0.052	0.052
2	0.160	0.108
3	0.280	0.120
4	0.393	0.113
5	0.492	0.099
...	...	...
19	0.954	0.008
20	0.961	0.007

Source: Swain et al., 1989. Reprinted by permission

occurred (state 11). The probability that we have arrived at state 4 at some time during the first n steps is represented by

$$p(1:4) = 1 - 0.85^n - 0.98(0.65^n - 0.55^n)$$

Solving this expression at various values of n produces the results shown in Table 12.4. Also displayed in that table are the incremental changes in probability from step to step, $f(1:4)$, which indicate the probability of reaching state 4 *in that step*. The table shows, for instance, that the probability of having reached state 4 *by* the fifth step is nearly 0.500, while the probability of reaching state 4 *in* the third step is 0.120. The table indicates that the most likely first passage time is three periods since $f(1:4)$ has the highest first passage probability, 0.120. The mean first passage time (that is, the mean number of steps to reach state 4 from state 1) would be $1(0.052) + 2(0.108) + \dots = 7.28$ periods.

The time-dependent CI approach adds considerable power to cross-impact studies. Using it, we can more realistically portray the variation of probabilities and impacts with time and handle the sequencing of event probabilities representative of real-world situations. The price we pay for this extension comes from the increased data requirements and manipulation complexity. The classic and time-dependent CI approaches are compared for a specific case in Exhibit 12.1.

■ Exhibit 12.1 Classic and Time-Dependent CI Compared—The AIMTECH Case

AIMTECH is the U.S. Air Force project acronym for a forecast of the development of artificial intelligence (AI) technology. Although the AIMTECH study was much broader, we will focus on a subset intended to augment pilot decision making in combat situations. An AI system capable of planning against actions by an adversary requires development of two other technological capabilities. Thus the subset of (technological) events that concern us here are

ANA—An AI system capable of reasoning by analogy

COM—An AI system embodying common sense

ADV—An AI system capable of planning action against an adversary

The bottom line of the AIMTECH study was to determine which AI targets were most attractive for R&D development.

The AIMTECH forecast (Reitman et al., 1985) was used by Swain et al. (1989) as an example of the Markov-based CI approach. We rely heavily on the Swain et al. (1989) study to compare and contrast analyses by the classic and time-dependent CI approaches. Data for this example are only illustrative and, in some cases, result from assumptions beyond the study's scope.

In classic CI, we would develop initial marginal probabilities for each event and conditional probabilities to display the effects of their occurrence or nonoccurrence on each other. Such a set of matrices is shown in Table 12.5.

These matrices represent very limited CI. For example, note that COM's occurrence does not change the initial marginal probability of ANA and vice versa. Thus ANA and COM are independent of each other. Likewise, the nonoccurrence of ADV does not affect either ANA or COM. Further, the occurrence of ADV implies the occurrence of both ANA and COM, and ADV cannot occur unless both ANA and COM do. Following the procedure outlined in Section 12.2.1, the CI matrices in Table 12.5 can be played using the Monte Carlo technique. After 10,000 plays the final marginal probabilities were found to be 0.59, 0.60, and 0.77 for ANA, COM, and ADV, respectively.

The problem with the classic CI solution is that since the model is entirely independent of time so are the results. Thus we cannot take into account that certain events must precede others. We have no way of knowing when any might occur; therefore, we have no way to determine if increased R&D funding, for instance, would cause an ANA or COM to be realized significantly earlier. Nor has the classic model allowed us to portray impacts between events that grow or decay over time. This is unrealistic. Such effects are apparent when we view the probabilities estimated for the three variables over time by the AIMTECH study (see Figure 12.1). For example, as Figure 12.1 shows, ANA cannot occur before 1986 or after 1994 and COM must occur in the window from 1993 to 1997. This time dimension is lost in classic CI. These are precisely the kind of dynamics for which time-dependent CI is intended.

TABLE 12.5 AIMTECH Cross-Impact Matrices

The Probability of Event Becomes:	If This Event Occurs			If This Event Does Not Occur		
	ANA	COM	ADV	ANA	COM	ADV
ANA (0.85)*	1.0	0.85	1.0	0.0	0.85	0.85
COM (0.65)*	0.65	1.0	1.0	0.65	0.0	0.65
ADV (0.37)*	0.44	0.47	1.0	0.0	0.0	0.0

Source: Swain et al., 1989 Reprinted by permission.

* Initial marginal probability.

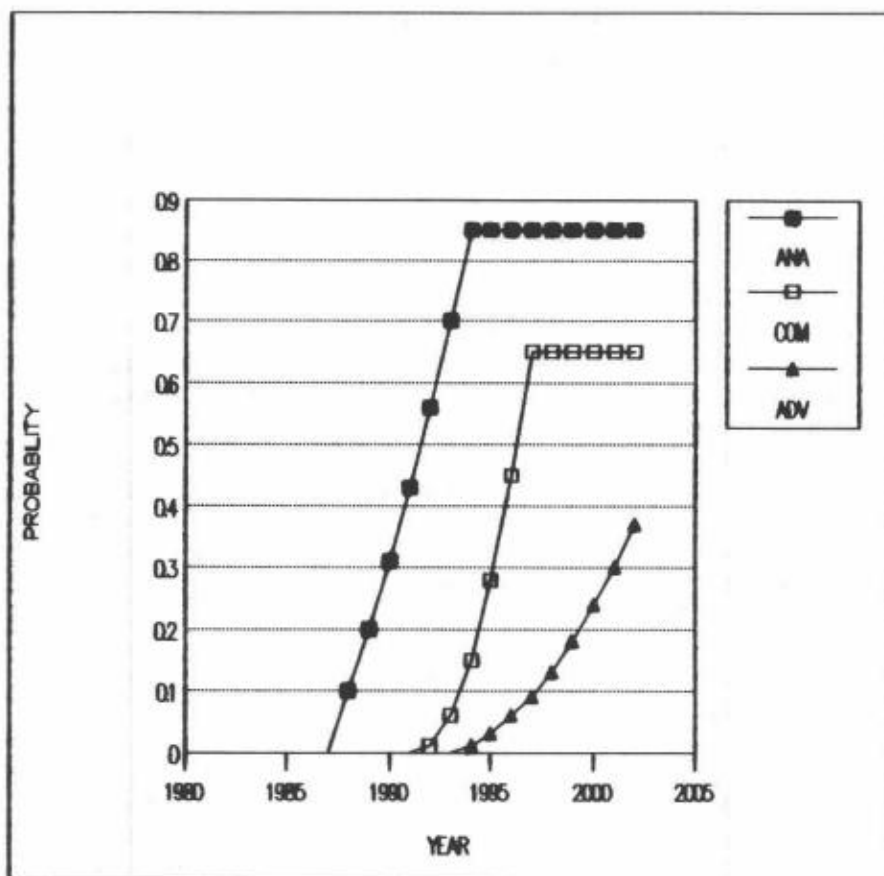


Figure 12.1 Cumulative probabilities forecast for AIMTECH. (Source: Swain et al., 1989. Reprinted by permission.)

Swain et al. (1989) have estimated the transition matrix data from information given in the AIMTECH study. Two of the transition matrices are given in Table 12.6. In studying that table, recall that Event 3, ADV, can only occur after Events 1 and 2, and no events un-occur. Because of the ADV constraint, only five of the possible eight theoretical states are modeled: the remaining three cannot occur. Ordering and numbering of the matrices in Table 12.6 follow the scheme used in Table 12.3. It is important to note from the table that, unlike the example given in the last section, the transition matrices are different for different times. Therefore, we must use Equation 12.6 to compute the n -step transition matrix.

Swain et al. (1989) estimate the state probabilities in 1993 to be $S^{1993} = (0.26, 0.04, 0.68, 0.02, 0)$ —that is, the probability of being in state 1 in 1993 is 0.26, and so forth. These probabilities, of course, are derived from those shown in Figure 12.1. For example, the probability of ANA having been reached by 1993 is 0.70 in that figure. In the example then, the probabilities for state 5 (0.68)

TABLE 12.6 Markov Transition Matrices, $p^{(1993,1)}$ and $p^{(1996,1)}$ for AIMTECH

$p^{(1993,1)}$	Transition from This State at Time t	To This State at Time $t + 1$				
		(000)* $N = 1$	(010) $N = 3$	(100) $N = 5$	(110) $N = 7$	(111) $N = 8$
	(000)* $N = 1$	0.2692	0.2154	0.5154	0	0
	(010) $N = 3$	0	0.6000	0	0.4000	0
	(100) $N = 5$	0	0	0.9500	0.0500	0
	(110) $N = 7$	0	0	0	0.5000	0.5000
	(111) $N = 8$	0	0	0	0	1
$p^{(1996,1)}$						
	(000)* $N = 1$	0.5000	0.5000	0	0	0
	(010) $N = 3$	0	1	0	0	0
	(100) $N = 5$	0	0	0.6415	0.3585	0
	(110) $N = 7$	0	0	0	0.8846	0.1154
	(111) $N = 8$	0	0	0	0	1

Source: Swain et al., 1989. Reprinted by permission.

* Order of events = (ANA, COM, ADV)

and state 7 (0.02) must total 0.70. Likewise, since the probability of achieving COM by 1993 is 0.06, the sum of the probabilities for states 2 and 7 also must be 0.06. As outlined in the previous section, the initial state probabilities and the transition matrices can be used to "bootstrap" our way through the time horizon, calculating state probabilities as we go. For instance, using Equation 12.5,

$$S^{1994} = S^{1993} P^{(1993,1)} = (0.07, 0.08, 0.78, 0.06, 0.01),$$

while S^{2002} can be shown to be (0.01, 0.14, 0.34, 0.14, 0.37). Thus the probability of reaching any state is dynamic, as shown by Figure 12.2.

In contrast to the classic approach, the time-dependent CI provides information about the process throughout the time span. A number of analyses can be performed. For example, using Equation 12.6 the nine-step transition matrix $P^{(1993,9)}$ can be obtained. We can examine the last column of this matrix to find the probability of attaining ADV by 2002, given the various initial states in 1993. These probabilities are 0.2041, 0.4184, 0.3377, 0.9221, and 1.0 for beginning states 1, 3, 5, 7, and 8, respectively. Thus, if we start in state 1 (000) in 1993, there is a probability of about 0.20 of attaining ADV by 2002; if we start in state 3 (010), there is a probability of about 0.42. This is contrasted with the probability of achieving ADV in 2002 of 0.37, which is averaged across all starting states. As outlined in the previous section, this analysis could be extended to probe the most likely or the mean first passage times to ADV.

In this section, the Markov-based, time-dependent CI analysis has been contrasted with classic CI. The classic approach provides the marginal probabilities for ANA, COM, and ADV in 2002. The Markov approach provides information

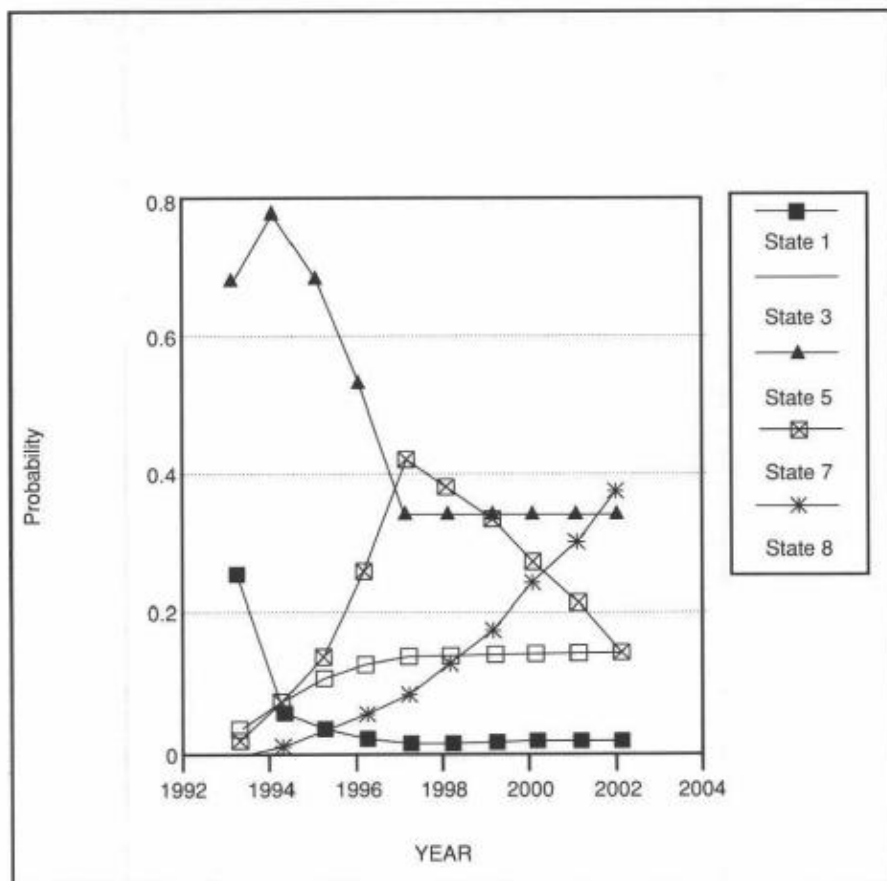


Figure 12.2 Probabilities of attaining states. (Source: Swain et al., 1989. Reprinted by permission.)

throughout the process and a means of dealing with impacts that vary with time—a major improvement. However, expansion of CI to this dynamic dimension could only be made at the expense of developing nine transition matrices spanning the period from 1993 to 2001—each one of which requires at least as much input as that provided for the classic CI process. Clearly, the increased resources implied are not appropriate to every topic. ■

12.2.3 Other Time-Dependent CI Schemes—XimpacT

Enzer and Leschinsky (1986) have extended classic CI to approximate time steps, although the causality implied is not explicitly modeled. This approach is incorporated in the XimpacT CI software package.² Inputs to XimpacT include a set of

²Distributed by XimpacT, contact Selwyn Enzer, IBEAR Program, University of Southern California, Los Angeles, CA 90007.

trends (including their nominal forecasts), a set of events (including their probabilities), and cross-impact factors. Outputs are termed *alternative futures* (or scenarios) to acknowledge the uncertainty or probabilistic nature of the problem. Their purpose is to encourage communication about alternatives among policy analysts and decision makers. In addition, following the scenarios year by year can help decision makers to understand the underlying dynamics.

Output information about scenarios is available both in tabular and graphical form. Tabular output includes summary reports about the scenarios and statistical reports about trends. The summary reports indicate which events occur and their year of occurrence as well as year-by-year values of the simulated trend. Statistical reports for each trend indicate the maximum, minimum, average, and standard deviation over all scenarios for each year. Graphical output includes the same types of data with some flexibility for the user in choosing data to be plotted.

12.2.4 Closure

While the discussion of CI presented in the preceding sections is extensive, the manager should by no means mistake it for the last word. Rather, it provides the background that he or she will need to understand ongoing CI development as well as one or two extensions of the classic approach. The manager is advised to refer to Sage (1990) and Martino (in press) for further information.

Sage (1990) points out four problems in CI analysis. First, there is no guarantee that those filling the matrix of CI relationships will not violate Bayes' rule. Second, the relationship between $P(i:j)$ and $P(i)$ should at least be symmetrical, but even this may not happen. While revised approaches using likelihood ratios instead of the simple enhancing/inhibiting/independent schemes can provide symmetry, they do not ensure that the probabilities so generated will be consistent with Bayesian probability theory. Third, if the occurrence of an event affects the probability of another event, then the nonoccurrence of that event should also affect the probability of the second event. However, the simplest approaches to CI assume that nonoccurrence has no impact. Finally, Sage notes that the time element is important, but CI often does not include a mechanism for handling the timing of events.

The following eight-step procedure for CI analysis is recommended:

1. Use collective inquiry methods to generate a set of appropriate events, elements and relations for a particular problem.
2. Use a structural modelling method to evolve a structure for the events under consideration.
3. Develop a scale so that one may translate expert opinion and knowledge concerning the likelihood of occurrence of events into probabilities. This scale should be discussed, understood and approved by the group supplying the probability estimates.
4. Obtain an initial set of probabilities of the type in which estimators have most confidence. Every effort should be made to obtain as many of the direct tree-related conditional probabilities as possible. For direct tree-related conditional probabilities

not estimated, obtain sufficient marginal and lower-order conditional probabilities, and impact assessments, to bound the direct tree-related conditional probabilities. These upper and lower bounds are computed.

5. Display the set of direct tree-related conditional probabilities or bounds computed on these probabilities if they are not directly estimated. Obtain final probabilities with which to determine future probabilities.

6. Determine future event probabilities.

7. Incorporate the results of the cross-impact analysis into a system simulation model of realities in which events are assumed to occur or not occur as predicted from the results of the cross-impact analysis. Inputs or policies connected with the events must be determined in order to accomplish this simulation. Iterate these simulation runs for other likely futures.

8. Incorporate the results of the cross-impact analysis and system simulation exercises as the systems analysis phase of systems engineering activities for the particular problem under consideration. (Sage, 1990, p. 262-263)

Sage then emphasizes the complexities of adjusting the analysis when new information is presented. The basic approach is Bayesian in that the prior probabilities are adjusted for posterior observations or perceptions stated as observations. However, human beings are not particularly good at adjusting hierarchical inference (that is, logical interconnections of elements in a net or tree). Sage shows how quantitative procedures based on a few assumptions and using the computational power of computers might improve the handling of new information for complex systems developed in the CI analysis. Developing an approach based upon hypothesis evaluation, he suggests that inference can be done on the basis of imprecise knowledge that can be either hypothetical or evidential.

Joseph Martino has developed an approach to handle both the probability and timing of events in the CI format (Martino, pending). He uses the timing of events to reduce the number of cross-impacts that must be considered. For example, if Event 1 must precede Event 2, the occurrence or nonoccurrence of the second event cannot affect the probability of the first. Martino also notes that implicit assumptions the manager makes about the occurrence of earlier events affect the estimates he or she makes about later ones. Thus some impacts are implicit in the estimates themselves and should not be accounted for a second time by CI. For instance, suppose we estimate a 70% probability that the New York Yankees will win the 1995 World Series. This estimate clearly assumes that they will win the American League pennant that year. If they do, it has no impact on winning the series because that has been assumed. We need only estimate the impact of not winning the pennant on the probability of winning the series. Combining timing and structural factors such as those outlined here can significantly decrease the amount of information that must be developed and improve the validity of a CI analysis.

The CI approach described by Martino recognizes both events that must occur (or not) at a specific time and those that can occur with varying probabilities over a span of years. It also accommodates sets of events that are yoked together in

such a way that one, but not more than one, event in the set may occur. Finally, the approach can be used to identify events that are especially pivotal or especially sensitive to the occurrence of other events.

12.3 KSIM

KSIM is a deterministic simulation model developed by Julius Kane (1972). KSIM extends the concepts of CI to produce a dynamic simulation that is easy to use yet sufficiently powerful to provide meaningful analysis of many real-world problems. The model retains the concept of the impacts of events on each other characteristic of CI. However, this concept is married to a differential equation that portrays an S-shaped (logistic) growth or decline of the variables being modeled. This equation provides the continuous, dynamic (time-dependent) characteristics of KSIM. The logistic variation is a loose analogy to biological system growth (the support and rationale for this analogy are presented in Chapters 9 and 10). Since impact magnitudes are estimated subjectively, KSIM in effect utilizes both "hard" (objective) and "soft" (subjective) input. Thus it is an appropriate implementation of Kane's premise that experience, opinions, and judgments control decision-making.

The variables modeled by KSIM, X_i , are first identified, defined, and quantified. The maximum value of each variable is determined so that each can be normalized on a scale of 0 to 1. The initial value of each is also estimated. The simulation marches forward from these initial values a step at a time using the differential equation

$$\frac{dX_i}{dt} = \sum_{j=1}^N (\alpha_{ij} + b_{ij} \frac{dX_j}{dt}) X_i \ln X_i \quad (12.7)$$

where

- X_i = the variable described
- N = the total number of variables considered
- X_j = the impacting variables
- α_{ij} = the long-term impact of X_j on X_i
- b_{ij} = the short-term impact of X_j on X_i

The solution to this logistic CI equation is

$$X_i(t + \Delta t) = X_i(t)^{P_i(t)} \quad (12.8)$$

where

- $X_i(t + \Delta t)$ = value of variable at end of the time period
- $X_i(t)$ = value of X_i at the start of time period
- Δt = the time period

and

$$P_i(t) = \frac{1 + \Delta t(\text{sum of inhibiting impacts on } X_i)}{1 + \Delta t(\text{sum of enhancing impacts on } X_i)}$$

or, mathematically,

$$P_i(t) = \frac{1 + 0.5t \sum_{j=1}^N [|I_{ij}(t)| - I_{ij}(t)]X_j(t)}{1 + 0.5t \sum_{j=1}^N [|I_{ij}(t)| + I_{ij}(t)]X_j(t)} \quad (12.9)$$

and

$$I_{ij} = \alpha_{ij} + \frac{b_{ij}}{X_j(t)} [dX_j(t)/dt]$$

While the equations appear formidable, operationally the concept is relatively straightforward. One must estimate the impacts of the level of each event (that is, level = value of X_j) on all other variables. This is the α_{ij} , which is determined in much the same manner as impacts in CI. Then, the impacts of the rates of change of each event (dX_j/dt) and the slope of the *trend* in X_j on the other events (b_{ij}) are estimated in the same fashion. Once these impact magnitudes have been determined, a relatively simple computer program (such as that included in the TOOLKIT) can be used to solve the equations and perform the forecast.

The characteristics of KSIM are pretty much what would be expected of a logistic curve. For example, when the sum of the inhibiting impacts is greater than that of the enhancing impacts, the power $P_i(t)$ in Equation 12.8 will be larger than one. And, since $0 \leq X_i(t) \leq 1.0$, $X_i(t + \Delta t)$ will be smaller than $X_i(t)$. Further, all else being equal, the larger the variable causing the impact, the greater the magnitude of that impact will be. Note also that a given value of $P_i(t)$ will have less effect on the magnitude of X_i if X_i is near either 0 or 1. This produces the S-shaped variation we expect of growth or logistic curves.

KSIM is one of the few dynamic models that can be constructed and used with relatively limited time and resources. The general procedure that a group of technology forecasters or managers would use is as follows:

1. Discuss the problem and agree on the scope and boundaries of the simulation (such as level of aggregation, spatial boundaries, and time frame).
2. Identify, define, and label the important variables and determine their initial values, ranges, and maximums. Normalize each variable on a 0 to 1 range.
3. Structure the long-term and short-term impact magnitudes and array them in matrix form as in CI. Impacts that increase the size of a variable (enhance it) are positive, those that inhibit it are negative. Numerical values for mag-

nitudes are proportional to the size of the impact. For example, if X_1 is not impacted by X_2 , then α_{12} and/or b_{12} will be 0. This work sometimes is cut in half by considering only short-term (b_{ij}) or long-term (α_{ij}) impacts.

4. Run the model and refine the impacts, variable definitions and/or values until the outcome is satisfactory. Usually a base case is run and the output is compared to a similar situation or to theoretical behavior. The process is repeated as often as necessary to produce acceptable results.

The model can now be run to investigate the effects of changing initial values or basic assumptions or of introducing new assumptions. In this way, alternative futures can be examined and forecasts and trade-offs can be determined. KSIM also provides for a very useful extension, allowing external events or policy decisions to be added to the model as variables in the CI matrix. This is done by formulating the impacts of, say policy options, as additional columns but not rows in the CI matrices. Thus a decision to invoke a policy option impacts the variables, although the option is not itself impacted by the variables. Using this approach, the decision maker can systematically investigate the effects of policy decisions on the behavior of the system.

KSIM should be viewed as a process as well as a product. The benefit of KSIM accrues from building the model, as much as from operating it and analyzing the results. Building the model provides the format in which a team can structure the discussion of a complex issue. In that format, experience, opinion, and judgment can be incorporated along with hard data. Further, a completed model allows alternatives to be quickly formulated and their consequences to be assessed. Thus KSIM can provide an environment within which the manager can study and learn about complex situations.

The process makes a number of assumptions that imply limitations as well. KSIM assumes that a satisfactory model can be devised and that the variables and their interactions can be accurately defined. It also assumes that realistic bounds can be placed on the variables; that a growth curve adequately represents the change patterns being studied; and that opinions, experience, and other subjective information can be formulated mathematically. Equally important, KSIM assumes that the pairwise relationships portrayed by the matrices adequately represent true causal interaction, a much more complex situation. Finally, KSIM models a deterministic world; however, the technology manager, the forecaster, and the rest of us live in a probabilistic one. For an example of how to apply the KSIM model, see Exhibit 12.2.

■ Exhibit 12.2 KSIM Applied to Model Fax Transmission

To clarify concepts, consider the example of transmission used earlier. We will examine four variables:

1. Number of fax machines (irrespective of sophistication), N

2. Median cost of fax machines purchased, C
3. Number of pages of fax transmission, T
4. Cost/transmission (regardless of length), $\$$

We also assume a single policy option—taxing fax transmissions. Suppose that impact magnitudes were estimated on a scale from 0 (no impact) to ± 3 (major impact) and that we have defined a major impact for each variable as one that causes a 10 percent change in the level of the variable. The short-term and long-term impact matrices that are estimated appear in Table 12.7. Note that these impacts are merely presented as representative for the purposes of this example; they are not careful estimates.

Note that the policy impacts are represented as long-term impacts only and that they are incorporated by adding a column to the long-term impact matrix. The results of running the base case (no policy intervention) are shown in Figure 12.3. Note that the model predicts that the number (N) of fax units in operation quickly approaches the maximum value. Other variables grow rather quickly as well. However, the median cost (C) of a fax unit grows slowly and then declines. Before this model is used for forecasting and decision making, it would be necessary to verify, insofar as possible, and modify the variable initial levels and impacts. This process might begin by setting year 1 as some time in the past and checking to see if model predictions track historical variable behavior. It is easy to see that even this simple check might be difficult because of problems associated with gathering the necessary historical data. These problems might cause us to redefine variables to more readily fit the available data.

Assume the model has been fine-tuned and verified to our satisfaction and now we are interested in finding the changes in variable behavior that might be caused by a policy intervention. In the example, that intervention is a tax imposed on the normal cost of fax transmission. The impacts portrayed in Table 12.7 for this policy option were constructed assuming that a tax of approximately 5 percent is imposed. We can model various degrees of policy intervention by choosing different values on the range 0 to 1 for the initial level of the policy variable. Note

TABLE 12.7 Long- and Short-term KSIM Impacts on FAX Transmission Variables

On This Variable	Long-term Impact of This Variable					Short-term Impact of the Rate of Change of This Variable			
	N	C	T	$\$$	Policy	N	C	T	$\$$
$N(0.2)^*$	+3	-1	+3	-1	-1	+2	+1	+2	-1
$C(0.35)^*$	-2	0	-1	0	0	0	0	0	0
$T(0.15)^*$	+2	0	+3	-2	-1	+2	+1	+2	-2
$\$(0.30)^*$	0	+1	-1	0	+1	0	0	-1	0

* Initial values; variables scaled from 0 to 1.

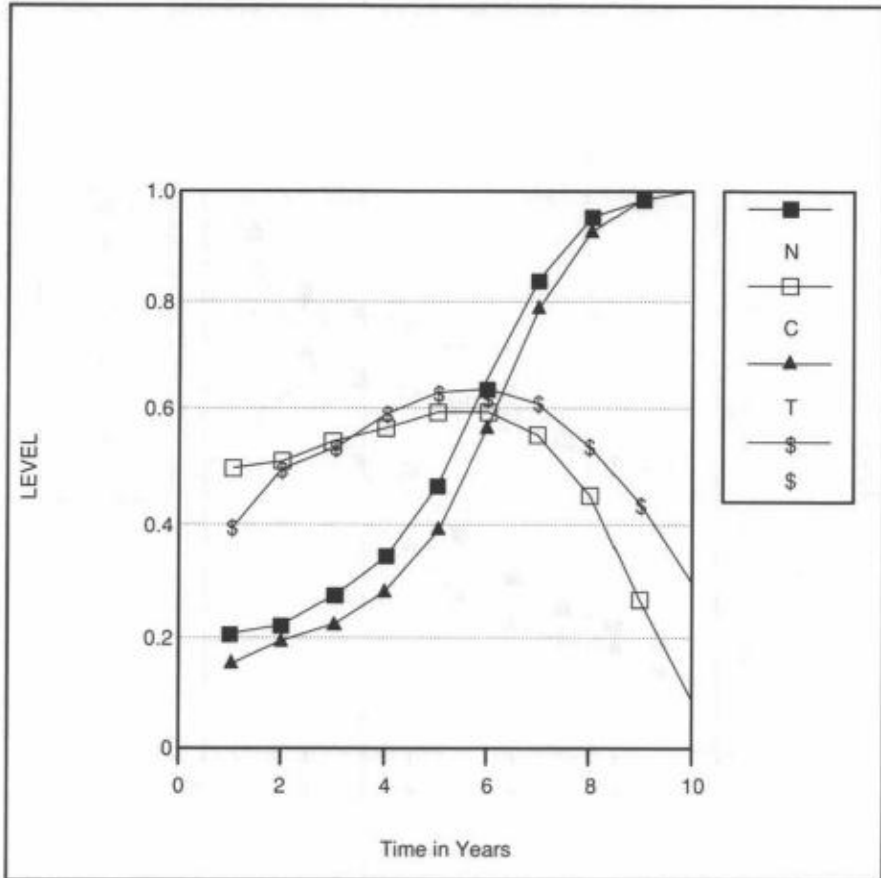


Figure 12.3 KSIM model of fax growth—base case.

from the matrices that policies are represented by columns but not rows in the matrix. Thus policies impact, but are not impacted by, other variables. For this reason, the value of the policy variable does not change with time. We might choose a policy implementation of 0.2, for instance, to indicate that we will tax only those transmissions that involve documents of 10 or more pages. The results of this policy intervention are shown in Figure 12.4.

Figure 12.4 shows that the policy intervention has significantly changed the picture from the base case. Neither the number of fax units (N) nor the pages of transmission (T) rise as fast as before. The median unit cost (C) rises to a higher value before falling, but the peak is delayed about two years. Even with the tax, the cost per transmission ($\$$) changes little over the first four years. However, it peaks at a much higher value about a year later than without policy intervention. It is important to note that the policy impacts must be verified and modified just like the model itself. Only when the behavior seems reasonable to the manager or forecaster can a degree of confidence be placed in the results. This point is

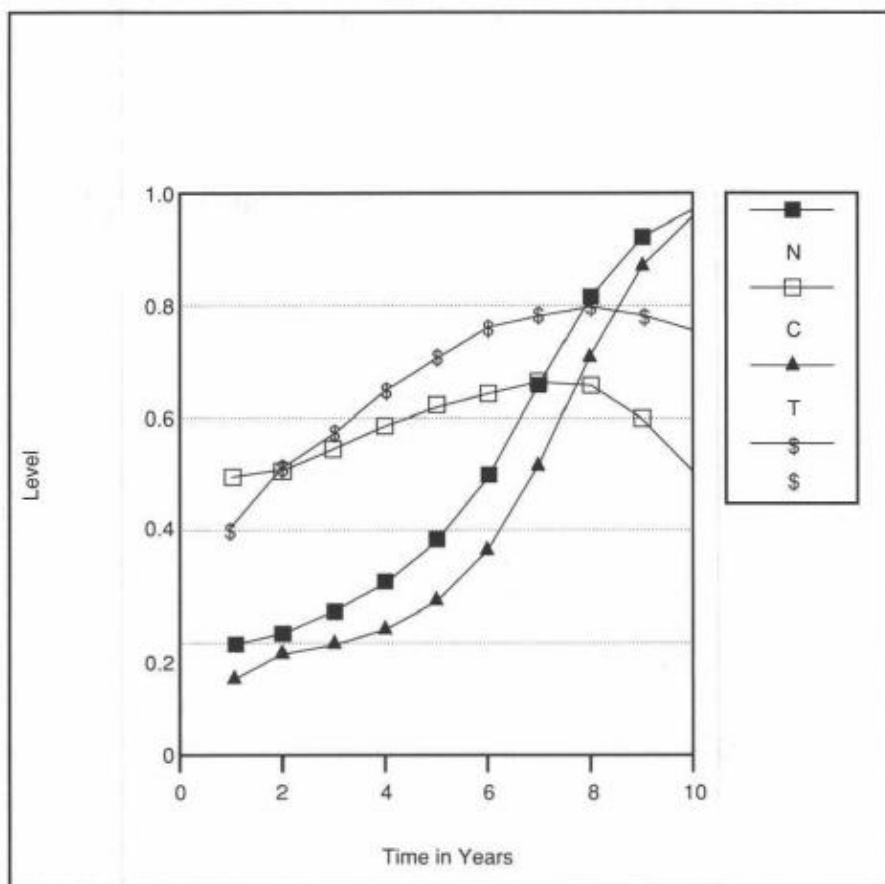


Figure 12.4 KSIM model of fax growth—policy intervention.

easily overlooked, for the model may assume a spurious credibility because of the sophistication of the computer equipment used to produce it. Beware!

The procedure could be extended to several policy options, or the policy option procedure could be used to model external events. ■

12.4 FORECASTING WITH SYSTEM DYNAMICS³

System dynamics (SD) is a quantitative simulation approach used to forecast and modify the dynamic behavior of important human systems. Both the philosophy and methods of SD are important. SD embodies a philosophy of physical and human

³ This section was contributed by Professor Willard R. Fey, Georgia Institute of Technology.

causality that is focused on systems that are complex, nonlinear, and aggregated and that involve information-feedback control processes with fuzzy goals, decisions, information gathering and transfers, production functions, and time delays. The variables that characterize the operation of such systems have time histories composed of complex combinations of trends, oscillations, and random variations.

SD methodology is used to: identify the dominant feedback loops; represent loop relationships with nonlinear difference equations; analyze loop dominance and forecast future performance; synthesize structure changes to improve patterns of behavior; and implement changes effectively in real systems. Features of the methodology flow from the nature of causality in human systems described by the philosophy.

SD most often is used to improve the performance patterns of real systems. Occasionally, it is used to design new systems or to forecast future system behavior of an existing system without modifying it. In either case, a model of system causal feedback structure must be built. Then this model must be analyzed to establish its validity. Only when sufficient confidence in the model has been established can it be used to modify or forecast system operation. This is done by simulating model performance over a period of time sufficient to reproduce a significant portion of system history and to project future behavior. Ordinarily, SD analysts do not attempt to forecast particular variable values at specific times. Instead, they try to forecast a pattern of variation over a period of time. In forecasting technology development, therefore, SD would not be used to forecast the appearance of particular technologies, breakthroughs, or production processes. Instead, the process within which the technology develops and the impacts on human systems would be represented.

This is accomplished by writing equations to model the technology creation process and the social and economic environments of interest. This model then is simulated to re-create past time histories for the variables that are affected by technology (such as productivity, material use efficiencies, costs of production/mining/growing, diversity of consumer products, speeds of transportation, etc.) to provide verification. Then these histories are projected to forecast future patterns. In this way the goals, resources, limits, and opportunities in the entire socioeconomic environment that affect technology development are taken into account.

12.4.1 The System Dynamics Philosophy of Human Systems

Human systems (families, companies, institutions, nations) are organized into goal-seeking, information-feedback loops as shown in Figure 12.5. Components of these systems incorporate the activities of perception, decision making, and action; accumulation of things, ideas, and emotions; and the concepts of goals, expectations, and beliefs. Loops do not function separately. They are connected to form complex, internally interacting feedback systems. Thus changes in one part of a system eventually are felt everywhere in it.

System performance cannot be produced by, nor deduced from, the parts separately. Some loops, however, are more influential than others, although dominance often shifts from one loop to another and all contribute and are affected. Most

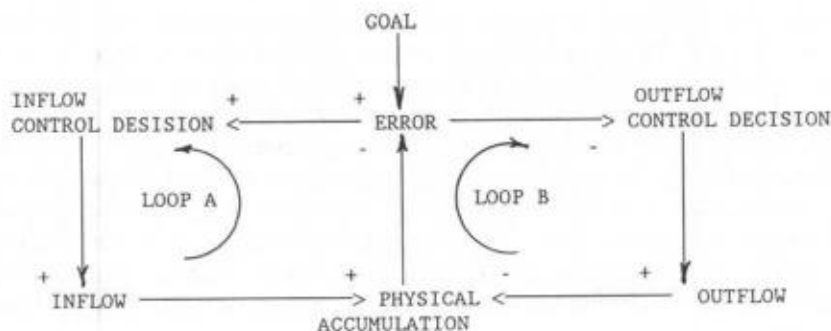


Figure 12.5 Feedback loops in a simple control system.

accumulations (information accumulations excepted) are deterministic. However, the deterministic behavior of system operations and concepts is modified by random forces, inconsistencies, and a partial lack of time continuity. In most real situations, control is imposed at the higher levels of aggregation for which goals are established. Therefore, SD models generally track aggregate flows rather than individuals or things. Most decisions in SD models are repetitive, and a fixed time between repetitions (or a continuous flow) is assumed even though the interval actually may vary. Thus computer-based SD simulations (such as DYNAMO and STELLA) use fixed time intervals between calculations instead of a variable interval based on random event occurrences, as do stochastic simulations.

Loop characteristics (such as number and types of accumulations, algebraic sign, time delay, strength of response, nonlinearities, etc.) determine the type and magnitude of the time pattern that an isolated loop produces. There also are aspects of the loop linkages that determine the complex combination of patterns that compose system time histories. Human systems are complex, nonlinear, and quasi-periodic with nonstandard objective functions. Therefore, neither rigorous deterministic nor stochastic mathematical analyses can be used to obtain closed form solutions (time responses), optimum parameter values, or optimum structure modifications. Thus simulation is necessary.

Since system behavior patterns are produced by interaction of all system components, they can be changed by altering the feedback structure. Structure change occurs often in real systems (for example, firing the coach, trading a player, or selling an unprofitable product line). However, there is seldom a loop analysis such as SD provides to justify it. Whether it can be analyzed or not, the structural change activity forms a metaloop, as shown in Figure 12.6, that may complicate the process considerably. The metaloop contains the system feedback loop structure that creates the performance patterns. These patterns are perceived and evaluated, and, when they are unsatisfactory, structure changes are developed and implemented. Figure 12.6 shows that SD becomes part of the metaloop when it is used for analysis and change in a real human system. Thus SD is a part of the system that is being studied; it is not an external, independent, objective, and passive activity,

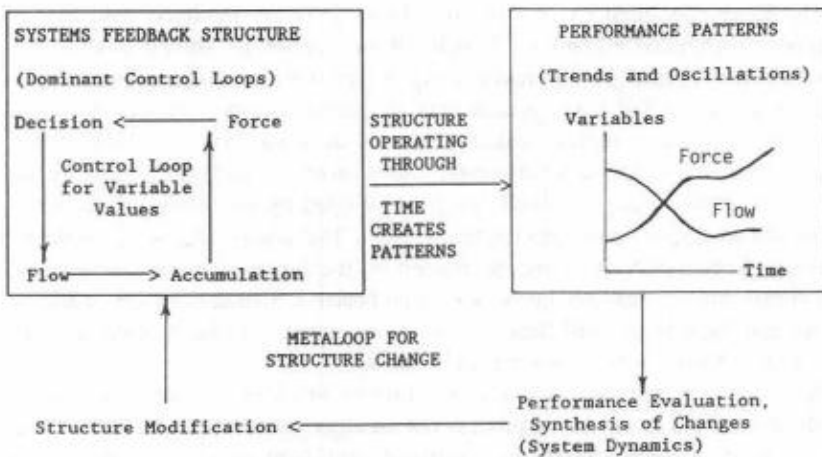


Figure 12.6 The metaloop that controls system structure change.

as scientific studies usually are assumed to be. Since it is an integral part of the system, the SD activity might be included in its own analysis.

12.4.2 System Dynamics Methodology

SD is a science-aided art usually practiced by a consultant or advisor to a human feedback system (company, city, nation, etc.). The consultant executes a series of prescribed activities in a logical sequence to solve a particular dynamic problem (for example, to create growth in a stagnant company or to reduce inventory oscillations). The usual SD problem involves changing the behavior pattern of an existing system by altering its feedback structure. However, occasionally the structure of a new system is synthesized or the future time patterns of an existing system are forecast without attempting to improve them. Whatever is done, the majority of the following SD activities will be required.

The first step is to identify the important variables about which the client is concerned. Then data are gathered to determine the historical time patterns for those variables. Next the operation of the system is observed and measured to determine how people, materials, products, money, and ideas are stored and flow through the system; how information is obtained and used; and how decisions are made and implemented. The goal is to understand the performance (time patterns) and operation (feedback structure). This understanding provides the basis for writing a hypothesis (the dynamic hypothesis) that identifies the feedback loops and linkages the analyst believes are most important. Since there is noise in the system and less important parts of the structure are omitted to simplify the analysis, the hypothesis is neither complete nor perfect. Therefore, the success of the analysis depends on the analyst's ability to distinguish between dominant and unimportant loops and between patterns caused by random forces and those caused by the operations of reasonably deterministic loops.

The loops identified by the dynamic hypothesis are modeled with first order, nonlinear difference equations. Usually these cannot be solved analytically, so system behavior must be simulated using one of the SD computer programs (such as DYNAMO or STELLA). Simulations are performed for a variety of conditions (different exogenous inputs, initial values of accumulation variables, parameter values and tables, decision functions, and/or noise sequences). These variations are specified by an experimental design developed by the analyst to determine if the model adequately emulates the real system. The analyst also seeks to determine if model performance patterns are created by the loops identified in the dynamic hypothesis for the reasons he or she hypothesized. Corrections are made to the model and hypothesis until they are "accurate enough" to be acceptable and until the causes of system behavior are understood.

Next modifications in feedback structure to improve system performance are synthesized. This is a creative process not an algorithmic one, unless the system is linear and very simple. Synthesis is difficult, and there are many errors that can be made, so it must be completed carefully. Structural changes (such as elimination of loops or creating new loops) usually are more effective than adjusting parameters. It is possible for different SD analysts to propose different changes for the same system or to overlook simple, highly effective ones. Nevertheless, reasonably effective changes do emerge from most SD studies.

At this point a "transition" analysis is performed to forecast system behavior from the time that recommended changes are made until improved system behavior is expected. This is necessary because system performance does not instantly improve when the structure is changed. Rather, in feedback systems, there is a transition period during which behavior differs from both the old and new patterns. This analysis involves simulation of the old and new system models together with the introduction of the changes to create one entire life cycle. This is an analysis of the metaloop shown in Figure 12.6. Once the transition analysis is complete, the recommended structure changes are introduced into the real system. Care must be taken to avoid undesired changes in other parts of the system and to see that the recommendations are implemented properly. Continuing observation and evaluation are necessary to correct implementation errors and to be sure that the expected transition and new system behavior patterns develop as anticipated.

When SD is used for forecasting only, the synthesis, transition analysis, implementation, and evaluation parts of the methodology are not required. However, all other activities described in this section still must be carried out.

12.4.3 Technology Forecasting with System Dynamics

Technology is particularly difficult to forecast. Previously unknown physical laws, materials, processes, and applications will be discovered and new technologies will affect society in many unprecedented ways both now and in the future. Changes in technology stem from research—a creative and expansive process. Basic research focuses on new discoveries about the laws of nature in physical or biological relationships. Advances occur in places where nature hides them, not necessarily where researchers seek them. Applied research concentrates on solutions to specific

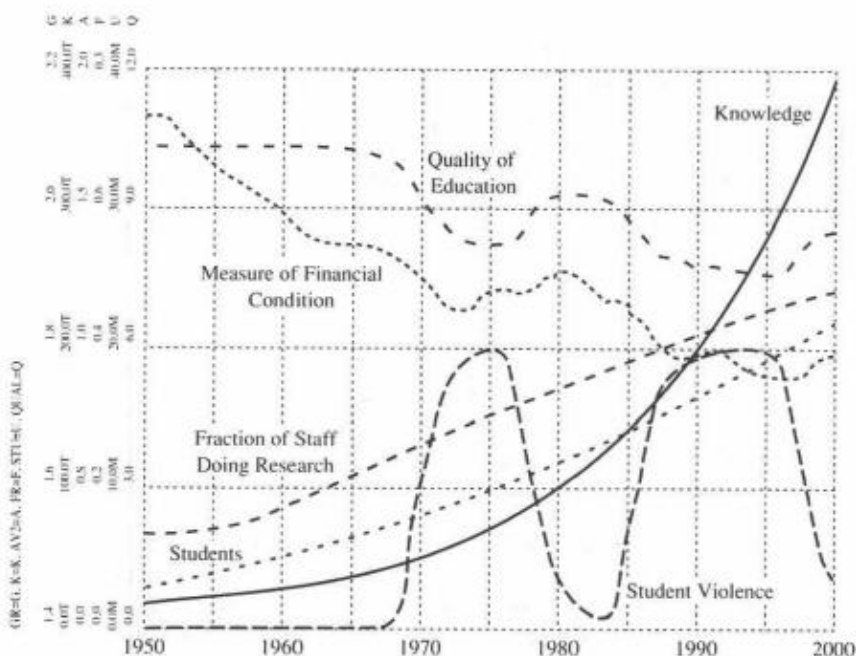


Figure 12.8 Simulated 50-year history (1950 to 2000) for Fey's education model. (Source: Fey, W.R. and Knight, J.E., *The Dynamics of Educational Institutions, Proceedings, 1971 Summer Computer Simulation Conference*, p. 1118, 1971. Reprinted by permission. © Simulation Councils, Inc.)

why the institutional setting has contributed so much to knowledge growth. Simulated time histories for several educational system variables are shown in Figure 12.8. The first 20 years of that simulation (1950 to 1970) approximate the historical patterns quite well. The last 30 years (1970 to 2000) constituted a forecast at the time the model was developed.

Morecroft (1986) examined the fledgling high-technology growth market for automated storage and retrieval systems (AS/RS) and developed a model to examine how marketing effort, capacity expansion, and industry reputation interact to produce eight- to nine-year growth cycles in the demand for these systems. In his model, three subsystems portray AS/RS market growth: (1) customer interest and industry reputation; (2) studies done by firms preparing to bid on systems; and (3) new competition. These are incorporated in an overall feedback model (see Figure 12.9) that can be used to design and interpret simulation experiments in market dynamics.

These three models illustrate the conventional SD principles associated with forecasting systems with important technological feedback loops. These principles include:

1. Technology cannot be studied by itself. It occurs in a context of social, political, financial, personnel, and educational factors that must be considered when the dominant control loops are identified.

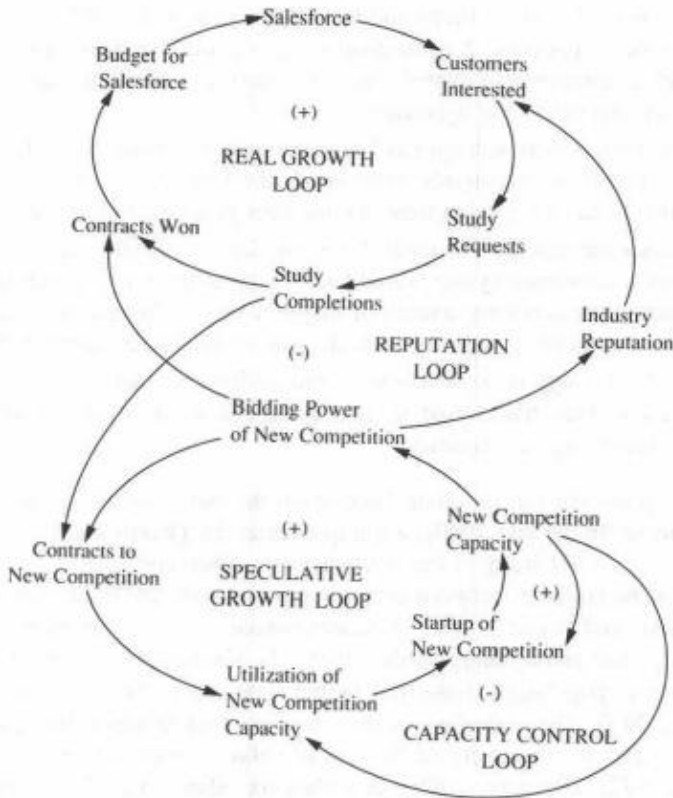


Figure 12.9 Morecroft's feedback loop structure of the AS/RS growth market. (Source: Morecroft, "Dynamics of a Fledgling High Technology Growth Market," *System Dynamics Review*, vol. 2, pp. 36–61, 1986. Reprinted by permission of the System Dynamics Society.)

2. Technological development is not a random process. While there is noise in the process and research outcomes are not perfectly predictable, the long-run socioeconomic performance patterns are created by feedback control processes.
3. Dominant control tends to be exercised at aggregate levels, so SD models for systems with technological aspects usually are highly aggregated. Individual research projects and individual researchers normally are not represented.
4. Forecasting is done with a causal model of the dynamic feedback control process within which technological development takes place. There is no extrapolation of past data time histories or sampling from probability distributions except when noise is used in the feedback system.
5. The objective is to forecast patterns of change over a period of time (even a century or more), not specific research outcomes or values of variables at particular times.
6. The standard SD methodology is used to determine past performance and system structure. A dynamic hypothesis is written to describe the domi-

nant feedback control loops and to indicate how their operation creates the performance patterns. A corresponding quantitative model is developed. The model is analyzed (simulated under a variety of conditions) to establish its validity and verify the hypothesis.

7. When the model is accepted as "valid enough" (it cannot be perfect) and the loop structure is thoroughly understood, the forecast is made by simulating the model over a period of time that includes past history and a future period.
8. Forecasts are usually not made for technological variables themselves, but rather for important system variables. In education, for instance, the growth of knowledge is not the pattern of major concern. The quality of education, financial condition, and student body size would be the subjects of study.
9. Since technology is so pervasive in our civilization, there is almost no significant system that can be studied using SD without the consideration of some technological variables.

This last point has an important bearing on the most widely known SD study ever conducted. In the late 1960s, a group directed by Dennis and Donella Meadows conducted an SD study of the world system. They attempted to represent the dynamics of the collision between expanding population and industrial output, on the one hand, and "fixed" limits of nonrenewable resources, pollution assimilation capacity, and arable land, on the other. The simulation forecasted continued rapid growth in population, industrial output, and most other variables until approximately 2050. Then one by one the variables peaked and collapsed to levels from 20 percent to 60 percent of their peak values within 30 years (see Meadows et al., 1972). (Bremer—1989—describes six other world-simulation models, although none are based on SD.)

Economists correctly criticized the Meadows world model for omitting financial variables such as prices, interest rates, money supply, and debt levels. However, omitting technological development was an even more serious error. Many of the constant values and constant transformation functions in the model (such as industrial capital/output ratio, service capital/output ratio, inherent land fertility, food processing loss, etc.) depend on technology. To change these constant factors to variables would require including the people and resources devoted to creating new knowledge as well as the ways in which they are used in the model. Clearly, such an addition is needed. It was the development of medical and industrial technology over the last few centuries that extended the limits that restrained population expansion and the quality of life. It is not surprising that the old limits and some new ones appeared with a vengeance in a model that froze technological development. There may be insurmountable limits of some kind but they cannot be assessed credibly unless realistic technological development factors are included in the analysis.

12.5 GAMING SIMULATION

It is possible to use simulation for technology forecasting and management through game playing. Gaming employs an entirely different approach than the methods

described in other sections of this chapter. It involves constructing a realistic set of rules and then observing the behavior of players who either compete or cooperate to achieve a goal within the context provided by the rules. Forecasts or management decisions are then conditioned on the behavior of the players and/or on the outcomes they are able to achieve.

The game Monopoly® provides such an example. In Monopoly the goal is to achieve economic dominance. The rules are loosely patterned on real-world business and economics, and uncertainty is introduced by the throw of the dice and randomly shuffled and drawn instruction cards. Thus Monopoly loosely can be styled as a simulation game. We certainly can learn something of the motivation of our friends (and ourselves) by observing the behavior of the players.

Many simulation games have been constructed to investigate organizational change, business development, war tactics and strategy, policy making, communications, learning, environmental planning, health care, diplomacy, gambling, and so forth. An International Simulation and Gaming Association publishes the proceedings of their annual conferences (see Crookall et al., 1988; Klabbers et al., 1989), and two journals publish developments in the field—*Simulation & Games* and *Simulation/Games for Learning*.

Gaming is a powerful method for treating issues that are complex and ambiguous. In many instances, it has proven an effective alternative or adjunct to strategic planning methods used by large institutions. The United Nations, Upjohn Pharmaceutical, Chase Manhattan Bank, Xerox, and Conrail, for instance, have used gaming to solve problems in a complex environment. Gaming is especially appropriate for situations characterized by

1. Numerous interacting variables
2. Variables and their interactions that are difficult or impossible to quantify
3. The absence of a conceptual model for decision making
4. Decision making in a sociopolitical context in which the players may be idiosyncratic and unpredictable

While it is not possible here to consider all the caveats associated with good gaming, there are a number of general rules that can be given. These are discussed in four general areas—designing, playing, evaluating, and correcting the game.

In designing a game, the manager must first set the objectives for the simulation and the goals toward which players will work jointly or competitively. Next the rules governing player actions and interactions and other aspects of the game must be devised. For the simulation to be effective, these should emulate the most important aspects of the situation being gamed. Measures of player success also must be provided. These are usually translated into indicators displaying the degree to which the goal has been achieved. Indicators might be token money (as in Monopoly), judgments made by other players, mathematical payoff functions, and so forth. Further, a display system must be devised to show the progress of the game. This might be a board, as in Monopoly, or a video display, as used by Nintendo® and other electronic games. In the business environment, a computer

screen often may be the most appropriate display. Finally, the game must include mechanisms for introducing the randomness that results from exogenous (outside) events.

It is also important that players represent as many of the different groups involved in the situation being simulated as is necessary to achieve reality. Each player will bring a somewhat different culture and perspective, which is important to the context of the game. The game's rules and goals should be carefully documented, each player should have a copy and should understand the goals and rules before beginning. A nonplaying observer should oversee the game and make decisions in cases not covered by the rules.

Once the game is completed, the results must be evaluated and compared with the objectives of the game. The indicators can be used to help determine how well the objectives of the simulation have been met. Players should be encouraged to provide feedback on the realism of the game, the clarity of the rules, and so on. Also, since one of the objectives usually is to learn something about the situation,

■ Exhibit 12.3 A Manufacturing Technology Game: ADVANTIG

It is useful to consider one of the many games in common use to clarify the contexts in which gaming can be useful. Exhibit 12.3 provides an example.

Advanced Manufacturing Technology, or ADVANTIG⁴ is a simulation that has been used by manufacturing companies trying to transform their operations using advanced manufacturing technology such as computer aided design/computer aided manufacturing (CAD/CAM), robotics, automated material handling, and material requirements planning (MRP). The goal of the game, as the name implies, is to win the automation game by gaining a manufacturing advantage over competitors.

To play ADVANTIG, a team from the company tries to transform a fictitious firm, F.L.I. Casting, Inc., from a low-technology to a high-technology organization. Key management plays the roles of F.L.I. CEO, production manager, head of engineering, head of finance, and so on. There typically are from 8 to 21 participants who should have some knowledge of CAD, computer numerical controlled machines, robotics, assembly techniques, and MRP. Decisions about strategy, technology, and implementation are made by the team. Their objective is to select advance manufacturing technology and to implement it profitably while meeting present customer obligations and capturing new business.

Four aspects of F.L.I. operations are simulated: production, management, engineering systems, and dealing with advanced manufacturing technology vendors. Management sets F.L.I. strategy, bids on contracts, and maintains financial records such as contract forms, balance sheets, and material requisitions. Engineering buys and implements new technology with the help of the vendors. The game typically lasts from four to five hours and is followed by a two-to-four-hour evaluation (debriefing). ■

⁴ADVANTIG is available from the Industrial Technology Institute, P.O. Box 1485, Ann Arbor, MI, 48106.

players should be encouraged to interact with each other and analyze their actions and interactions.

The evaluation likely will show that the objectives, rules, and/or conduct of the game need to be revised for it to more nearly represent the situation being simulated. Once revisions are completed, the revised game should be conducted again with new players and revised again based on a new evaluation. Thus the game is iteratively made to more nearly simulate the situation it portrays. Once the manager is satisfied with the simulation, it may be possible to replay the game with a set of historical circumstances. Although player response and exogenous events may occur differently, it should be possible to calibrate and fine-tune rules and objectives somewhat to increase confidence in the game.

12.6 CONCLUSION

Simulation is a powerful tool for the forecaster and technology manager to apply in the proper circumstances. However, he or she must realize that simulation generally requires a significant expenditure of time and resources. It is important when weighing the strengths and weaknesses of various approaches to view the building of simulation models as a significant aid to understanding complex situations. Therefore, simulation must be considered for its value as a process as well as its usefulness as a predictive and/or analytical method.

Simulation has different meanings and can be accomplished by applying many different techniques. All simulation, however, seeks to simplify complex situations while retaining those aspects that are important for understanding system behavior. Thus all simulation models contain critical core assumptions that must be carefully tested. Moreover, all simulation models require modification and fine-tuning when first completed to produce system emulation that is satisfactory to the user—that is, validate the model. One means of validation is to see if the model can “predict” historical behavior patterns. It is important to note, however, that no model of a complex situation such as technology growth can be completely validated. Success in predicting historical behavior is no sure guide to successful prediction of the future. Thus the technology manager or forecaster must retain a healthy skepticism when employing simulation as a tool.

EXERCISES

- 12.1 Play the CI game in the fax example (see Table 12.1) for 100, 200, 500, 1,000, 2,000, 5,000, and 10,000 runs using the computer program in the TOOLKIT. Plot the values of the cross-impacted probabilities. Draw inferences from this exercise.
- 12.2 Suppose we want to forecast the future of an electric car that reaches a speed of 60 miles per hour, travels 120 miles between recharges, and requires 12 hours for a full recharge. Prepare occurrence and nonoccurrence matrices similar to those in Tables 12.1 and 12.2.

- 12.3** Manually play the AIMTECH matrices shown in Table 12.5 20 times. Compare the cross-impacted probabilities to the original values.
- 12.4** If you have access to XimpacT, execute the program for a 10-year horizon using PERSONAL.CIM as input. Display all the tabular and graphical outputs.
- 12.5** Consider the Markov transition matrix information given in the AIMTECH Exhibit 12.1.
- (A) What is the probability of remaining in State 1 for another step?
 - (B) What is the probability of going from State 1 to State 3 on the next step?
 - (C) What is the probability of going from State 3 to State 2 on the next step?
- 12.6** Using the same information as in Exercise 5,
- (A) What is the probability of going from State 1 to State 3 *on* the third step?
 - (B) What is the probability of going from State 1 to State 3 *by* the third step?
- 12.7** For the AIMTECH example, compute S^{1997} if $S^{1996} = (0.02, 0.13, 0.53, 0.26, 0.06)$.
- 12.8** Consider Exhibit 12.2. Using the program in the TOOLKIT, vary the impact of the median cost, C , of a fax unit on the number of fax transmissions, T , by increments of 0.5 and compare your results to those in Figure 12.3. Investigate the impact of different initial variable levels on the result.
- 12.9** Develop a policy implementation of your own and investigate the effects that varying its degree of implementation (initial value) has on the base case shown in Figure 12.3. Repeat this using an external event instead of a policy.
- 12.10** Develop a KSIM model for some older technology (such as television). Choose the variables and initial values. Ask your classmates to help you estimate the cross-impacts. Use the KSIM program in the TOOLKIT to run the model and compare the forecast behavior with the historical trends of development. Modify the model until you have what you consider to be an acceptable match.
- 12.11** Obtain a copy of Morecroft (1986) and analyze the article.
- 12.12** Morecroft (1986) contains a copy of the DYNAMO program used to draw the inferences given in the article. If you have access to DYNAMO, first enter the code and obtain output similar to that presented. Modify the inputs and examine the output to gain greater insight into the model.

--- 13

SCENARIOS

OVERVIEW

Scenarios are outlines of some aspect of the world—in our case, the future world. This chapter discusses the nature of scenarios, presents the various types that can be useful to the technology manager, and suggests how the manager might use them. Scenario construction is also considered in some detail. Utility and validity, which are useful criteria by which to evaluate scenarios, are presented.

13.1 INTRODUCTION—WHAT ARE SCENARIOS?

Webster's New World Dictionary defines a scenario as: "an outline or synopsis of the plot of a drama, opera, etc., indicating scenes, characters, etc.; or an outline of a motion picture indicating the action in the order of its development, the scenes, the cast of characters and their appearances, etc."

Rather than outlines of dramas, operas, or movies, for our purposes, scenarios are outlines or synopses of some aspect of the future. Although they may be presented in narrative form, they often include numerical data and graphics. Scenarios even may incorporate multimedia presentations of still and motion pictures, audio, and so forth. There is no reason why scenarios should not be as appealing to their audience as *Star Wars* was to its.

Scenarios incorporate and emphasize those aspects of the world that are important to the forecast. These are the *dimensions* or important factors of the scenario. Scenarios may integrate several specific forecasts concerning individual dimensions to create a multidimensional forecast.

Some of the most interesting and well-known future scenarios are the better works of science fiction. By imaginatively presenting stories of the future, authors encourage their readers to open their minds and understand conditions or environments that may appear alien or "far out." While Jules Verne's descriptions of long undersea voyages and travel to the moon must have struck nineteenth-century readers as speculative, to say the least, analogous technologies now are commonplace. Aldous Huxley's vision of a genetically programmed society in *Brave New World* has not yet been realized. With each passing year, however, that sort of society becomes increasingly possible as biomedical technologies at the molecular and molar levels develop rapidly.

The systematic use of scenarios in forecasting, planning, and strategic analysis began after World War II. Becker (1988) divides the period from 1950 to 1986 into three parts. From 1950 to 1965 scenarios were used in the United States to set the context within which action took place. In France they were taken as portrayals of strategies and their outcomes. During this period the predictive power of forecasts was thought to be high. However, from 1965 to 1975 users began to realize that scenario-based predictions were not as accurate as formerly believed. From 1975 to 1986 forecasters began to integrate strategies and contexts into single scenarios. Thus more realistic views of the use and accuracy of scenarios for forecasting have eventually prevailed. During this later period, impact assessments (such as technology assessments and environmental impact assessments) were especially common scenario forms. Evaluation of old forecasts began during this period as a means of improving the quality of forecasts (see, for example, Ascher, 1978).

Scenarios usually incorporate uncertainty. This is expressed by describing a range of possibilities, usually by a few individual scenarios that, taken together, constitute the range of futures considered likely. Alternately, uncertainty can be expressed as the probability of a single scenario or, at a micro level, as a range of projections for the dimensions within a single scenario. It is important for forecasters to avoid the bane of scenarios—fantasy. While science fiction may incorporate fantasy, the forecaster must maintain close contact with reality. The discussion of critiquing scenarios presented in Section 13.5 should help maintain a firm basis in reality.

13.2 TYPES OF SCENARIOS

Scenarios can be distinguished based on their temporal orientation—that is, as future histories, snapshots, or combinations of both. *Future histories* map a trajectory from the present state to some future time, focusing on a particular part of the world and tracking its development during some period. "The Energy Age" (Mathisen, 1977), for instance, presents a hundred years of future energy history from 1977 to 2077. A classic work on conflict escalation emphasizes stages or "rungs" in developing escalation scenarios (Kahn, 1965). Future histories stress the dynamics of a system, possibly at the expense of its structural or cross-sectional characteristics. Asimov's *Foundation* series is a fictional example of a future his-

tory. *Snapshots*, on the other hand, present a cross-sectional view at a single point in time, and structural, rather than dynamic, characteristics are emphasized. Orwell's *1984* and Huxley's *Brave New World* are fictional snapshots. Considered from a planning perspective, snapshots portray goals or end states. Future histories portray plans or paths to end states. These views can be integrated in a single scenario consisting of a trajectory that leads to an end state.

Another distinction between scenarios is whether they are descriptive (that is, extrapolative) or normative. Descriptive scenarios address the issue of what *will* come to pass. Normative scenarios describe what *should* come to pass. These scenarios respond to planning and policy-making concerns by representing goals and plans. Descriptive scenarios simply lay out the range of possible futures determined in the forecast without attempting to evaluate them as desirable or undesirable.

Scenarios very often are constructed in sets. When a single dimension is dominant, the most common grouping is a set of three. The *baseline*, or most likely scenario, results from extrapolating current conditions and trends. This scenario tracks business as usual. The set also includes two extreme scenarios, referred to as *optimistic* and *pessimistic*. These evaluative designations refer to changes in the central dimension from the manager's perspective—positive changes in the optimistic and negative changes in the pessimistic scenario. Of course, when there are many important dimensions, the "space" in which the scenarios are embedded becomes multidimensional, and the selection of appropriate scenarios is more complicated since all relevant dimensions must be considered.

13.3 HOW ARE SCENARIOS USED?

Scenarios can provide intellectual stimulation for imagining the range of possibilities that the future may hold. This is exemplified by science fiction literature and by "visionary futures." This stimulation is valuable. It can provide forecasters, forecast users, and the public with a range of images that will open their minds and expand their imaginations to encompass a future that may be surprising and unexpected. Exhibit 13.1 shows a potential use of trends and scenarios to help form public policy.

As discussed, scenarios may present descriptive forecasts or they may offer normatively based plans or goal states. However, they have other uses in the planning and policy-making processes as well. Becker (1983) points out three distinct uses for scenario forecasts. First, scenarios can be used to assess whether various policies or actions will assist or inhibit the realization of conditions described in them. This approach can be used to construct future histories that link the implementation of policy options to desired future states. Second, scenarios can be used to assess how well alternate policies and strategies would perform under different conditions. This involves creating snapshots of future policy impacts and future histories leading to them. Finally, scenarios can provide a common background for individuals involved in planning within an organization, an exercise in environment creation using scenarios as the vehicle for creation and presentation. In the case

of normatively based plans and goal states, snapshots can be used to identify the goal states and future histories can be used to describe the workings of the plans designed to achieve them.

Scenarios have been used widely in business and government applications. Businesses can use them to pull together the data and assumptions underlying their strategic plans. This can help to communicate the plan to employees, to financing sources, to stockholders, and, in some cases, to the public. Governments can obviously use them to explain plans for the future and to build public support for their realization. Exhibit 13.1 describes legislation that will increase the formal use of scenarios in the work of the U.S. Congress.

■ Exhibit 13.1 Alternative Scenarios in the U.S. Government

Senator Albert Gore has proposed a bill, the Critical Trends Assessment Act, that would give the federal government a highly visible source of major alternative futures, or scenarios (Gore, 1990). The bill recognizes the complexity and interdependence of the modern world, exhibited by issues such as global environmental change, homelessness, and international economic competition. It notes the richness of information sources available on many critical trends, and it targets the need to integrate existing sources of information to posit alternative, long-range (20-year) futures.

The bill would provide for ongoing assessment through a new Office of Critical Trends Analysis, which would report on alternative futures every four years. In addition, the Joint Economic Committee of Congress would produce a similar, independent report every two years.

This initiative extends a pattern of increasing use of scenarios by various levels of government and by large corporations. They have proven to be a valuable tool for understanding the uncertainties and the options of the years ahead (Bezold, 1990). ■

The scenario form has two generic uses that can be exploited within the forecasting, policy, and planning processes—*integration* and *communication*. First, the scenario offers the potential to integrate information from diverse sources and of different character into a single forecast. The results of trend extrapolation and expert opinion studies, for instance, can be woven together in narrative form. Further, quantitative data and qualitative information can appear in a scenario side by side with values. Thus the limitations of structure that trend analysis, expert opinion studies, and models exhibit individually can be overcome by incorporating their forecasts in a scenario. Broad features of a situation can be expressed clearly and in an integrated and understandable way, which brings us to the second generic use of scenarios—communication. The technical report does not convey information effectively to nontechnical audiences. In format, length, content, and style the technical report is not usually a good vehicle for mass communication. The scenario, which expresses a technical analysis in literary form, is a far better

approach to such an audience. It can engage their imagination without sacrificing sound technical analysis.

In cases for which there are no data, credible experts are lacking, and the assumptions required to develop good models are not available, scenarios usually are the only way to create a forecast. In such cases, a good deal of imagination is required to build on whatever information is available. The pitfall is that these scenarios may be more fantasy than forecast since their basis in fact, while the best available, may not be sufficiently robust. Yet even such an effort should narrow uncertainty about the future and focus the attention of decision makers and other involved parties more precisely than if scenarios were not attempted.

13.4 CONSTRUCTING SCENARIOS

The procedure for constructing scenarios presented here should not be viewed as a rigid prescription, but rather as a checklist or outline that systematically guides forecasters without rigidly binding them. Scenarios are an opportunity for creativity both in thought and expression, but they are likely to be more useful if they have structure. Exhibit 13.2 outlines a typical set of steps for successful scenario construction.

■ Exhibit 13.2 Checklist for Scenario Construction

1. Identify topical dimensions
 2. Identify intended users' interests and the appropriate style of information presentation
 3. Specify time frame
 4. Specify general societal contextual assumptions and specific (technology) assumptions
 5. Set out the key dimensions
 6. Decide on the number of scenarios and their emphases
 7. Build and present the scenarios
-

13.4.1 The Topical Dimensions

The first consideration in constructing a scenario is to identify the dimensions that it will incorporate. These are typically determined by the purposes for which the scenario is being constructed. However, there are situations in which the forecaster has some flexibility in identifying the dimensions to be addressed. This flexibility should be used (1) to lay out dimensions that are relevant for the purposes of the forecast and (2) to assess the forecaster's ability to build a sound forecast from the information that is available or that can be developed in the course of scenario construction.

13.4.2 The User

Scenarios are constructed for a specific user or group of users. These might be the forecasters themselves, a government agency, a corporation, a public interest group, an individual, or the public. The focus of the scenario must relate to the needs and interests of the user. The format in which it is organized and presented should reflect not only the user's interests, but also the type of information and style of presentation that are most appropriate. For example, scenarios intended for a popular audience should not be presented in a technical report format; a literary or multimedia presentation would be more appropriate. By the same token, scenarios intended for a professional audience, such as economists, should incorporate all the economic sophistication and technical style required to satisfy their critical demands. Although the forecaster should be guided by the needs and interests of the user, professionalism sometimes demands that factors the forecaster believes are important should be considered even if the user may believe them to be irrelevant or wish to ignore them.

13.4.3 The Time Frame

The time frame places limits on the character of a scenario. For example, a very short frame, say a year, usually implies that current conditions and trends will dominate. Unless conditions are extremely unstable, this leads to immediate extrapolation of current conditions and to scenarios that tend to bunch around the baseline. On the other hand, a very long time frame, such as a century, almost ensures that accurate information and forecasts will be lacking even for relatively well-defined parameters, such as population. Thus the task of the scenario writer becomes speculative; success depends on his or her ability to integrate imagination and judgment with a very narrow information base. The intermediate case, perhaps 5 to 20 years, is more common for the technology manager. It allows a variety of forecasting techniques, such as trend analysis, expert opinion, and modeling, to be integrated into the scenario. Current conditions and trends may, of course, change even over the time frame of an intermediate forecast. Therefore, the uncertainty inherent in them is higher than that in very short term forecasts.

13.4.4 The Assumptions

The assumptions on which scenarios are based are the most important factors in determining the quality of the forecast (Ascher, 1978). Time spent examining and refining assumptions at the beginning is time well spent, as it can help eliminate errors that may force the entire forecast to be redone. The time frame is a major determinant of the stability of assumptions. The longer the time frame, the less likely it will be that current conditions and trends will continue. Therefore, as the time frame increases, progressively greater care is required to formulate the underlying assumptions.

Assumptions usually fall into two categories. The first, *general societal assumptions*, usually relate to broad conditions of political decision making and stability,

the state of economies, societal values, global environmental conditions, and demographic trends. They also may involve the rates of development of various families of technologies and the rate at which they are implemented on a broad scale. The second category consists of *specific assumptions* relating to the forecast at hand. For example, in an energy forecast for the United States, these might involve governmental energy policies, the conditions in certain industrial sectors, societal attitudes toward energy use and conservation, and the status of critical technologies (including the possibility of technical breakthroughs).

Assumptions of both types will vary with the time frame of the forecast and with the perspectives of interest groups involved in the forecast as actors, sponsors, or users. Different interest groups generally have different biases. For example, there is an obvious difference in perspective on the relationship of smoking to cancer for the American Cancer Society and the tobacco industry. The forecaster will find it impossible to eliminate every bias—nothing would be left. However, it is worthwhile to make the biases contained in the assumptions as explicit as possible.

13.4.5 The Dimensions

The users and the purposes of the scenarios determine the dimensions of their structure. Scenarios can be one-dimensional or multidimensional. The more dimensions, the more complicated the scenarios and the more complex their construction and presentation. Although completeness is important, unneeded complexity can dull presentation clarity. Therefore, the forecaster must decide how each dimension will be expressed and what level of detail will be required for each.

The structure of each scenario must define interrelationships between dimensions as well as relationships between dimensions and the boundary conditions of the scenario. These can be formatted as a kind of template or pattern and used for all scenarios in the set. This template also must incorporate the underlying assumptions. Dimensions can be represented quantitatively or qualitatively, and the level of aggregation at which each is expressed must be determined. For example, a set of scenarios dealing with future U.S. energy use might incorporate three dimensions: yearly energy use, principal sources of energy, and major end uses of energy. The first dimension could be expressed quantitatively in quads of energy; the second and third dimensions could be expressed qualitatively. Sources could include coal, petroleum, renewable sources, nuclear fission, and nuclear fusion. End uses might include commercial, industrial, transportation, and residential uses.

Representations of the dimensions may involve forecasts. Thus the scenarios may integrate a number of forecasts dealing with different topics and incorporating different methodologies. In the preceding energy scenarios, energy use could be forecast by trend extrapolation and demand modeling (see Chapter 10) with some consideration of the potential impact of major events with significant probability of occurrence (see Chapter 12). Energy sources could be described by combining trend extrapolation based on time series data that describe the amount of energy produced by each (see Chapter 10). Expert opinion about their relative future importance, perhaps gathered with a Delphi poll (see Chapter 11), could be added. The sources of

data and forecasts for each dimension must be considered in light of available time and budget. For a quick, low-budget effort, new forecasts are precluded; however, when significant time and money are available, sophisticated models may be built as part of the forecast.

13.4.6 The Number and Emphasis of Scenarios

The number and emphasis of scenarios can be determined by a procedure that uses morphological analysis (Chapter 7). First, consider each dimension of the scenario. For those to be represented qualitatively, determine the major emphases that the dimension may assume. For dimensions that are quantitative in character, select appropriate ranges of values including the baseline and extremes as well as various intermediate ranges if they are appropriate. The emphases and/or appropriate value ranges of each dimension establish the morphology of the scenario space. This morphology defines the maximum number of possible scenarios. The problem now is to identify the scenarios that will be significant in meeting the goals of the forecast.

For forecasting a single quantitative dimension, one to three scenarios almost always are adequate. The scenarios for a single qualitative dimension depend on the major emphases of that dimension. Multi-dimensional scenarios require more complex representations in which the sets of values for the various dimensions are considered as wholes. The most useful combinations of these are selected as the final set of scenarios. The major concern is to represent a complex situation without using too many scenarios.

Selecting scenarios from the morphology requires linking potential scenarios to the goals of the forecast. Initially it is best to choose high-probability scenarios whose outcomes will make a substantial impact. Often this set is satisfactory. Sometimes, scenarios that are too similar must be eliminated. At times, arbitrary judgments will be necessary to arrive at a final set.

For example, in the energy scenarios discussed previously, it may be convenient to largely ignore the mix of end uses by assuming that it will be more or less stable in all scenarios and similar to the current mix. Absolute values for energy use could be selected as high, medium, and low with suitable quantifications for each. Potentially dominant sources of energy then could be selected. In this instance, we might choose petroleum, coal, solar and renewable, and nuclear fission as candidates for dominance and reject nuclear fusion as unlikely. Thus we would have a maximum of 12 scenarios—three values $\times$ four sources. Next we need to decide how to eliminate unlikely or redundant cases. We might select petroleum/medium as the baseline scenario. In addition, we probably would want at least one high and one low energy use scenario as extremes. Likewise, each energy source should be highlighted. Thus we would have a minimum of four scenarios to work from. Now we rely on the background information and forecasts available to make the final selection. For example, in addition to petroleum/medium (baseline), the scenario set might include nuclear fission/high, solar/renewable/low, and coal/medium. The procedure for selecting scenarios is similar in other cases.

13.4.7 Building and Presenting Scenarios

The first requirement is to construct the template that each scenario will follow so that all will be comparable. Next the template must be filled in for each scenario using available forecasts and background information. The resulting scenarios then must be checked for utility and validity (see Section 13.5). When the scenarios meet the criteria established for them, they can be presented to the users.

At this point, the style and media used for presentation become important considerations. Whether as a technical report, a set of short stories, or a movie, the intellectual content of the scenarios should be packaged to effectively reach the intended audience.

13.4.8 Examples of Scenarios

A couple of examples should help clarify the process of building scenarios. Lough and White (1988) discuss the issue of decommissioning commercial nuclear power plants and consider two options, DECON and SAFSTOR. DECON involves immediate decontamination and dismantling. In this option, the plant and site are completely cleaned up, all radioactive materials are removed to a radioactive waste disposal facility, and the plant and site are released from all licensing restrictions. The SAFSTOR option involves safe storage and mothballing the facility, followed by deferred dismantling. Here, the plant and site undergo minimal clean up and then remain under physical security and licensing restrictions for a period of, typically, 20 to 50 years. The overall impact on reactors of either of these alternatives is similar. SAFSTOR generates lower occupational doses of radioactivity and lower levels of radioactive waste, however, which compensate for the costs and uncertainties involved in controlling the site for a long period.

Lough and White developed three decommissioning scenarios. The most important dimension is the institution that dominates decommissioning policies. The policies and outlook of the institution determine the criteria for selecting among the decommissioning alternatives. These criteria constitute the second dimension of the scenarios. The preferred alternative is the third dimension, taken as optional.

The federal government, in particular the Nuclear Regulatory Commission (NRC), dominates decommissioning policies in the first scenario. In accordance with the NRC staff preferences regarding timeliness and planning criteria, the two options that dominate the technical choice set are DECON and the SAFSTOR option with dismantling deferred up to the maximum of 50 years.

In the second scenario, state authority dominates nuclear activities. Because state regulators place high value on timeliness and on the planning criteria associated with decommissioning, DECON is the only acceptable technical choice. Also, the radioactivity levels of the site are expected to be returned to preconstruction values.

The utility dominates decommissioning policies in the third scenario. All technical options are considered by the utility with the criteria for choice being low cost.

Another example is provided by a set of contextual scenarios that deal with the health care of the elderly in the Netherlands from 1984 to 2000 (Becker et al., 1986). Medical, biological, and technological developments are assumed to be identical for all three scenarios of the set. Further, economic conditions serve as a backdrop and do not greatly influence any scenario. Social, political, and demographic conditions are the dimensions that vary from scenario to scenario. The scenarios are abstracted in Exhibit 13.3.

■ **Exhibit 13.3 Three Alternative Scenarios of Health Care for the Elderly through 2000 in the Netherlands**
(Based on Becker et al., 1986).

Scenario 1 (Baseline): Autonomous developments resulting in about the same pressure on health services for the elderly as currently felt:

1. Continuation of the emancipation of women resulting in a decrease in informal help from women to the elderly
2. Emancipation of the elderly as a result of higher levels of education
3. Increasing political influence of the elderly
4. Increase in the proportion of elderly belonging to ethnic minorities
5. Continuation of the debate with regard to euthanasia and an increase in passive euthanasia
6. Decrease in the number of children (important with regard to children helping their parents when they are old)
7. No major changes in norms and values with regard to health and health services

Scenario 2: Autonomous developments resulting in more pressure on health services for the elderly:

1. Growing professionalization of care for the elderly
2. An attitude of the population toward health services that results in a relatively high consumption of these services
3. Less solidarity between generations (less care by children, etc.)
4. Less solidarity within the elderly generation (less informal help from seniors to seniors)
5. Norms and values in society put the elderly more and more into an isolated position (disengagement, mandatory retirement)
6. The elderly are less financially dependent than in 1984, so they can pay for professional help
7. No new types of living together of the elderly

Scenario 3: Autonomous developments resulting in changes in pressure on health services for the elderly.

1. A more critical attitude toward health services diminishing the quest for medical consumption
2. Growth of a sense of responsibility of individuals for their own health, resulting in changes in lifestyle involving less high-risk behavior
3. Less pressure on institutional care and cure; the elderly live on their own for a longer time
4. Less solidarity between generations
5. More solidarity within the generation of the elderly; help from senior to senior increases
6. Norms and values in society incorporate the elderly more fully into the dynamics of social life
7. The elderly are less dependent than in 1984, not because of financial position, but because the social climate has changed
8. New types of living together by the elderly grow in importance ■

Critical incidents could be added to these elderly healthcare scenarios to produce substantial perturbations to the contexts described in them. Scenarios, with impacts of critical incidents considered, need to be linked to common background assumptions in medicine, technology, and economics to provide the basis for developing options to guide health care policy for the elderly in the Netherlands.

13.5 CRITIQUING SCENARIOS

There are two instances when it is necessary to critique a set of scenarios: (1) forecasters critique scenarios before presenting them to a user group or (2) technology managers critique scenarios presented to them. The task in either instance is to determine whether the scenarios form an adequate basis for making decisions about the future—decisions that may require resources to be committed today. As usual, such decisions involve risk.

Forecasts cannot be evaluated on the basis of their outcomes, since these are in the future. The only viable option is to assume that the inputs—in this case, the scenarios—are surrogates for the output. Thus the manager must establish criteria on the input side for evaluating the scenarios and then apply them. The result of this evaluation determines the credence to be placed in the scenarios.

Following Porter and Rossini (1977), the major criteria for evaluation are utility and validity. *Utility* addresses the usefulness of the forecasts to the user group. It is determined by the goodness of fit between what the forecaster has produced and what the user needs. *Validity* is determined by whether the forecast predicts what it claims to predict. Useful subcriteria for validity include:

1. The degree to which techniques used in the scenarios reflect the state of the art

2. The quality of information used in the scenarios, as compared to the best information available
3. The internal consistency of the scenarios
4. The plausibility of the scenarios (Do the future histories used make sense? Do they lead to snapshots that are credible? Can the snapshots be reached by plausible future histories?)

Using these criteria does not guarantee the validity of any scenario. However, if applied carefully, they should allow a manager to assess the level of confidence that can be placed in the forecast. When applied by the forecasters, they will provide insight into revisions, if any, that should be made.

13.6 CONCLUSION

Scenarios are outline sketches of some aspects of the future that are useful in forecasting. They may be future histories—that is, dynamic time paths from the present to some point in the future. They also may be snapshots—structural cross-sections of a future state. Combinations of these are possible as well.

Scenarios are useful as stand-alone (independent) forecasts if data for time series are lacking, if expertise is weak or nonexistent, and if there are no solid bases for model building. Even if other forecasts are possible, scenarios may be used to integrate disparate forecasts and all sorts of qualitative and quantitative information. They also can be extremely effective vehicles for communicating forecasts to non-technical audiences because they can use a wide variety of literary and multimedia techniques.

Scenarios provide a basic technique for the forecaster, planner, technology manager, or policy maker because of their flexibility in dealing with a wide range of forecasting contexts. Scenarios always must be checked for utility and validity. They should be based on the best available information and techniques and prepared so that they will be consistent and plausible to the user.

EXERCISES

- 13.1 Changes in the political and economic order in Eastern Europe and the Soviet Union imply short- and long-term impacts on farmers and grain dealers. Construct a morphology for a set of scenarios requested by a major commodities exporter.
- 13.2 High-definition television technology can be a major factor in determining the direction of home electronics producers and retailers operating in the United States. Construct a template and a set of three scenarios to be used by a major home electronics manufacturer operating in the United States in deciding what resources should be invested in the R&D of this technology today. Consider regulating and marketing as well as technological issues.

- 13.3** In many midwestern states the vast majority of electrical power generation is fired by coal (for example, in Indiana over 90 percent). Local coal resources are primarily of medium- and high-sulfur content. Environmental legislation, especially the 1990 Clean Air Act, will have a major impact on power generation and coal mining in these states. Prepare a set of scenarios for a technology manager charged with making decisions about the expansion of a major surface mine in the Midwest.

--- 14

ECONOMIC FORECASTING AND ANALYSIS

OVERVIEW

Technology and economic activity are inextricably intertwined. Technological change is perhaps the most significant source of economic growth, and economic factors are crucial to the success of new technology. This chapter describes the interrelationship between economics and technology and shows how economic forecasting can aid technology forecasting and management. Tools of economic analysis and their use by managers are presented. The chapter begins by describing how technological progress has enhanced economic well-being. Demand, its determinants, and forecasts are considered next and illustrated by example. Finally, input/output analysis is introduced as a powerful tool to aid in decision making.

14.1 INTRODUCTION

World resources are scarce relative to human needs and desires; therefore, choices are necessary. Economics is really a science of choices—how they are made and how they ought to be made. Because economics must deal with the complexities of human behavior, accurate prediction is difficult. The tools of the economist and the technology manager are not as precise as those used by the physical scientist. Nonetheless, economic tools can be used to aid in decision making. Economic inquiry is crucial to understanding how technologies become integrated into the fabric of economic activity.

To the economist, demand is the willingness to buy plus the ability to pay. Technological innovation only occurs when consumers, businesses, and governments

demand it. On the one hand, a technology launched when profits are high and unemployment is low has a greater chance of success than one offered when consumers are insecure and businesses are struggling to survive. On the other hand, new technology has led societies out of recessions and depressions when production has stimulated investment and recovery. Technology managers must understand these relationships and how to integrate economic analysis with management decisions.

14.2 TECHNOLOGY AND THE ECONOMY

Over 30 years ago, Nobel laureate economist Robert Solow (1957) concluded that most of the increase in the standard of living in the United States was due to technological progress. Specifically, he concluded that 87.5 percent of increased output per capita from 1909 to 1949 was due to technical change; the remainder was due to capital investment. Subsequent studies have refined this analysis. However, most economists would agree with Edwin Mansfield (1989, p. 700) that "the rate of technological change is perhaps the most important single determinant of a nation's rate of economic growth." Clearly, technology has created new products to enhance the quality of life. It also has dramatically improved the productivity of labor and capital resources.

People can produce more in less time if they know how to use resources more effectively. In the early twentieth century, for instance, Henry Ford could raise the wages of automobile assemblers to \$5 a day by production innovations. Later, American factory workers could reduce their work week to 40 hours while continuing to improve their incomes. More recently, Japan, South Korea, Taiwan, and other Asian countries have gone from devastated nations to world economic powers thanks to the modern technology used in their production processes.

New technology can overcome inherent limitations imposed by the *law of diminishing marginal returns*. This law says that if the input of a resource is increased in a production process for which supplies of some other resources are fixed, at some point the additional output that results from one more unit of the input will decline. For instance, machinery and equipment are scarce (fixed resources) in China, but there are a lot of workers. Thus simply adding workers to the production process will not add much to output. On the other hand, Japan has a large inventory of capital equipment in relation to the number of workers. Therefore, adding more of the same equipment could stretch the limits of the labor force and not increase output very much. With technological improvement, both capital and labor resources can be used more efficiently. Thus increased inputs together with the improved technology can mean greater productivity and higher worker income.

It is important to stress that a new idea must be incorporated into a new product or process to produce a benefit. This process of innovation does not automatically follow the discovery of new knowledge. In fact, *Business Week* (1989, p. 15) noted that according to Rustum Roy "The more inventive ideas the U.S. dreams up, the further it will fall behind. Each one will be just another opportunity for a foreign rival to out-innovate a U.S. company in producing it." Evidently if you invent a

better mousetrap the world does not necessarily beat a path to your door— it may take your idea and beat you to the marketplace.

A conclusion that inventions are counterproductive is too strong. However, it illustrates that merely increasing research and development (R&D) expenditures to produce more inventions is too simplistic. In fact, Drucker (1985) stresses that new knowledge is not the most attractive source of innovation. The chances of success are lower and lead times longer for technology based on new knowledge than for that based on other sources of innovation. He suggests that the characteristics of the market and the production process are better indicators of opportunity than the newness of the knowledge base. Drucker sees the entrepreneur as one who systematically analyzes opportunities and develops strategies in response to inevitable change.

Understanding the role of technology in the economy is crucial to the success of firms in that economy. The technology manager needs an entrepreneurial perspective that welcomes change as a source of opportunity. No markets are static, and their dynamic nature can topple dominant firms and raise new, innovative enterprises. Further, the manager must understand the need for technology to be appropriate.

The phrase "appropriate technology" was coined by economist E. F. Shumacher (1973). His purpose was to adapt the thinking of Ghandi to the subject of economics. India and many other underdeveloped countries had suffered because their leaders had imposed Western technology. Shumacher called for smaller-scale production processes that take into account a people's characteristics and culture. Including this human dimension broadens the view taken by traditional economics; however, the lessons are not inconsistent. Conventional economic theory would agree that the large, unskilled, underemployed labor force of an underdeveloped area is not effectively utilized by large-scale, sophisticated, labor-saving production approaches. Furthermore, the effects of products and manufacturing processes on the people who run them are known to affect productivity and quality as well as workers' feelings about themselves and their lives. Using a multidimensional perspective in selecting and developing technology has been part of the reason that the Japanese have succeeded so well in the world market. Therefore, this perspective is clearly very relevant to technology managers.

14.3 MARKETS AND INNOVATION

Traditional economics asserts that market conditions determine the success of a technology. Resource prices, for instance, are guides to the current scarcity of the inputs to production. If labor is becoming scarce (and therefore more expensive), innovations that make labor more productive will succeed. Robotics, for example, may be widely adopted because labor costs are high in developed economies such as those of Japan, Western Europe, and the United States. On the other hand, if interest rates (the resource price of money) are high, then innovations that improve or extend the life of existing equipment are more attractive.

Successful technology responds to market conditions. The development of high-rise building technology, for instance, was a response to high rents in congested central business districts. Likewise, many energy conservation technologies (such as light-weight, high-strength materials in automobiles) were adopted because of large petroleum price increases in the 1970s. As raw material and disposal costs rise, innovations to facilitate recycling will become increasingly attractive.

Changes on the demand side of markets also invite innovation. For example, the exercise craze of the 1980s created an attractive environment for the Sony Walkman®. And high-definition television may become attractive in the 1990s if families prefer entertainment at home and are willing to pay for enhanced picture quality. Such changes in tastes may be hard to quantify, but other determinants of demand are more quantifiable and predictable. Rising incomes, for instance, increase the demand for luxury goods and gadgets. Likewise, the demographic reality of increasing numbers of the elderly will make innovations in prosthetic devices (such as effective artificial hips) attractive. In a similar fashion, falling computer hardware prices make developments in software application more lucrative.

Knowing market conditions obviously is not enough. A manager must be able to visualize how a new technology will fit into future conditions. This is difficult. In fact, Drucker (1985) advises against innovating for the future because an economically effective technology needs quick application to produce returns on investment. However, the manager who does not innovate for the future forfeits the opportunity to leapfrog competitors and be the first to take advantage of emerging opportunities.

Forecasting techniques (see Chapter 10) provide ways to extrapolate current trends in important market variables. Hilbrink (1989) suggested decomposing technologies into their component parts and examining trends in each component to give indications of potential for growth and economic impact. For example, data processing systems could be divided into components such as the central processor unit, transmission systems, memory, storage, and so forth. Hilbrink notes that large data processing system architecture is sensitive to the balance between memory and data transmission costs. His analysis suggests that transmission costs will fall faster than memory costs. If so, the design of system architecture may take advantage of this future development by emphasizing telecommunications connections rather than on-site processing. Since changes are occurring very rapidly in many technologies in addition to computing, approaches such as the decomposition method suggested by Hilbrink offer advantages over just innovating for now.

Another approach to analyzing future conditions was suggested by McIntyre (1988), who noted that innovators should examine the infrastructure needed to support a new technology, in addition to examining the usual growth curves for it. In the short run, there may be no market for a new product because the infrastructure is not available for its effective use. Television, for example, was not very attractive to consumers before broadcasting facilities had been widely established. More recently, while home computers have not met the ambitious sales projected for them, McIntyre says that they, like the automobile, could become a necessity if some reasonable applications are widely adopted.

Actually, the importance of infrastructure to new technology was recognized much earlier by Schumpeter (1934). Among the many important contributions of this economist and innovation scholar was an insight into the way markets for new products develop. In his study of the U.S. aluminum industry, Peck (1961) drew on Schumpeter's insight to show how the market for a new product develops more rapidly if its uses are demonstrated. Alcoa used this demonstration effect to help expand its market in the early twentieth century. The firm developed new applications for aluminum and priced them attractively to get manufacturers to use the metal. The increased demand made Alcoa's monopoly on primary ingot production much more profitable.

McIntyre (1988) proposed analogous strategies for today's high-tech firms. He stressed the need to stimulate market adaptation and described some specific approaches to do it. One approach is to ensure that product design includes the potential for growth or enhancement. For example, IBM designed expansion slots into its early PCs so that other vendors could manufacture plug-in circuit cards to provide additional features. Similarly, firms (such as computer manufacturers) can help create infrastructure by stimulating the development of other firms that make complementary products (such as software). At times, innovators may wish to form strategic alliances to establish compatible standards, a form of infrastructure.

The growth of a new technology also critically depends on competing and complementary technologies. Therefore, it is important to carefully analyze the markets for these associated technologies and to understand the markets on which the new technology itself depends. While the complete development of a market analysis is beyond the scope of this text, there are some basic tools of economic analysis with which the technology manager should be familiar. Economic forecasts, for example, are important to the timing of product introduction, and input/output analysis is very helpful in understanding the interdependencies of the various markets. The following sections provide an introduction to these concepts.

14.4 FORECASTING THE ECONOMY

New technology can stimulate the economy and help bring about higher levels of investment and economic growth. However, an individual innovator must respond to existing and projected economic conditions because, alone, he or she cannot have much impact on them. Therefore, it is important to be aware of economic conditions and how they will affect the development of a new technology.

Over the past century, industrial economies have generally grown; however, surrounding this basic growth trend there have been business cycles of varying severity. In an expansion phase, output grows, employment is high, and businesses see rising profits. Consumers have good incomes and are likely to be more receptive to new products. Business profitability makes it possible to invest in new capacity that incorporates new technology. On the other hand, a recession can lead to business losses and rising unemployment. Consumers tend to limit buying to necessities and to make do with old durable products. Businesses are faced with

excess production capacity, and, as a result of reduced consumer demand, they feel less pressure to install new technology. It is dangerous to generalize because other factors may influence the success of a new technology. However, it is clear that recessions are probably not good times to launch new technologies; economic expansions, on the other hand, may help establish new technologies.

There are many theories about business cycles and their control (see, for example, Dornbush and Fischer, 1987, especially Chapter 10; Volland, 1987). Although the technology manager does not need the depth of analysis required of the corporate economist or financial manager, he or she should have a basic understanding of what affects the level of economic activity and how to get reasonable forecasts. This is important information as the vitality of the economy, the level of inflation, or changes in interest rates could have significant effects on a new technology. Although economic forecasting, like weather or technology forecasting, is not an exact science, it can help.

The basis of most forecasting of national economies is the model given by the following:

$$\text{Output} = \text{Consumption} + \text{Investment} + \text{Government} + \text{Net exports} \\ \text{spending}$$

or

$$Y = C + I + G + N \quad (14.1)$$

Consumption (C) includes everything from *services* (such as haircuts) to *nondurable goods* (such as food) to *consumer durables* (such as cars and appliances). Investment (I) is output that goes into structures, (such as homes, factories, office buildings), equipment (such as computers, robots, rolling mills), and inventories of materials or products that are held for later sale. Government spending (G) includes federal, state, and local purchases. These may be financed by taxes or by borrowing. The former directly affect consumers and investors. The latter affect credit markets and interest rates directly and the levels of C and I indirectly. Net exports (N) are sales of production to foreign buyers, less purchases of goods and services from abroad. Higher exports mean higher production and incomes in the domestic economy, while imports replace domestic production with foreign-made goods and services. Since the level of demand effectively determines the level of output (Y), the terms "output" and "demand" often are used interchangeably.

Most forecasts directly or indirectly estimate future levels of the four components (C , I , G , and N) of the aggregate demand for production output. Some forecasters (such as Data Resources, Inc., Chase Econometrics, Federal Reserve Board/MIT, and The Brookings Institution) use huge, computer-based, econometric models to make these estimates. These models employ many equations developed statistically in a manner analogous to the regression techniques discussed in Chapter 9. However, such models have not demonstrated significant superiority over more qualitative approaches.

For this reason, judgment or opportunistic forecasting is still very common. In such forecasts, the analyst uses the best information available to estimate each component and sums the estimates to predict demand, as in Equation 14.1. However, the individual components are not independent of one another. For example, personal consumption depends on disposable income, which, in turn, is affected by the level of government spending. The more the government spends, the higher incomes will be. In other words, consumption depends on government spending. Thus estimates made independently of one another can lead to internal inconsistencies and an erroneous forecast. To avoid this, the forecaster must iterate estimates to establish the internal consistency of the forecast.

Other forecasts use indirect approaches. Table 14.1 lists the leading, coincident, and lagging indicators provided by the U.S. Department of Commerce in its *Business Conditions Digest*. As the names imply, these data provide some indirect insight into what is happening to the components of economic activity over time.

TABLE 14.1 Economic Indicators

<i>Leading Index Components</i>
Average workweek, production workers, manufacturing
Layoff rate, manufacturing
Net orders for consumer goods and materials (1972 dollars)
Vendor performance, percentage of companies receiving slower deliveries
Net business formation (index: 1967 = 100)
Contracts and orders for plant and equipment (1972 dollars)
New building permits, private housing units (index: 1967 = 100)
Net change in inventories on hand and on order (1972 dollars)
Change in sensitive crude materials prices
Change in liquid assets
Stock prices, 500 common stocks (index: 1941–1943 = 100)
<i>Roughly Coincident Index Components</i>
Employees on nonagricultural payrolls
Personal income less transfer payments (1972 dollars)
Industrial production, total (index: 1967 = 100)
Manufacturing and trade sales (1972 dollars)
<i>Lagging Index Components</i>
Average duration of unemployment
Manufacturing and trade inventories (1972 dollars)
Labor cost per unit of output, manufacturing (index: 1967 = 100)
Average prime rate charged by banks
Commercial and industrial loans outstanding, reported weekly by large commercial banks
Ratio of consumer installment credit to personal income

Source: Compiled from indicators provided by the U.S. Department of Commerce, *Business Conditions Digest*.

For example, if the average workweek is lower and the layoff rate is higher, one would expect this to lead to less demand for, and hence production of, consumer goods at some later time. Changes in money supply also can precede economic change. For instance, monetarist economists, such as Nobel laureate Milton Friedman, place great importance on the role of money in economic activity. Therefore, they maintain that changes in money supply can be powerful predictors of later levels of prices and economic activity.

Technology managers are unlikely to find it in their best interest to generate their own economic forecast. However, relying on forecasts by a single expert is unwise because none have amassed consistent records of superior accuracy. Perhaps the best approach is to invest small amounts of time and money to examine the consensus outlooks that are readily available. *Business Week*, *Fortune*, *Forbes*, *Industry Week*, and other business magazines present forecasts, as do the *Wall Street Journal*, *New York Times*, and other major newspapers. Generally, each quotes a range of economists, and collectively they provide a broad view of the economic outlook. Businesses also can buy a consensus forecasting service such as *Blue Chip Economic Indicators*. This can be a cost-effective way to get a balanced picture of predictions by recognized people.

14.5 INPUT-OUTPUT ANALYSIS

The preceding section described tools for understanding the general economic climate for technology. Recall that the discussion of forecasting (see Chapter 4) noted the interaction between the various parts of the economy. There are tools that explicitly take account of these interactions and provide information about the way in which changes in one sector can affect others. One tool, input-output analysis, is useful both to identify the market potential for products and to show the impacts of new technologies on the economic system.

Input-output analysis was created and continues to be developed by Wassily Leontief (1985). In this analysis, the economy is modeled as a static structure that represents the flows of production from one industry to another and to the final demand components described in Equation 14.1. Input-output analysis breaks down the interacting transactions (flows) necessary to make a product to show which components of the economy must supply what types of production. The transactions must be measured in some unit, with monetary terms the most convenient. While it may be possible, for example, to show how tons of steel and glass and kilowatt hours of electricity can be added to make one car, it is much easier to add so many dollars worth of steel, glass, and electricity to the costs of other components to arrive at the cost of producing a car.

Consider an economy that is composed of only two industries, X_1 and X_2 . Table 14.2 shows what happens to the output of these two industries. (Table 14.2 is similar to the interindustry transactions tables published by the U.S. Department of Commerce in the *Survey of Current Business*. While our example is limited to two industries for simplicity, the same principles apply to the government's data.)

TABLE 14.2 Interindustry Transactions

	Transactions in \$						Total Output
	X_1	X_2	C	I	G	N	
X_1	100	200	400	100	100	100	1,000
X_2	500	600	200	300	200	200	2,000
Wages	300	700					
Other	100	500					
Total	1,000	2,000					

The first row, for example, shows how much of the total output of X_1 (\$1,000) is sold from one firm to another within industry X_1 (\$100), to industry X_2 (\$200), and to the four final demand sectors indicated in Equation 14.1. The second row gives a similar breakdown for industry X_2 (\$2,000 total output). Although the analysis could be made in terms of physical units (such as tons of steel), measurement in monetary units will be very useful in subsequent steps of the analysis.

The "Wages" and "Other" entries in the table are incomes that represent the distribution of the value added by each industry. *Value added* is determined by sales minus the cost of the intermediate goods purchased from other industries. Note that the total of each industry's column equals the total of that industry's row.

The direct requirements of one industry can be calculated for the output of another from the information in the interindustry transaction table. Table 14.3 presents these requirements for our hypothetical two-industry economy as computed from the information in Table 14.2. Entries show the share of each dollar of output of an industry (column) represented by input to the other (row). For example, a dollar of output from X_1 requires \$0.10 ($100/1,000$) of its own output as an intermediate input plus \$0.50 of industry X_2 output plus \$0.30 of labor and \$0.10 of other income, such as rents and profits. (The direct requirements table for the United States also is published in the *Survey of Current Business*.)

While the direct requirements table is useful, it does not tell the whole story. The limitations of the direct requirements table can be seen if one asks: "How

TABLE 14.3 Direct Requirements Table

Input To	Output of Industry in \$	
	X_1	X_2
X_1	0.10	0.10
X_2	0.50	0.30
Wages	0.30	0.35
Other	0.10	0.25
Total	1.00	1.00

much would industry X_1 have to produce to sell an additional \$100 of output to consumers?" Since \$0.10 of each dollar of X_1 output goes for intermediate goods produced within the industry (see Table 14.3), its total output must be at least \$110 to produce an additional \$100 of output for consumers. However, there is more to consider. Table 14.3 shows that each dollar of output by X_1 uses an input of \$0.50 from industry X_2 and that X_2 requires $0.10 \times 0.50 = \$0.05$ of output from industry X_1 to produce it. Thus the total output of X_1 must be at least \$115—but that requires additional input from X_2 —and so forth. A convergent pattern of increasing output levels has clearly been established. This simple example helps explain why the direct requirements table is a building block for the more useful total requirements table.

This problem can be better understood by examining it in terms of two equations in two unknowns. First, assume that the final demands for the outputs of industries X_1 and X_2 are as portrayed by the interindustry transactions table (Table 14.2), *except* that the level of consumer purchases of X_1 output is \$100 higher (that is, $C = \$500$ instead of \$400). This results in the following equations:

$$X_1 = (100/1,000)X_1 + (200/2,000)X_2 + 500 + 100 + 100 + 100$$

or

$$X_1 = 0.10X_1 + 0.10X_2 + 800 \quad (14.2)$$

Likewise,

$$X_2 = 0.50X_1 + 0.30X_2 + 900 \quad (14.3)$$

These equations are constructed using direct requirements data from Table 14.2 and the increased level of consumer purchases. They define actual industry output levels for the \$100 increase in consumer demand for output of X_1 .

Solving Equations 14.2 and 14.3 simultaneously gives values for X_1 and X_2 of \$1,120.69 and \$2,086.21, respectively. Comparing these values with those in Table 14.2, the result of a \$100 higher level of consumer sales by X_1 would be to increase the output required of X_1 by \$120.69 and of X_2 by \$86.21. Thus every dollar of increased final demand for output from X_1 would increase the output for that industry by \$1.2069 and would increase the output of X_2 by \$0.8621. A similar exercise, assuming that final demand for X_2 increases, reveals that the direct and indirect effects on outputs total \$0.1724 of additional output for X_1 and \$1.5517 of additional output for X_2 .

The solutions to these equations can be used to construct a total requirements table (see Table 14.4). This table shows the direct and indirect effects of a one-dollar increase in the final demand for each industry's output. The total effects of a one-dollar increase in final demand for the output of an industry can be found by reading down the column for that industry in Table 14.4.

TABLE 14.4 Total Requirements Table: Direct and Indirect Effects per Dollar of Final Demand

An Increase in Demand for Output of This Industry Is Produced	By an Increase of \$1.00 of Demand for Output of This Industry	
	X_1	X_2
X_1	1.2069	0.1724
X_2	0.8612	1.5517

Matrix algebra can be used to apply these principles to the more realistic case in which the number of industries is quite large. Equations 14.2 and 14.3 can be written as

$$X = AX + Y \quad (14.4)$$

where X is a 2×1 matrix, the elements of which are X_1 and X_2 in the example; A is a 2×2 matrix, the elements of which are the entries of the first four cells in the direct requirements table (Table 14.3) expressed as fractions of industry total output; and Y is a 2×1 vector made up of the sums of the final demands for each industry's output from the components C , I , G , and N . For the two-industry example,

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} 0.10 & 0.10 \\ 0.50 & 0.30 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} + \begin{bmatrix} 800 \\ 900 \end{bmatrix}$$

Rearranging, the final demand, Y , is related to total output, X , by subtracting out intermediate goods (AX) so that

$$Y = (I - A)X \quad (14.5)$$

where I is the identity matrix. Solving for X uses the standard approach of inverting the $(I - A)$ matrix to get

$$X = (I - A)^{-1}Y \quad (14.6)$$

The elements of the $(I - A)^{-1}$ matrix are the entries in the total requirements table. They show the direct and indirect effects of an increase in demand for the industry heading each column. For our simple example,

$$(I - A)^{-1} = \begin{bmatrix} 1.2069 & 0.1724 \\ 0.8621 & 1.5517 \end{bmatrix}$$

Since computers are well suited to this kind of matrix manipulation, it is possible for the U.S. Department of Commerce to consider not just two industries but nearly

100 industry groups. Their computations produce the total requirements table in the February 1979 issue of the *Survey of Current Business* in the manner shown here. For example, that table reveals that a one-dollar increase in final demand for paper board containers and boxes leads to direct and indirect output increases of \$1.05608 for the industry itself, a \$0.51828 increase in paper and allied products output and a total increase of \$0.06008 in the output of the chemicals and selected chemical products industry. Effects on all the other industry groups could be examined as well.

The implications of this total requirements table for those doing market planning should be clear. Each element in the row for the industry reveals the relationship of that industry to the final demands for output from each industry listed in the table. For example, suppose a new box-making technology is expected to double the output of boxes in the next decade. Since 6 percent (\$0.06008/\$1) of box-making costs are due to the chemicals industry, this change will be important for at least some firms that produce chemicals. Total requirements tables have been used to analyze the effects of economic changes ranging from decreases in defense spending to expanded investment in environmental protection technology. Isard (1972), for example, showed how input-output analysis could be used to analyze the environmental effects of economic development. This analysis involved an extension of the model to explicitly include ecological inputs in a manner analogous to industrial and factor inputs.

There are some severe limitations on the use of input-output tables by technology managers. Perhaps most important is that the table assumes that the proportions of inputs to outputs remain constant and that relative producer prices do not change. These problems are somewhat alleviated by frequent updates of the tabulated data, but as the technology manager is well aware, change can be rapid. This is not to say that the information provided is not helpful, but rather that the conclusions from input-output analysis should be treated as approximate. Leontief and others are actively working to show how technological change can be incorporated in the analysis and better understood (see Leontief, 1985, for discussion of the analysis of technology).

14.6 CONCLUSION

Technology and the economy are closely interrelated. In fact, the rate of technological change is considered by many to be the primary variable in economic growth. However, from the technology manager's perspective, the influence of the economy on the success of a new technology is of more central concern. It is important for the manager to understand how economic conditions affect the chances for success of the technology he or she manages, and it is crucial that decisions take into account the present and expected future states of the economy.

Economists study how scarce resources are allocated to satisfy human needs. It is unlikely that many technology managers can or should generate their own economic forecasts. The track record of economic forecasts suggests that managers should follow consensus forecasts for the short term made by relatively simple

methods and frequently updated. These are widely available from several sources. Managers must understand economic factors sufficiently well to determine what the forecasts portend for their technology. Long-term economic forecasting is not well established. "Long wave theories," mentioned in Section 2.1 and 10.4.2, deserve consideration.

Managers must visualize how new technology will fit into present and future conditions, and they must understand how changes in the economy can offer opportunities. Therefore, the technology manager should develop strategies to take advantage of inevitable changes in the economy.

Finally, managers must develop their own philosophy about the time horizon for which they will develop new technologies—or understand and apply the philosophy of the corporation for which they work. Although innovating for the future implies greater risk than developing for the present, it also offers significantly greater rewards.

EXERCISES

- 14.1 Technology can increase productivity. List the technological developments of the past 30 years that make students more productive.
- 14.2 Events of the 1970s dramatically increased the price of energy. How have higher energy prices affected the technological characteristics of three products in use in the 1990s?
- 14.3. Pick a familiar technology. Conceptually break it into component parts. How are the prospects for technological progress different for these parts? How might you take advantage of these differences as a manager in an industry that produces this technology? In an industry that uses it?
- 14.4 Trace the important infrastructure and related technologies that developed to make the automobile so central to modern life in industrialized countries. How would a firm use this information when considering the development of a vehicle for use in third-world nations?
- 14.5 Use the sources suggested in Section 14.4 to develop a forecast of the U.S. economy for the next year. Identify several new technologies whose introduction and growth will be especially inhibited or helped by these economic conditions.
- 14.6 Find the most recent total requirements input-output table for the United States (examine the most recent issues of the *Survey of Current Business*, published by the U.S. Department of Commerce). Specify information from that table to use in market planning in an industry of interest.

14.7 Construct a total requirements table:

(A) If the interindustry transaction table is as follows:

Transactions (in \$)							Total Output
	X_1	X_2	C	I	G	N	
X_1	150	150	300	160	80	160	1,000
X_2	550	350	150	250	300	200	1,800
Wages	200	700					
Other	100	600					
Total	1,000	1,800					

- (B) If $I = \$300$ for X_1 output and everything else remains the same, what are the new values of X_1 and X_2 ?
- (C) If $C = \$270$ for X_2 output and everything else remains the same, what are the new values of X_1 and X_2 ?
- (D) If $G = \$250$ for X_2 output and everything else remains the same, what are the new values of X_1 and X_2 ?